Tuesday, July 28, 2026, 8:00 AM - 9:15 AM

TA01

301A

LLM & Reinforcement Learning for Healthcare Applications

Invited Session

AI in Healthcare

Chair: Esmaeil Keyvanshokooh, Mays Business School, Texas A&M University, College Station, TX, United States

Co-Chair: Gian-Gabriel Garcia, University of Washington, Seattle, WA, United States

1 - Rubric-Aligned Fine-Tuning via Influence-Guided Optimization

Sean Xu, Carnegie Mellon University, Pittsburgh, PA, United States, Holly Wiberg, Ramayya Krishnan, Cathy Jiao, Chenyan Xiong

Domain adaptation of large language models (LLMs) requires alignment with domain-specific capabilities that are inherently multidimensional, such as factual accuracy, completeness, and domain relevance. However, commonly used automatic evaluation metrics for generative tasks fail to adequately capture these requirements, while expert-based evaluation is costly and difficult to scale, particularly in high-stakes domains. Recent advances in LLM-as-a-judge, equipped with structured, domain-specific rubrics, enable automated and scalable evaluation of LLM-generated outputs, producing interpretable feedback and rubric-based scores that approximate a domain-aligned reward signal. Beyond quantifying alignment, this capability creates an opportunity to incorporate evaluation directly into the training process to guide domain adaptation. Leveraging these evaluation signals, we propose a rubric-aligned fine-tuning framework that formulates adaptation as a budget-constrained optimization problem aimed at maximizing improvement on a multidimensional domain-specific rubric. The framework introduces an influence-guided mechanism to estimate each training example's contribution to rubric improvement, enabling efficient data selection for fine-tuning. It further incorporates iterative rubric reweighting based on evaluator-identified deficiencies to promote balanced improvements across dimensions. We instantiate the approach on clinical note generation from doctor–patient dialogues as a use case in the healthcare domain adaptation. This work establishes a general paradigm for aligning LLM adaptation with domain-specific evaluation through automated feedback.

2 - Leveraging Observational Data for Adaptive Clinical Trials: Deconfounded Warm Starts and Adaptive Posterior Fusion

Prateek Jaiswal, Purdue University, West Lafayette, IN, United States, Esmaeil Keyvanshokooh, Junyu Cao

Randomized clinical trials often require large patient cohorts before reliable conclusions can be drawn, while abundant observational data remains underutilized due to confounding and hidden biases. In this work, we propose a principled framework for incorporating such data into sequential clinical decision-making. We first introduce Deconfounded Warm-Start Learning (DWSL), which leverages a Doubly Debiased LASSO (DDL) procedure to identify a sparse set of reliable measured covariates and construct a reduced context. These estimates are used to initialize Thompson Sampling with informative priors on measured features while retaining uninformative priors on hidden covariates, enabling effective and robust warm-starting.

Building on this foundation, we develop an extension based on an Adaptive Product-of-Experts (APoE) posterior that dynamically combines offline and online information. The method uses a mismatch score to modulate the influence of offline data over time, allowing the algorithm to exploit high-quality observational data while remaining robust to distributional shifts and model misspecification. We establish regret guarantees that recover the best of both worlds: when offline and online environments are aligned, the algorithm achieves improved performance reflecting the effective offline sample size; under mismatch, it gracefully falls back to standard online learning rates. Empirical evaluations on synthetic and semi-synthetic healthcare datasets demonstrate

consistent gains over existing methods, highlighting the potential of our approach for improving the efficiency and personalization of clinical decision-making.

3 - From Restless to Contextual: A Thresholding Bandit Reformulation for Finite-Horizon Improvement

Jiamin Xu, Cornell University, New York, NY, United States, Ivan Nazarov, Aditya Rastogi, África Periáñez , Kyra Gan

This paper addresses the poor finite-horizon performance of existing online \emph{restless bandit} (RB) algorithms, which stems from the prohibitive sample complexity of learning a full \emph{Markov decision process} (MDP) for each agent. We argue that superior finite-horizon performance requires \emph{rapid convergence} to a \emph{high-quality} policy. Thus motivated, we introduce a reformulation of online RBs as a \emph{budgeted thresholding contextual bandit}, which simplifies the learning problem by encoding long-term state transitions into a scalar reward. We prove the first non-asymptotic optimality of an oracle policy for a simplified finite-horizon setting. We propose a practical learning policy under a heterogeneous-agent, multi-state setting, and show that it achieves a sublinear regret, achieving \emph{faster convergence} than existing methods. This directly translates to higher cumulative reward, as empirically validated by significant gains over state-of-the-art algorithms in large-scale heterogeneous environments. The code is provided in \href{https://github.com/jamie01713/EGT}{github}. Our work provides a new pathway for achieving practical, sample-efficient learning in finite-horizon RBs.

4 - Constraint-Aware Self-Improving Large Language Model for Clinical Role Model Generation

Esmaeil Keyvanshokooh, Mays Business School, Texas A&M University, College Station, TX, United States, Che-Yi Liao, Gian Garcia

Personalized treatment planning requires clinicians to identify feasible treatment trajectories tailored to each patient's profile. Clinical Role Models (CRMs), i.e., individuals with similar health profiles whose successful treatment trajectories serve as clinician-facing reference points, can anchor these decisions in concrete, patient-relevant examples. However, identifying suitable CRMs is fundamentally limited by data scarcity; even in data-rich settings, underrepresented patient subpopulations may have few or no reliable matches in the local database. Foundation Medical Large Language Models (MLLMs) can synthesize realistic CRMs to bridge this gap, but risk generating clinically invalid outputs and are susceptible to inconsistencies from evolving risk-scoring tools. This raises a fundamental question: can generative AI reliably produce synthetic patient profiles that satisfy verifiable clinical constraints? We develop CASE (Constraint-Aware active Self-improving fine-tuning), a framework that pairs an MLLM with an optimization-based verifier to ensure generated CRMs are clinically valid and robust to risk-scoring uncertainty. Verified examples are fed back to fine-tune the MLLM in a self-improving loop, guided by an active learning algorithm that prioritizes the most informative patient cases. We design a loss function that penalizes unreliable outputs, encourages lower patient effort-to-change, and preserves linguistic quality, and prove that the learning algorithm achieves sublinear regret with patient-specific safety guarantees. Our framework demonstrates that pairing LLM generation with optimization-based verification can overcome data scarcity bottlenecks in data-driven clinical optimization, improving care for underserved populations.

TA02

301B

Managing Waits in Healthcare Operations

Invited Session

Applied Probability

Chair: Opher Baron, University of Toronto/ SiMLQ, Toronto, ON, Canada

Co-Chair: Dmitry Krass, Rotman School of Management, University of Toronto, 105 St George St, Toronto, ON, M5E1S7, Canada

1 - Learning from Clinical Notes: LLM-Derived Insights for Capacity Planning and Scheduling in Multi-Appointment Care

Larissa Troper, University of Waterloo, Waterloo, Ontario, Canada, Hossein Abouee Mehrizi, Hamid Arzani, Parsa Behmanesh, Brendan Wylie-Toal

Healthcare centers generate large volumes of clinical notes, particularly in multi-appointment care settings such as pediatric rehabilitation, where patients receive treatment over multiple sessions and clinicians document progress at each visit. These notes form rich longitudinal records capturing patient progress and contextual factors affecting treatment trajectories and resource needs, yet they remain largely underused for operational decision-making because they are unstructured and difficult to analyze at scale. Leveraging advances in Large Language Models (LLMs), we develop a scalable framework that transforms narrative clinical notes into structured signals for operational planning. The pipeline extracts patient-level indicators, such as disorder types, progress trends, and contextual factors, and is evaluated using standard AI metrics with human-in-the-loop validation by clinical partners. Using these signals, we construct patient learning curves summarizing progress across sessions and integrate them into a mathematical optimization model that determines reassessment policies, where reassessments serve as checkpoints for continuation or discharge decisions. The resulting policy balances the capacity cost of reassessment sessions with the risk of incorrect stopping decisions under uncertainty, providing a data-driven approach for capacity planning and scheduling in multi-appointment care systems.

2 - Impact of Shorter Emergency Department Waits on Patient Arrivals

Opher Baron, University of Toronto/ SiMLQ, Toronto, ON, Canada, Dmitry Krass

The impact of long waiting times in Emergency Departments (ED) on poor health outcomes has been well established, as has the linkage between long waits and Left Without Being Seen (LWBS). Less well understood is the extent to which shorter ED waits are causally linked to increased patient volumes. In this talk we discuss several initiatives focused on improved ED patient flows at several different Canadian hospitals. We discuss the specific interventions implemented, as well as the analytical tools we have developed to support these initiatives and analyze their impacts.

To support on-going ED operations we introduce SiMLQ – a data-driven automated process simulator that allows ED staff to identify key patient pathways, potential bottlenecks, and possible process interventions. The user can then simulate the impacts of these interventions under different ED states.

To analyze the long-term impact of ED improvement projects we use a simultaneous equation model allowing us to estimate complex feedback mechanisms in the system.

We demonstrate empirically that flow-improving initiatives lead to significant service improvements, which, in turn, lead to higher downstream patient arrivals. Moreover, the improved system handles higher patient volumes without worsening wait times or need for additional resources.

3 - Shift Effects in the Emergency Department: Threshold Routing and Tail Access Performance

Hongsong Yuan, Shanghai University of Finance and Economics, shanghai, China, People's Republic of, Opher Baron, Simai He, Dmitry Krass, Parvin Malekzadeh

Emergency departments (EDs) frequently face shift effects, where physician behavior and patient access to care measures, such as Time to Physician Initial Assessment (TPIA) and Left Without Being Seen (LWBS), vary within a shift. In EDs physicians choose between serving newly arriving patients and returning patients. We model this system as a finite-horizon two-station tandem queue with abandonment of new patients and a buffer on the in-system queue length. We characterize the throughput-maximizing policy in a deterministic fluid approximation and show it has a time-threshold structure. Furthermore, we conduct full scenario analysis for different traffic intensities, and derive closed-form solutions for access to care measures in each scenario. We evaluate the mean and upper tails of these measures under the optimal policy and benchmark them against the baseline strategy that prioritizes returning patients baseline. We further show that applying an analogous time-threshold policy to the stochastic system is asymptotically optimal.

Numerical results reveal a surprising, non-monotone V-shaped relationship between the buffer level and the upper tails of TPIA and LWBS rate. Managerially, this implies that to minimize the upper quantiles of TPIA and

LWBS rate, EDs can tune the buffer level to a moderate value. In practice, this can be implemented through flow coordinators, triage liaison roles, or similar release-control protocols.

4 - Tackling Geographical Health Disparities: Hospital Collaborations in the Age of Telemedicine

Nan Liu, Boston College, Chestnut Hill, MA, United States, Yunting Shi, Sergei Savin, Guohua Wan

Rural populations often lack access to quality care compared to their urban counterparts, and health disparities between urban and rural residents are well documented. Telemedicine offers a promising solution to bridge this gap by enabling remote collaboration between urban and rural hospitals. Although such collaboration can improve care quality at the rural hospital, it may also strain the urban hospital, which must reallocate part of its capacity to support rural care delivery. This creates a tension between improving rural care access and managing urban hospital resources.

This study explores telemedicine-based collaboration between an urban and a rural hospital. We develop a three-stage Stackelberg-game model to characterize the interactions among the social planner, the two hospitals and patients. In our model, the social planner sets the compensation amounts for the two hospitals, the urban hospital proposes a capacity and revenue-sharing contract to the rural hospital, and patients choose their care facility.

Our analysis reveals that the extent of telemedicine collaboration depends jointly on reimbursement rates, hospital capacity levels and demand dynamics. Importantly, telemedicine collaboration does not always lead to improved social welfare. However, a social planner can induce welfare-maximizing strategies by appropriately adjusting the relative per-patient reimbursement rates across the two hospitals. Our research highlights the significance of both financial incentives and operational characteristics in shaping telemedicine-based collaboration. Both economic and operational levers are critical for designing sustainable telemedicine strategies that improve rural healthcare access.

TA03

302A

Empirical Healthcare Management: Capacity and Throughput Under Pressure

Invited Session

Healthcare Operations Management

Chair: Max Yakovlev, Vanderbilt University, Nashville, TN, United States

1 - Maximizing Throughput: Selective Admissions in Nursing Homes

Adriana Corredor Waldron, North Carolina State University, Raleigh, NC, United States

How do nursing homes respond to price increases? This paper studies a large, federally mandated increase in Medicare reimbursement for short-term nursing and rehabilitation services. I exploit variation across facilities in pre-reform patient case mix and the Medicare payment schedule to construct a simulated revenue shock. Using linked Medicare cost reports and resident-level nursing home assessments, I find that a 10 percentage point increase in simulated revenue raises Medicare admissions by 18.7 percent, with no evidence of crowd-out for non-Medicare residents. The expansion is driven by a shift toward short-stay rehabilitation patients and away from clinically complex patients, even though the reform raised rates more generously for the latter. Facilities accommodate this shift by increasing therapy spending and substituting contracted labor for salaried labor, without expanding bed capacity. These findings show that Medicare reimbursement shapes not only the volume of post-acute care, but also its composition and organization.

2 - On Time-varying State Dependent Service Processes in Healthcare

Hao Ding, University of Miami, Miami, FL, United States, Sokol Tushe, Ruomeng Cui, Donald Lee

Understanding the drivers of efficiency in service processes has long been an important focus for the Operations Management (OM) community. A key finding in the literature is that service speeds are state-dependent. For example, in the context of emergency departments (ED), the productivity of a physician---and therefore the service speed experienced by her patient---depends on the workload of the physician as well as the overall census of the ED (the total number of patients present). As physicians manage multiple patients at the same time,

their workloads naturally require multitasking across patients. Studies in the healthcare OM literature often empirically treat workload and multitasking as the same measure, which is the number of patients a physician attends to at the same time. While it is now well accepted in the OM literature that workload, a key state variable of service systems, has a significant impact on service speed, there is less agreement on how: A wide spectrum of relationships have been reported that range from slow down as workload increases, to speed up, or to even more complicated patterns. In addition, these findings contrast sharply with those in the neuroscience and psychology literature, which consistently show that multitasking reduces productivity

3 - Strategic Substitution under Staffing Shortages: Use of Chemical Restraints in U.S. Nursing Homes

Kenneth Moon, Cornell SC Johnson College of Business, Cornell University, Ithaca, NY, United States

Approximately 20% of U.S. nursing home residents receive antipsychotic medications, imparting serious risks of stroke, heart failure, and mortality in elderly dementia patients. Despite decades of attention, the causes of this practice remain poorly understood. We investigate whether staffing shortages cause nursing homes to strategically sedate difficult patients, substituting drugs for labor-intensive care. Through FOIA requests, we obtain nursing homes' total antipsychotic usage rates free of the diagnostic exclusions that underreport in public data. We construct a panel of 10,453 U.S. nursing homes from 2014 to 2019 and use two independent sources of variation in staffing—local labor market wages and changes in state minimum staffing regulations—to establish causality. A one-hour increase in nurse staffing per resident day reduces antipsychotic use by 6.0 residents per home. This effect is concentrated among understaffed facilities and for-profit homes. Staffing affects usage without a qualifying clinical diagnosis, indicating that operational pressure not clinical need drives these decisions. Certified nursing assistant staffing has the largest impact, as drugs substitute for frontline hands-on care. State minimum staffing regulations effectively reduce antipsychotic use; we estimate that the federal staffing standards adopted in April 2024 and repealed in December 2025 would have reduced nationwide antipsychotic use by 27,952 residents (12.67%). Policies addressing facility-specific shortages are twice as cost-effective as uniform mandates.

4 - Impact of Travel Nurse Participation in Surgical Operations

Max Yakovlev, Vanderbilt University, Nashville, TN, United States, Eric Johnson, Vikram Tiwari

Under financial pressures and persistent nursing shortages, US hospitals increasingly rely on travel nurses to meet demand for care. While nursing has been an active focus of researchers, little is known about the operational effect of travel nurse participation on operating room (OR) performance. We aim to establish whether travel scrub nurses impact surgical throughput and process flow. Using OR case-level data from 270,000 surgeries linked to unique HR documentation from a large academic medical center, we analyze the impact of travel nurse participation on surgical performance. To address potential endogeneity, we instrument travel nurse participation using exogenous shifts in labor supply measured by state licensing approvals. Travel scrub nurse participation is associated with a 7-8% increase in surgical case lengths, underscoring the slowdown effect of travel nursing. Moreover, slowdowns are concentrated in specialties where surgeries are more equipment-intensive and settings where surgeries are planned. The performance gap between staff and travel nurses narrows as flexible workers accumulate institutional familiarity, yet remains meaningful even after six months of working at the hospital. As extensions, we also investigate travel scrub nurses' effects on surgical process flow. Between-surgery room transitions take longer when a travel nurse participates, and first-scheduled cases of the day are less likely to start on time. Rebalancing travel nurse staffing away from standardized services and encouraging longer contracts can reduce performance penalties associated with travel nurse participation. These policies can let hospital managers capitalize on the benefits of a flexible workforce while minimizing efficiency losses in the OR.

TA04

302B

Healthcare Data Infrastructure Modernization

Invited Session

Healthcare Operations Management

Chair: Mohammed Omer Shakeel Ahmed, UTHealth Houston, Irving, TX, United States

1 - Perationalizing Cybersecurity in Healthcare Systems: a Risk-Based Framework for Secure Medical Device Integration and Patient Data Operations

Shekar Munirathnam, IEEE, Cumming, GA, United States

Healthcare organizations face an uncommonly dangerous trifecta of cyber threats: an increasing threat surface due to the proliferation of network-connected medical devices, regulatory demands that necessitate both accessibility and protectability of health data, and an increasing threat level against hospitals and health systems because the disruption of operations directly impacts patient safety. The article introduces a risk-based framework for operationalizing cybersecurity in health environments, with specific attention paid to the integration of medical devices, operations of patient data, and the preservation of clinical workflow. The article discusses the threat landscape faced by healthcare delivery organizations, assesses the structural limitations of traditional IT security constructs in the context of clinical operations, and introduces a multilayer construct that meets the unique constraints of medical devices—legacy systems that cannot be patched, real-time monitoring that cannot experience latency, and life-critical operations that prioritize availability over confidentiality. The framework addresses HIPAA Security Rule compliance, FDA premarket and postmarket cybersecurity guidance, and newly introduced requirements under the PATCH Act and HHS Cybersecurity Performance Goals. The framework's effectiveness is demonstrated against published breach reports, and implementation advice is offered to healthcare security professionals working under the resource-constrained environment that characterizes the industry.

2 - Healthcare Data Analytics Using Microsoft Fabric: from Fragmented Data to Actionable Insights

Chinni Krishna Abburi, Communities Foundations of Texas, Irving, TX, United States

Healthcare organizations generate large volumes of data across clinical, operational, and financial systems, yet these data are often siloed, inconsistently modeled, and difficult to analyze at scale. This presentation demonstrates how Microsoft Fabric can be used as an end-to-end analytics platform to integrate, model, and analyze healthcare data to support data-driven decision-making.

We present a real-world analytics framework built using Microsoft Fabric that combines data ingestion, transformation, semantic modeling, and reporting within a unified environment. The approach leverages Fabric’s lakehouse architecture to centralize heterogeneous healthcare data sources, applies dimensional modeling to support analytical queries, and enables scalable analytics through integrated SQL, Spark, and Power BI capabilities.

The presentation highlights key design decisions, including data modeling strategies, governance considerations, and performance optimization techniques specific to healthcare analytics use cases. We also demonstrate how Fabric simplifies analytics workflows by reducing tool fragmentation while maintaining flexibility for advanced analytics and reporting.

Results from this implementation show improvements in data accessibility, reporting consistency, and time-to-insight for healthcare stakeholders. The session concludes with practical lessons learned, common pitfalls, and recommendations for organizations considering Fabric as a modern analytics platform in healthcare.

This talk is intended for analytics practitioners, data engineers, and decision scientists interested in scalable healthcare data analytics architectures and applied data engineering solutions.

3 - Beyond Efficiency: a Scoping Review of Patient Safety, Data Privacy, and Medico-Legal Challenges in AI-Driven Clinical Documentation

Sara Ali, Tele Geriatric Research fellowship, Okemos, MI, United States, Leena Khalafallah, Eiman Ahmed, Nadia Elbaghdadi, Hajer Ali, Azza Ahmed, Mohammed Omer, Nadir Abdelrahman

Background: AI-driven clinical documentation (CD) tools generate clinical records to reduce administrative burden and improve doctor–patient interactions but raise concerns regarding accuracy, safety, and legality.

Objectives: This scoping review aimed to map current evidence on AI-based CD tools and their implications for documentation accuracy, patient safety, data privacy, and medico-legal accountability.

Methods: Following PRISMA-ScR guidance, a systematic search of PubMed, Embase, Web of Science, IEEE Xplore, and Google Scholar was conducted from 2020-2025. Studies on AI use in CD in healthcare settings were included.

Results: 27 studies with diverse designs were included. Findings were synthesized across four domains. Large language models (LLMs) and ambient scribes were most frequently evaluated. Omission errors were the most commonly reported, comprising up to 83.8% of errors, and were strongly associated with LLMs, highlighting completeness as a key system design challenge. Hallucination rates ranged from 1.47% in optimized pipelines to 36% in general-purpose applications. Privacy concerns included third-party data sharing, limited transparency, and weak regulatory oversight. Multiple biases - algorithmic, linguistic, demographic, and automation - highlighted the complexity of AI integration. Evidence on medico-legal accountability was limited, although several studies emphasized the need for clearer legal frameworks.

Conclusion: AI-assisted documentation tools may improve workflow, but variability in error rates, particularly omissions, raises concerns about reliability and patient safety. Persistent privacy concerns, regulatory uncertainty, and limited medico-legal evidence underscore the need for human-in-the-loop validation, enforceable governance, and cautious implementation. Future research should prioritize standardized evaluations of error rates and omission risks, alongside clearer regulatory and legal frameworks.

TA05

302C

Frontier in AI and Healthcare

Invited Session

Public Health Policy

Chair: Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

Co-Chair: Adrian (Yuan) Ma, Ross School of Business, University of Michigan, Ann Arbor, United States

1 - From Trial-and-Error to Precision Mental Health: Evidence on Antidepressant Response

Adrian (Yuan) Ma, Ross School of Business, University of Michigan, Ann Arbor, MI, United States, Jun Li, Wallace Hopp

Depression is common, costly, and difficult to treat effectively. It affects an estimated 5.7% of adults worldwide, and nearly 48 million U.S. adults currently have or are being treated for the condition in 2025. Despite its scale, antidepressant treatment remains highly trial-and-error: first-line remission in clinical trials is often only 28–33%, and real-world nonadherence ranges from 46% to 83%. A key empirical challenge is that important behavioral confounders—such as medication adherence and major life events—are rarely observed in structured electronic health record (EHR) data. To address this limitation, we combine information extracted from unstructured clinical notes with longitudinal EHR data to study heterogeneous treatment effects and identify ways to shorten costly trial-and-error prescribing.

Our outcome is the Patient Health Questionnaire-9 (PHQ-9), a standard nine-item measure of depression severity. We first show that regressions using a static indicator for whether the PHQ-9 assessment occurs during an active antidepressant treatment episode yield insignificant estimates, suggesting that static specifications mask treatment-response dynamics. We then construct patient-level treatment trajectories that track antidepressant initiation, continuation, discontinuation, and switching. We use fixed-effects event-time models to evaluate how treatment responses vary across patient profiles**.** We find symptoms peak immediately before treatment initiation, worsen slightly in the first 14 treatment days, and then improve significantly thereafter. We also find heterogeneity in response, with demographics and behavioral factors such as major life changes and medication adherence shifting the magnitude and timing of treatment effects.

2 - Regulating Adaptive New Products: Can Less Oversight Lead to Better Development Practices?

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States, Jiayi Lai, Liang (Leon) Xu, Xin Fang

The literature on new product introduction pays little attention to how regulatory policies shape which products reach the market, despite strict oversight in many multibillion-dollar industries. We study this interaction for adaptive new products whose performance is hard to verify and evolves post-launch, such as AI-enabled medical devices. Motivated by ongoing regulatory reforms, we develop a two-stage game-theoretic model to compare a conventional pathway, which requires reclearance for post-launch updates, with a streamlined pathway, which allows certain planned updates to proceed without additional review. Our analysis focuses on products that are both adaptive (capable of post-launch improvement) and credence-like (whose true performance is difficult to observe). We identify a counter-intuitive "Goldilocks zone" in which waiving reclearance yields higher welfare: when the technology's improvement potential is moderate or regulatory detection capability is limited, streamlined oversight not only enables innovation effort that would otherwise be discouraged but also strengthens incentives for truthful reporting at launch by tying the value of future improvement opportunities to initial approval. By contrast, when oversight is strong and improvement potential is high, reclearance remains the preferred approach because it deters and detects noncompliance without unduly slowing beneficial updates. By jointly examining regulatory pathways and post-launch improvement incentives, the paper highlights how oversight design shapes innovation outcomes and user welfare for adaptive, credence-like products, and provides a conceptual framework for tailoring regulatory stringency to the underlying innovation and monitoring environment.

3 - When Clinicians Edits AI: Classifying and Measuring Documentation Errors in Primary Care

Ruozhu Wang, Boston University, Boston, MA, United States, Anita Carson

AI-powered clinical documentation tools that generate notes from clinician-patient conversations are rapidly adopted. Research on note quality yields mixed results, and most focus directly on the quality of AI-generated notes, without investigating the impact of clinicians' edits to AI-generated notes. This gap suggests the need for rigorous evaluation of AI-generated clinical notes that account for the full human-AI interplay.

We conduct a study on a dataset of 57 clinician–patient consultations and the corresponding medical notes transcribed by 4 out of 10 different AI tools. Each AI-generated note is reviewed independently by five physicians to flag errors, yielding 1,140 evaluated notes. This scale enables us to develop an error taxonomy through grounded theory, reaching saturation via iterative open coding with a four-person team.

We develop a taxonomy of error types identified in our dataset. This taxonomy is based on two primary dimensions: (1) whether or not the content was stated during the clinical consultation dialogue; and (2) whether or not the content is included in the AI-generated clinical documentation. These two dimensions yield four mutually exclusive and collectively exhaustive error categories: "Translation" error (stated but incorrectly recorded), "Omission" errors (stated but missing), "Fabrication" errors (not stated but included), and "Consultation Failures" (clinically relevant but never discussed). Then, we also characterize the structural and contextual patterns associated with each error type, and triangulate our findings using newer AI tools.

Our taxonomy provides actionable guidance for clinicians editing AI notes, health system managers overseeing deployment, and researchers aiming to understand tool reliability.

4 - Micro-Level Impact of Collaborating with LLMs on Productivity and Fatigue

Xufei Liu, Wharton, University of Pennsylvania, Philadelphia, PA, United States, Hummy Song, Christian Terwiesch

Large language models are deployed to reduce documentation burden, yet measured productivity gains are heterogeneous. We examine how LLM assistance impacts the worker, the type of work done, and the overall change to work productivity.

We begin with a field experiment in which licensed clinicians complete summarization and diagnosis tasks remotely. Although mean completion times are directionally lower when working with LLMs-created drafts, we

do not detect statistically significant average time savings. Open-ended responses suggest that clinicians use LLM-generated drafts as a scaffold, but they often reinvest initial time savings into revision work.

Motivated by these findings, we conduct a controlled lab study in which participants complete repeated summarization tasks both independently and with LLM assistance. We collect high-frequency interaction data, including keystrokes, click behavior, typing rates, physiological signals, and eye-tracking measures. Interestingly, while LLM assistance reduces task completion time, there are systematic changes in behavioral and physiological patterns over time. Metrics such as pupil dilation decline significantly during LLM-assisted work, consistent with increasing fatigue - an effect significantly higher when compared to writing without LLM support.

Taken together, our results show that LLMs shift documentation work from authoring toward monitoring and revision. These changes do not uniformly translate into time savings and may even introduce accumulating cognitive costs.

TA06

303

Applications of OR at a Cancer Center

Invited Session

Medical Decision Making

Chair: Iakovos Toumazis, MD Anderson Cancer Center, Houston, TX, United States

1 - Decision Analytics for Health at MD Anderson

Iakovos Toumazis, The University of Texas MD Anderson Cancer Center, Houston, TX, United States

The Institute for Data Science in Oncology (IDSO) at the University of Texas MD Anderson Cancer Center integrates advanced analytics, clinical expertise, and large-scale oncology data to accelerate evidence-based improvements in cancer care. Cancer care is increasingly complex mainly due to stochasticity in patient outcomes, resource-constrained environments, and rapidly evolving treatment modalities. There is a critical need for decision analytics rooted in optimization, simulation, and machine learning to guide decision-making. The **Decision Analytics for Health** focus area within IDSO advances quantitative methodologies that support complex decisions under uncertainty in oncology, from individualized treatment planning to population-level policy.

Industrial engineers working within IDSO collaborate closely with clinicians and data scientists to design models that address uncertainty in diagnosis and treatment, evaluate trade-offs among competing objectives, and inform local, regional, national, and international policy. Ongoing projects include predictive modeling for individualized care planning, discrete-event and agent-based simulations to optimize care and services, and multi-objective optimization frameworks that integrate clinical outcomes with operational performance metrics. These analytic tools enable stakeholders to evaluate trade-offs, anticipate system-level bottlenecks, and design resilient, effective, and efficient cancer care processes.

IDSO offers a unique environment where engineering rigor is paired with meaningful, high-impact applications. We invite industrial engineers and OR/MS researchers to learn more about the Decision Analytics for Health focus area within IDSO and contribute to the development of scalable, data-driven solutions that advance the efficiency and effectiveness of cancer care.

2 - Evaluation of Engage-Derived Lung Cancer Screening Recommendations on the National Lung Cancer Screening Trial

Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Maryam Eghbalizarch, Seyyed Mostafa Mousavi Janbeh Sarayi, Nicholson Collier, Martin Tammemagi, Rafael Meza, Jonathan Ozik, Iakovos Toumazis

The ENGAGE framework optimizes lung cancer screening for ever smokers based on their personal lung cancer risk and life-expectancy at the individual-level. We updated ENGAGE to consider additional non-smoking-related risk factors, including race, chronic obstructive pulmonary disease (COPD), personal history cancer and family history of lung cancer, education level, body mass index (BMI), in addition to sex and smoking history. This study evaluates the performance of the original and enhanced ENGAGE-derived lung cancer screening on the Natural Lung Cancer Screening Trial (NLST).

We obtained individual-level information of NLST participants . We then used both the original and enhanced versions of the ENGAGE framework to derive an optimal screening schedule for these participants. The efficiency and effectiveness of ENGAGE were measured by the number of screenings, true-positive cases, false-positive cases, true-negative cases, and false-negative cases. Additionally, we investigated the lung cancer cases detected by NLST for which their diagnosis was delayed by the ENGAGE-derived screening strategies.

The original ENGAGE recommended strategies reduced the total number of screenings in the NLST by 18% and number of false-positives by 16%, while delaying diagnosis in 3% of the lung cancer cases. Enhanced ENGAGE-derived screening reduced the total number of screenings by 20% and the number of false positives by 18%, while maintaining the number of true positives at 97% of those detected in the NLST. ENGAGE-derived schedules, especially the enhanced ENGAGE framework, maintained satisfactory level of effectiveness comparable to that of annual screening, despite significantly reducing the number of screening exams and false-positive cases.

3 - Advancing Safety, Quality, and Access to Cancer Care with Operations Research

Jeffrey Siewerdsen, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Andrew Schaefer, Eric Bickel, David Jaffray

The rising burden of cancer and the corresponding need for high-efficiency clinical operations present a major opportunity for impact through operations research (OR). To accelerate translation of OR methods into measurable clinical benefit, The University of Texas MD Anderson Cancer Center (MDACC) has launched new academic partnerships that integrate OR expertise with real-world healthcare challenges. A joint Center for Operations Research in Cancer (CORC) has been established between MDACC and Rice University, complemented by a research internship program with The University of Texas at Austin to expose students to clinical operations, workflow, real data, and decision-support needs. These initiatives unite OR researchers from Rice and UT with clinical domain experts and data scientists in the Institute for Data Science in Oncology at MDACC.

Collaborative activities include jointly funded postdoctoral fellowships, multi-institutional symposia, and competitive pilot grants to advance research with scientific rigor and clinical relevance. Current projects address operational challenges across the cancer care continuum, including: surgical scheduling optimization using improved case-duration models; workflow optimization for MR and breast imaging centers; operational planning / prioritization leveraging novel predictive models for length-of-stay after surgery and risk stratification for unplanned readmission after chemotherapy; and data-intensive capacity planning for expansion of operating rooms, inpatient units, and radiation therapy resources. Emerging areas include development of digital twin models at scales ranging from individual procedures to enterprise-level patient and data flow. This growing collaborative ecosystem at the interface of OR and healthcare demonstrates a transformative potential to improve access, efficiency, and outcomes in cancer care.

4 - Operating Room and Inpatient Capacity Planning with a Data-Driven Approach

Niloufar Mirzavand Boroujeni, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Jeffrey Siewerdsen, Kham Nguyen, Abigail Caudle, Mark Clemens, David Jaffray, Sivaramakrishnan Ramani, Puyao Ge, Andrew Schaefer, Robert Ghafar

Healthcare organizations increasingly use data-driven models to guide capacity planning for operating rooms (ORs) and inpatient beds, especially in cancer surgery where demand is rising and patient care involves numerous, multidisciplinary pathways. This work presents an analytical framework translating surgical demand into quantitative projections of perioperative capacity using institutional data and/or population-based sources.

The models account for procedure-level variations in case volumes, case duration, length of stay (LOS), operational constraints, and growth projections to estimate required ORs and beds under various scenarios and planning strategies. Validation against institutional data demonstrated close agreement, showing predictions within 6% of actual OR and bed capacity. Two use cases exercised the framework. Use Case 1 addressed planning a newly established cancer surgery center in a region without local operational data, using registry-derived cancer incidence and surgical conversion rates. Projections captured sensitivity to market share, service-line mix, and assumptions on case duration and LOS. Use Case 2 evaluated scenarios involving short-stay inpatient (SSI) surgery, which extends conventional outpatient LOS from 0–1 midnights to 0–2 midnights. Results for the SSI scenario showed a modest increase in OR demand accompanied by a higher, nonlinearly dependent increase in bed requirements, reflecting the asymmetric impact of extended LOS. The model allowed diverse stakeholders to run "what-if" scenarios and evaluate operational strategies with actionable insight for strategic planning. Ongoing work extends the planning models to incorporate temporal dynamics, stochastic variability, and operational constraints across clinical settings.

TA07

304

Applications of Analytics in Organ Transplant

Invited Session

Health Applications Society

Chair: Reza Skandari, Imperial College London, London, United Kingdom

1 - Developing An Data-Driven Adaptive Policy to Expedite Placement of Kidneys at Risk of Discard

Sanjay Mehrotra, NORTHWESTERN UNIVERSITY, EVANSTON, IL, United States, Masoud Barah, Vikram Kilambi, Richard Formica

The nonuse rate of deceased donor kidneys has increased significantly. Nota requires sequential organ allocation organs, which is inefficient for medically complex organs because it causes increased cold ischemic time and contributes to organ declines. A novel Sliding scale AdaptiVe Expedited (SAVE) Rescue algorithm for deceased donor kidney transplantation was developed that uses the nonuse rate to determine a rescue trigger threshold. Once triggered, allocation occurs in sequential batches, sized proportionately to the nonuse rate. Using a validated organ allocation simulation program, KSIM 2.0, the algorithm was evaluated using pre and post-KAS250 OPTN data.

Nonuse was reduced across all KDPI strata. Overall, the nonuse of the recovered KDPI>0.35 kidneys decreased from 38% to as low as 19%. For all kidneys the non-use rate reduced from 28% to 14%. Using lower thresholds and larger batch sizes resulted in progressive decreases in nonuse rates of KDPI>0.65 kidneys. Increasing the batch size has its largest effect at higher KDPI deciles.

The use of the SAVE Rescue Algorithm for Deceased Donor Kidney Transplantation significantly reduces the nonuse rate of deceased donor kidneys while adhering to the principle of sequential allocation.

2 - Impact of Regulation on Deceased-Donor Liver Offer Acceptance and Patient Welfare

Shubham Akshat, University of Houston, Houston, TX, United States, Liye Ma, Christopher Hughes

The public reporting of a transplant program (TP)'s post-transplant outcomes has been a contentious issue in the transplant community in the United States. TPs are also regulated by the Department of Health and Human Services, which periodically monitors their post-transplant outcomes. While they aim to uplift the underperforming TPs and enable informed decision-making (by the patients and insurance companies), there can be important unintended consequences. To maintain good public image and comply with the regulations, a TP might deviate from fulfilling its core responsibility of maximizing patient welfare when deciding to accept an organ offer for a patient. We study the impact of regulations on deceased-donor liver offer acceptance decisions at a TP. Our empirical framework integrates patient-centric and regulatory considerations, whereas existing literature has focused only on the former. We use a discrete-event simulation model to evaluate the impact of

regulations on organ non-utilization and patient welfare. We find that a TP becomes increasingly selective in its offer acceptance decisions as its regulatory standing worsens. This behavior exhibits a near-monotonic pattern for marginal quality organs. Broadly speaking, a TP deviates from a patient-optimal decision more when an organ is of marginal quality. When a TP was in poor regulatory standing, the non-utilization rate of marginal quality organs more than doubled (w.r.t. intermediate standing). More patients died on the waiting list (40 more deaths annually), disproportionately affecting high-risk patients. We underscore awareness about the distortionary effects of public reporting of a TP's outcomes, causing TPs to deviate from patient-optimal decisions.

3 - Modernizing the Development of Organ Allocation Policy

Elijah Pivo, MIT, Cambridge, MA, United States, Thomas Athey, Maura Hegarty, Nikolaos Trichakis, Dimitris Bertsimas

The U.S. organ transplant system has long contended with the challenge of fairly and effectively allocating the limited supply of transplantable organs. National organ allocation policies determine how patients are prioritized for each donated organ. However, understanding the implications of a proposed allocation policy, much less designing one capable of improving the system, is a formidable task. The waitlist consists of over 100,000 individuals spread across the country with compatible donor pools that vary drastically, each transplant involves complex and time-critical logistics, and policy changes may influence how patients and transplant centers judge transplant offers. This talk will focus on new analytical tools our group has been developing in partnership with UNOS and the OPTN, the main organizations that manage the national transplant system, in order to better support policymakers. These tools include clinical, logistical and behavioral models, a high-speed transplant system simulator, and multi-objective policy optimization. Together, these methods represent a major advance in the computational methods available for the design of organ allocation policy and an opportunity to improve the system for hundreds of thousands of current and future patients.

4 - Data to Dose: Efficient Synthetic Data Generation with Expert Guidance for Personalized Dosing

Hansraj Satyam Verma, Carnegie Mellon University, Pittsburgh, PA, United States, Holly Wiberg, Shixiang Zhu, Sridhar Tayur

Personalized treatment optimization is often constrained by data scarcity, particularly in settings involving heterogeneous patient profiles where fully labeled clinical outcomes are limited. Standard data-driven approaches struggle in these regimes, limiting their ability to provide individualized dosing recommendations. We introduce GenEx, a hybrid framework that integrates preference-based Bayesian optimization with a generative model refined through expert feedback. Expert-provided pairwise dose rankings guide updates to both a surrogate model and a generative component, which proposes synthetic data points in uncertain regions of the patient-dose space when the generator is sufficiently aligned with expert preferences. A trust gate controls synthetic injection, mitigating the accumulation of bias from generated samples. Theoretical analysis establishes sublinear regret bounds and bias-aware confidence intervals that account for the interaction between expert feedback and synthetic augmentation. Computational experiments on simulated and retrospective clinical data demonstrate that GenEx achieves faster convergence and stronger dosing utility relative to preference-only and synthetic-only baselines. These results suggest that integrating expert preferences with generative augmentation can meaningfully improve personalized decision-making in data-limited clinical contexts.

5 - Learning Sequential Donor Acceptance Policies for Heart Transplantation Using Counterfactual Reinforcement Learning

Yingtao Luo, Carnegie Mellon University, Pittsburgh, PA, United States, Reza Skandari, Arman Kilic, Rema Padman

Heart transplantation requires clinicians to make rapid, high-stakes decisions about whether to accept or decline donor organ offers. These decisions are inherently sequential and must balance uncertain future donor availability, evolving patient health status, and competing risks between waitlist mortality and post-transplant outcomes. Despite the availability of large-scale transplant registry data, existing decision-support tools largely rely on static risk prediction models that evaluate outcomes in isolation and do not capture the dynamic nature of donor acceptance decisions.

In this study, we formulate the donor acceptance problem as a finite-horizon semi-Markov decision process (SMDP) that models the irregular event-driven dynamics of transplant waitlists. Using longitudinal data from the United Network for Organ Sharing (UNOS) spanning 2018–2024, we construct a comprehensive sequential decision dataset containing over **111,000 donor–candidate offer events from 16,851 candidates**. We integrate multiple modeling components, including physician behavior policy model, time-to-censoring models, and a **doubly robust offline reinforcement learning framework** to estimate counterfactual treatment policies from observational data. The proposed framework corrects for both treatment-selection bias and informative censoring through inverse propensity weighting and inverse probability of censoring weighting, enabling unbiased estimation of long-term survival outcomes under alternative acceptance strategies.

Our results demonstrate how reinforcement learning can identify policies that better balance the trade-off between immediate transplantation and the option value of waiting for future offers. This work provides the first unified framework for optimizing heart donor acceptance policies as a sequential decision-making problem and highlights the potential to support data-driven clinical decision-making in organ transplantation.

TA08

305A

Role of Computing and Optimization in Healthcare Decision

Contributed Session

Optimization

Chair: Emily Tucker, Clemson University, Clemson, SC, United States

1 - Personalized Wellness Optimization: a Constrained Programming Approach to Cycle-Based Remedy Recommendations

Apurva Arni, UC Berkeley, Bellevue, WA, United States

Women's wellness applications have proliferated, yet remain limited to cycle prediction without actionable intervention. This work presents the first binary integer programming formulation for personalized cycle-based wellness recommendation**,** bridging this gap by selecting interventions from a multi-domain remedy database while respecting individual time budgets and clinical safety constraints.

The framework operates on a daily user state vector capturing six dimensions: cycle phase, mood, symptoms, sleep quality, energy, and stress. Each remedy in the database carries three values 1) clinical effectiveness derived from systematic review, 2) phase suitability reflecting hormonal physiology, and 3) a personalized response rate that updates with user feedback. The optimizer combines these via dynamic weights inferred from the user's current state, then selects an optimal recommendation set subject to time, conflict, diversity, and cardinality constraints.

The poster illustrates the full pipeline: state vector input, weight inference logic, the BIP formulation with all five constraint types, and the constraint relaxation hierarchy that guarantees a feasible recommendation even when constraints conflict. A worked example demonstrates how the model handles a luteal-phase user with high stress and a 60-minute time budget, producing a five-remedy recommendation across nutrition, movement, lifestyle, and herbal categories.

Implementation in Python with PuLP and the open-source CBC solver achieves sub-50-millisecond solve times on mobile hardware. The framework prioritizes interpretability, hard constraint satisfaction, feasibility guarantees and properties essential for any clinical decision support context where transparent reasoning and safety must hold.

2 - Equitable Prioritization of Liver Transplant Waitlist Policies via Stable Matching Recourse

Puyao Ge, Rice University, Houston, TX, United States, Seth Brown, Andrew Schaefer

Allocating scarce resources equitably across heterogeneous recipient groups is a central challenge in organ transplantation. In the particular case of liver transplantation, the current U.S. prioritization system produces significant disparities in expected outcomes between patients with end-stage liver disease (ESLD) and those with hepatocellular carcinoma (HCC). We develop a two-stage stochastic mixed-integer programming framework in which the first-stage decisions determine recipient group prioritization and the second-stage recourse solves a stable matching problem between recipients and resources based on the realization of uncertainty in health trajectories and arrivals. Our model captures key features of real waitlist settings: resources vary in quality, recipients may rationally decline offers in anticipation of better future options, and recipient health states evolve stochastically. We establish that the stable matching recourse polytope remains integral despite the modifications required by our two-stage setting, and we introduce fairness cuts that iteratively improve equity across recipient groups. Using a novel multi-agent simulation calibrated to U.S. liver transplant data, we show that our recommended prioritization schemes substantially reduce the outcome gap between ESLD and HCC patients while simultaneously improving organ utilization and reducing average waitlist time relative to current policy.

3 - Gnn-Enhanced Column Generation for Dynamic Volunteer Assignment and Routing in Non-Profit Operations

Hiruni Niwunhella, North Carolina State University, Raleigh, NC, United States, Tess Schaller, Leila Hajibabai, Faisal Alkaabneh

Non-profit food banks rely heavily on volunteers to operate their distribution programs, often coordinating vehicle routes to transport volunteers between their homes and service locations. Existing optimization approaches typically assume that volunteer availability is known; however, in practice, volunteers confirm participation dynamically over time. This study addresses a dynamic variant of the capacitated vehicle routing problem with pickup and delivery, time windows, and equity considerations (CVRPPD-TW-e). We develop a mixed-integer linear program (MILP) that jointly minimizes transportation costs and maximizes volunteer allocation, while enforcing equitable service across locations. To capture the dynamic nature of volunteer participation, we introduce phase-gated availability constraints that allow routing decisions to evolve as new volunteers become available over successive time periods. Due to the computational complexity of the MILP for large-scale instances, we propose a hybrid solution approach that integrates graph neural networks (GNNs) with column generation (CG). The GNN is trained to generate high-quality candidate routes, which are used to warm-start the CG process and reduce the burden of solving the pricing subproblem. By leveraging data, the GNN also captures patterns in volunteer arrivals, enabling more proactive and adaptive routing decisions. The proposed framework is evaluated against benchmark heuristic methods to assess solution quality and computational scalability. Computational experiments are conducted on both vehicle routing test instances and a real-world operational network derived from a food bank in Durham, North Carolina. The results demonstrate the effectiveness of the proposed approach in supporting scalable and data-driven coordination of volunteer-based logistics in resource-constrained, dynamic environments.

4 - Predicting Nursing Home Five-Star Rating Transitions with Hierarchical Attention Networks

Zhiguo Yang, Missouri State University, Springfield, MO, United States, Zongying Jiang, Yizhuo Zhao, Xiang Guo

Predicting changes in CMS Nursing Home Five-Star ratings can enable earlier detection of quality issues, improve regulatory oversight, and support better-informed decisions by patients and families. Although the rating is generated through a rule-based scoring methodology using health inspections, staffing, and quality measures, forecasting future rating transitions remains challenging due to dynamically evolving inputs and significant reporting delays.

This study develops machine learning models to predict Five-Star rating transitions at three horizons—3 months, 6 months, and 12 months ahead—using a seven-year monthly panel dataset from the CMS Provider Data Catalog. The models employ a hierarchical attention network to capture multi-level facility characteristics and temporal dynamics, with earlier periods used for training and later periods reserved for out-of-sample validation.

Comparative analysis with baseline models (linear regression and XGBoost) shows that the hierarchical attention network better captures temporal dependencies in nursing home performance. Initial experiments indicate that

while shorter horizons provide useful early-warning signals for tactical monitoring, the 12-month model offers the highest practical value for strategic decision-making. Because meaningful improvements in staffing, care processes, and compliance typically require 6–12 months to influence official ratings and align with annual planning cycles, one-year-ahead predictions enable proactive interventions well before the next major rating refresh.

These results demonstrate how hierarchical machine learning models can deliver timely, interpretable forecasts that support proactive quality management, more efficient regulatory resource allocation, and improved resident outcomes.

5 - Work Systems Approach to Support Innovative and Secure Quantum Health Operations Research

Emily Tucker, Clemson University, Clemson, SC, United States, Jackie Cha

Quantum computers may one day be transformative for human health research. But to get from an emerging technology to clinical impact, researchers across medicine, operations research, and biomedical sciences will need to work with researchers in quantum computing to develop the next generation of quantum-enhanced health innovations. This is not an easy bridge to cross. Establishing shared expectations and norms across health operations and quantum research communities will be essential to maintain data security and advance fundamental quantum innovations for improved health decisions. We discuss major structural barriers to success and opportunities for innovation across the health operations research lifecycle using a systematic work-systems approach. These stages include research planning and problem formulation, data access and preparation, modeling and methods development, and results interpretation and dissemination. We also highlight quantum-related problem domains of interest within healthcare operations to encourage further research from the health operations research community. This work helps develop the foundation for innovations in quantum health operations as these conventionally separate fields rapidly come together.

TA09

305B

Simulation and Applied Probability in Healthcare

Invited Session

Simulation Society

Chair: Shane Henderson, Cornell University, Ithaca, NY, United States

1 - Optimizing Peer Matching for Performance Improvement

Shanshan Luo, University of British Columbia, Vancouver, BC, Canada, Steven Shechter

Motivation plays a key role in driving performance, yet maintaining it over time remains a challenge in domains such as fitness tracking and digital learning platforms. Research suggests that individuals are more motivated when paired with slightly stronger peers, but motivation declines when the performance gap becomes too large. Building on this principle, we develop a motivation effect model to optimize peer matching based on baseline performance levels. Our goal is to maximize motivated effort and sustained performance through structured matching. Using real-world fitness data, we evaluate the optimal matching strategy against other commonly used pairing methods. Our findings provide insights into how structured peer matching can be implemented in fitness programs to improve engagement and performance consistency.

2 - Data-Driven Stochastic Modeling Using Autoregressive Sequence Models

Jing Dong, Columbia University, New York, NY, United States, Daksh Mittal, Shunri Zheng, Hongseok Namkoong

While queueing network models are powerful tools for analyzing healthcare delivery systems, they traditionally require substantial human effort and domain expertise to construct. To make this modeling approach more scalable and accessible, we propose a data-driven framework for queueing network modeling and simulation based on autoregressive sequence models trained on event-stream data. Instead of explicitly specifying arrival

processes, service mechanisms, or routing logic, our approach learns the conditional distributions of event types and event times, recasting the modeling task as a problem of sequence distribution learning. We show that Transformer-style architectures can effectively parameterize these distributions, enabling automated construction of high-fidelity simulators. Leveraging advances in artificial intelligence and the growing availability of data, our framework takes a step toward more automated, data-driven modeling pipelines to support broader adoption of queueing network models in healthcare.

3 - Optimizing Equitable Resource Allocation in Parallel Any-Scale Queues and Its Application to Liver Transplant

Sanjay Mehrotra, NORTHWESTERN UNIVERSITY, EVANSTON, IL, United States, Shukai Li

We study the problem of equitably and efficiently allocating an arriving resource to multiple queues with customer abandonment. The problem is motivated by the cadaveric liver allocation system of the United States, which includes a large number of small-scale (in terms of yearly arrival intensities) patient waitlists with the possibility of patients abandoning (due to death) until the required service is completed (matched donor liver arrives). We model each waitlist as a GI/MI/1+GI queue, in which a virtual server receives a donor liver for the patient at the top of the waitlist, and patients may abandon while waiting or during service. To evaluate the performance of each queue, we develop a finite approximation technique as an alternative to fluid or diffusion approximations, which are inaccurate unless the queue's arrival intensity is large. This finite approximation for hundreds of queues is used within an optimization model to optimally allocate donor livers to each waitlist. Computational results show that solutions obtained in this way provide greater flexibility, and improve system performance when compared to solutions from the fluid models. Importantly, we find that appropriately increasing the proportion of livers allocated to waitlists with small scales or high mortality risks improves the allocation equity. This suggests a proportionately greater allocation of organs to smaller transplant centers and/or those with more vulnerable populations in an allocation policy. While our motivation is from liver allocation, the solution approach developed in this paper is applicable in other operational contexts with similar modeling frameworks.

4 - Dispatching Community First Responders to Out-of-Hospital Cardiac Arrests

Shane Henderson, Cornell University, Ithaca, NY, United States, Pieter van den Berg, Oceane Fourmentraux, Caroline Jagtenberg, Hemeng Li, Giulia Montagna

Patient survival from out-of-hospital cardiac arrest (OHCA) can be improved by augmenting traditional ambulance response with the dispatch of community first responders (CFRs) who are alerted via an app. Dispatch of CFRs is complicated; for the best medical outcome, all nearby CFRs should be alerted, but this can lead to redundant CFR arrivals on scene, potentially decreasing CFR engagement with the CFR program in future OHCAs. We explore the use of dynamic programming, decision trees learned by bootstrapping data, and ideas related to information design to design the dispatch process. A case study for the city of Auckland, New Zealand using data from Hato Hone St John’s Ambulance Service demonstrates the ideas.

TA10

306A

Resource Allocation in Healthcare

Contributed Session

Optimization

Chair: Huijae Kim, KAIST, Yuseong-gu, Korea, Republic of

1 - Time-Varying Ambulance Relocation Under a Compliance Table Policy

Yiwen Hu, University of Nottingham Ningbo China, Ningbo, China, People's Republic of, Zujun Ma, Fei Yang, Zhao Cai

Ambulances often need to be repositioned across stations to maintain service coverage in emergency medical service (EMS) systems. A key challenge is that call demand, travel times, and ambulance availability vary over time. The compliance table policy specifies ambulance postings according to the number of available

ambulances. In this study, we address a time-varying ambulance relocation problem under this policy. To capture temporal variations in operating conditions, we adopt a data-driven approach to divide the day into multiple periods. For each period, an optimal compliance table is determined. In addition, a cross-period relocation penalty is incorporated to guide ambulance movements between adjacent periods, thereby reducing unnecessary relocations while preserving coverage performance. The objective is to maximize expected coverage over the entire day. To solve the problem efficiently at a practical scale, we develop a solution framework that combines an exact branch-and-price algorithm with heuristic procedures. Finally, a case study based on real data from the Ningbo Emergency Center is conducted to demonstrate the effectiveness and practical value of the proposed method.

2 - Developing a Heuristic Algorithm for Reallocating Outpatient Surgeries from the Main or to the Ambulatory Surgery Center

Ahmed Qasem, University of Alabama at Birmingham, Birmingham, AL, United States, Thomas Kayal, Lori Pelletier, Abdulaziz Ahmed

Operating room (OR) is among the most capital-intensive resources in health systems, yet inefficiencies persist in allocating outpatient surgery across facility types. Although ambulatory surgery centers (ASCs) are optimized for lower-acuity procedures, many eligible cases remain scheduled in higher-cost main hospital ORs, consuming prime-time capacity needed for emergent, complex, or inpatient surgeries and contributing to access delays. This study proposes a greedy first-fit scheduling heuristic to identify outpatient surgical cases eligible for transfer from the main OR and assign them to available ASC OR time slots. The approach is evaluated retrospectively using ten months of fiscal year 2025 EHR-derived surgical data from a pediatric health system with eight main ORs and four ASC ORs. Across the study period, the algorithm identified 2,071 movable outpatient cases and reassigned each without creating scheduling conflicts. Simulated reallocation increased aggregate ASC block utilization from 50.2% to 73.1% across the four ASC rooms, with room-level gains of 18.1–30.8%. These findings demonstrate that the proposed heuristic can serve as a prospective decision-support tool for optimizing surgical case allocation between the main OR and ASC. The algorithm is currently being used prospectively to support future scheduling decisions.

3 - Risk-Based Inspection Prioritization for U.S. Nursing Homes Using Machine Learning

Qirun Chen, Missouri State University, Springfield MO, MO, United States, Zhiguo Yang, Yiqing Zhang, Yuncheng Li, Xiang Guo

Regulatory agencies must prioritize inspections across approximately 15,000 Medicare- and Medicaid-certified nursing homes in the United States while operating under limited budgets and personnel. Despite mandates from the Centers for Medicare & Medicaid Services requiring unannounced surveys on a 9-to-15-month cycle, many state survey agencies struggle to meet timeliness standards due to high complaint volumes and surveyor shortages. As a result, determining which facilities should receive earlier inspection within the allowable survey window remains a critical operational challenge. Improving inspection prioritization can help regulators allocate limited survey resources more effectively, detect serious deficiencies earlier, and reduce risks to nursing home residents.

This study develops a machine learning–based risk scoring framework to identify facilities at elevated risk of serious regulatory noncompliance, including severe inspection deficiencies and Denial of Payment for New Admissions (DPNA). Using yearly records from the CMS Provider Data Catalog (2018–2024), including quality ratings, staffing levels, and inspection histories, we train a hierarchical attention network (HAN) model to capture temporal patterns of facility performance deterioration while accounting for nonlinear relationships among facility characteristics and the class imbalance associated with rare enforcement events.

The resulting risk scores enable regulators to rank facilities by enforcement risk and prioritize inspections under limited survey capacity. Preliminary results show meaningful improvements in AUC and recall over baseline models. SHAP analysis further provides interpretable insights into key drivers of enforcement risk.

4 - From Feasible to Valuable Infeasible: a Dual-Population Framework for Healthcare Districting

Huijae Kim, KAIST, Yuseong-gu, Korea, Republic of, Taesik Lee

Healthcare service area delineation is an important policy problem because it shapes how accessibility is evaluated and how resources are allocated. Existing optimization-based approaches often impose hard constraints such as minimum population and self-containment thresholds. This leaves only fully feasible solutions in the Pareto set and excludes alternatives that may offer better accessibility or spatial coherence while only slightly violating selected thresholds. As a result, it becomes difficult to understand how strict threshold satisfaction may conflict with policy-relevant performance and which constraints truly act as bottlenecks to Pareto improvement.

To address this limitation, we propose a dual-population framework that jointly explores and interprets both feasible and promising infeasible solutions. One population maintains implementable feasible Pareto solutions under current policy rules, while the other preserves valuable infeasible solutions that violate selected constraints but remain competitive in the objective space relative to a feasible reference set. To support this, we define a Pareto-relaxation dominance rule for second-population selection and treat infeasible solutions not as failures, but as policy-relevant information near the feasibility boundary.

We apply the proposed method to healthcare service area delineation in South Korea to identify binding constraints, characterize valuable infeasible alternatives, and examine whether the dual-population framework also improves feasible Pareto search over a single-population constrained genetic algorithm. More broadly, the study reframes infeasible solutions as an analytical resource for revealing policy trade-offs and the potential value of targeted constraint relaxation, thereby supporting a more flexible decision-making perspective for healthcare districting.

TA11

306B

Advancing Healthcare Delivery Through Data Science and Technology

Invited Session

Healthcare Technology and Bioinformatics

Chair: Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

1 - Dax Ambient AI Scribing: Maximizing Utility in Teaching Hospitals

Jennifer Lobo, University of Virginia, Charlottesville, VA, United States, Rahwa Ghenbot, Marieke Jones

Ambient AI scribes have demonstrated reductions in documentation time and improved clinician satisfaction across diverse settings. Benefits of same-day note closure and reduced after hours work appear strongest in documentation-intensive fields such as primary care and emergency medicine. Our study aims to fill the literature gap on how clinic size, trainee presence, or visit type influence documentation efficiency gains. We assess the impact of DAX Copilot. De-identified encounter level data were pooled from January 1 to May 31 2025. Using a negative binomial mixed effects model, we analyzed overall pre-DAX note time and DAX-utilized note time, notes time savings by provider groups (non-medical, primary care, internal medicine, pediatric subspecialties, general surgery, and surgical subspecialties/anesthesia), visit type (initial vs. follow-up visit), clinic size, and trainee presence. Additionally, data were analysed at the encounter level for time spent in each section of the chart (i.e., notes, imaging, medication reconciliation) to assess behavioral effects of DAX on charting tasks. Further, we use discrete event simulation to estimate documentation time savings at the day level with different assumptions in specialty, visit type mix, and DAX usage. We analyzed 745,734 encounters, and 16% of those used DAX. DAX reduced the predicted time spent on notes by 3.3 minutes overall. Assessing differences between visit type and specialty, primary (1.56 min), pediatric subspecialty (1.49), and general surgery (1.51) clinics saw the greatest documentation time reduction in initial visits compared to follow-up visits.

2 - Patterns of Maternal Transport Network in a State with Levels of Maternal Care and No Formal Perinatal Regions

Lauren Steimle, Georgia Institute of Technology, Atlanta, GA, United States, Jingyu Li, Margaret Carrel, Riley Byrd, Stephanie Radke

Maternal transport plays a critical role in ensuring that pregnant people receive timely, risk-appropriate care. Unlike some states, Iowa does not have formalized perinatal regions for the coordination of maternal transport. It remains unknown how facilities coordinate maternal transport in the absence of defined perinatal regions and whether these transport patterns reflect efficient, risk-appropriate coordination. Using network analysis, we constructed maternal transport network to represent maternal transport routes (network edges) among obstetric facilities (network nodes) as directional network and derive de facto "communities" that represent clusters of obstetric facilities that disproportionately transport amongst themselves. We characterized potential inefficient transport patterns, such as bypassing a closer risk-appropriate facility for a farther facility. We analyzed maternal transports as indicated on 2010-2023 birth records from the state of Iowa. Three "communities" of obstetric facilities largely shaped by geographic proximity were identified. 25.2% of transports bypassed a closer facility offering an equivalent or higher level of care than the destination facility. Health system affiliation was associated with bypassing transport, indicating potential organizational rather than purely geographic drivers of transport decisions. These results reveal patterns of maternal transport that effectively form de facto regionalized systems, even without formal perinatal regions. Our study provides actionable insights for designing perinatal care system in a state without formal perinatal regions. These findings may help identify opportunities to enhance care coordination among obstetric facilities, optimize maternal transport, and improve the efficiency of the regionalized systems of maternal care.

3 - A New Randomized Clustering Algorithm for Uncovering Multimorbidity Patterns

Hari Balasubramanian, University of Massachusetts, Amherst, Amherst, MA, United States, Ali Jafari

Multimorbidity refers the coexistence of two or more long-term conditions in an individual, and has been recognized as a global challenge. Chronic condition combinations observed in patients also exhibit a high degree of heterogeneity. We propose a new clustering algorithm -- Randomized Length Sequential Clustering Algorithm -- to systematically generate disease clusters in different age and gender demographics of the US population. Unlike off-the-shelf clustering approaches, our algorithm does not produce a single set of clusters, rather it encourages exploratory pattern discovery and is highly customized to the multimorbidity context. We demonstrate the clusters produced by the algorithm using patient-level data in the Medical Expenditure Panel Survey (MEPS).

4 - Early Treatment Instability After Psychotropic Initiation in Autism Spectrum Disorder

Paula Andrea Jaimes Buitron, The University of Texas at San Antonio, San Antonio, TX, United States, Leslie Neely, Arthur Westover, Carolina Vivas-Valencia

Individuals with autism spectrum disorder (ASD) frequently take psychotropic medications to manage co-occurring psychiatric conditions and behavioral symptoms. However, treatment response is heterogeneous, and modifications to the initial regimens (defined as the initial set of one or more concurrent psychotropic medications) are common due to limited effectiveness, adverse effects, or clinical complexity. Although prior research has documented high rates of psychotropic use and polypharmacy in ASD (the concurrent use of two or more medications), limited evidence exists on early treatment trajectories following initiation, including the timing of first regimen modification and which individuals are most likely to experience early changes. Characterizing these initial treatment dynamics is critical to understanding medication complexity and potential progression to polypharmacy. This study examines the timing of first psychotropic medication regimen change following initiation among children, adolescents, and adults with ASD and identifies patient-, treatment-, and clinical-level factors associated with earlier modification. We conducted a retrospective longitudinal cohort study using administrative claims data from individuals aged 1-84 years with at least two ASD-related claims who initiated a psychotropic medication. The index date was defined as the first psychotropic initiation following diagnosis. Individuals were followed until first regimen change, disenrollment, or the end of the observation period. Regimen changes were identified using pharmacy claims based on medication discontinuation or overlap. Time-to-event methods estimated the hazard of first treatment modification, accounting for censoring and varying follow-up. By quantifying the timing and predictors of early regimen change, this study provides population-level evidence on initial treatment instability in ASD.

TA12

306C

PhD Mentoring Session

Panel Session

Health Applications Society

Co-Chair: Yu Ma, University of Wisconsin Madison, Madison, WI, United States

1 - Panelist

Xiaoquan Gao, Singapore Management University, Singapore, Singapore

2 - Panelist

Isabelle Rao, University of Toronto, Toronto, ON, Canada

3 - Panelist

Arlen Dean, Washington University in St. Louis, St. Louis, MO, United States

4 - Moderator Panelist

Yu Ma, University of Wisconsin Madison, Madison, WI, United States

TA13

307

Implementing AI and Healthcare Technologies: Challenges, Adoption, and System Impact

Contributed Session

Healthcare Technology

Chair: Emily Garcia, Purdue University, West Lafayette, IN, United States

1 - Understanding MRI Overuse: a Physician Survey from a Tertiary Hospital

Yesim Altan, Başkent University, Ankara, Turkey, Bayram Sahin

Objective: This study aims to identify the factors that contribute to the overuse of magnetic resonance imaging (MRI) from the perspective of physicians. The study aims to reveal the factors that influence physicians to request unnecessary MRI scans.

Design: This study was designed as a cross-sectional survey. The study was conducted at Hacettepe University Adult Hospital. To determine the reasons that lead physicians to inappropriate MRI use, a structured questionnaire form consisting of two sections was used. Participants were asked 20 questions about factors thought to influence medical imaging requests using a 5-point Likert scale. The study population consisted of a total of 116 physicians working in the Emergency Medicine, Pain Medicine, Physical Medicine and Rehabilitation, Neurosurgery, and Orthopedics and Traumatology outpatient clinics at Hospital. The survey was distributed to all physicians, but only 69 responded.

Results: According to the study findings, leading factors to unnecessary MRI use include physicians' fear of being sued for medical malpractice (75.4%), the fear of missing a diagnosis (75.4%), the desire to protect oneself based on previous negative clinical experiences (71%), and not being able to devote enough time to patients due to an increased workload (68.1%).

Conclusions: This study shows that MRI overuse is influenced by factors such as defensive medical practices, diagnostic uncertainty, and workload.

Implications for Policy or Practice: Policies and decision support systems based on clinical guidelines should be developed to guide physicians in the appropriate use of imaging modalities and to assist them in making decisions regarding imaging requests.

2 - Visionconnect: Mixed Reality Road-Crossing Training for Children with Autism

Ross Niswanger, University of Houston - Clear Lake, Houston, TX, United States, Candace Kendrick, Tega Eshaturi

VisionConnect presents a mixed-reality simulation platform for therapist-guided road-crossing practice for children with autism. Real-world street-crossing training can be difficult because the skill is high stakes, environmental conditions vary, and unsafe actions can have immediate consequences. Built for visionOS on Apple Vision Pro, the platform creates a shared immersive environment in which therapist and learner remain visually co-present, supporting dyadic interaction rather than fully isolated immersion. The design is also motivated by mixed-reality use cases in which preserving mutual visual access between participants can support guided training. A central requirement of the platform is that the child should be able to enter and experience the simulation without needing to initiate the experience through GUI elements or manage scene controls. To meet this requirement, VisionConnect uses role-based entry and enforcement: the therapist initiates session start, alignment, and scene control, while the learner is auto-joined into the immersive experience with no HUD and no global scene controls. The platform combines fiducial-based coordinate alignment, floor anchoring, SharePlay-based state synchronization, collision-aware safety feedback, and observer mirroring for external supervision. Functional verification demonstrated reliable shared scene alignment, synchronized events across devices, warning activation during unsafe proximity to virtual vehicles, and stable casting to external devices. These results suggest that mixed-reality simulation can provide a practical healthcare technology framework for controlled, repeatable, and therapist-mediated safety training.

3 - Predicting Long-Term Opioid use Risk via Interpretable Machine Learning

Elisa Long, UCLA Anderson School of Management, Los Angeles, CA, United States, Fernanda Bravo, Jingyuan Hu

Opioids remain widely prescribed for chronic pain despite limited evidence of long-term benefit and serious risks, including addiction, overdose, and death. Identifying patients at high risk of sustained use is critical for early intervention. This study develops an interpretable risk-scoring model---Long-Term Opioid Use Risk (LTOUR)---to predict long-term use (>90-day supply in 180-day period) using an optimization-based machine learning algorithm. Interpretability is vital in clinical settings to build trust, support adoption, and reduce errors. Using deidentified data on 12 million prescriptions from California's prescription drug monitoring program (PDMP), LTOUR selects six clinically relevant variables: prior opioid prescription, days of supply, daily dose, opioid type, long-acting formulation, and prescriber history. LTOUR achieves high predictive accuracy (AUC=0.898) comparable to eight state-of-the-art algorithms, well-calibrated risk estimates, and robust performance across subgroups. By using only PDMP data, LTOUR could be readily integrated into existing monitoring systems, without requiring access to confidential health records.

4 - Physician Algorithm Concordance Shapes the Translation of Algorithmic Fairness and Decision Accuracy in Simulated Acute Myocardial Infarction Care

Emily Garcia, Purdue University, West Lafayette, IN, United States, Zachary J. Hass

Algorithmic risk scores used to inform clinical decision-making have been shown to exhibit fairness disparities in static evaluations, but it remains unclear whether those disparities persist once physicians follow or override the model’s recommendation. One example is the CADILLAC Risk Score, a machine learning-enabled model designed to predict one-year risk of major adverse cardiac events (MACE) and inform hospital discharge timing decisions after an acute myocardial infarction (AMI) hospitalization. To address this gap, this study examines how variable physician-algorithm concordance shapes the translation of previously observed fairness disparities in the CADILLAC Risk Score and decision accuracy in simulated AMI care. We constructed a simulated AMI cohort based on 4,364 patients from the MIMIC-IV database and modeled physician risk perceptions to evaluate concordance with CADILLAC risk estimates. Concordance was defined using probability-distance thresholds between physician and CADILLAC risk estimates. CADILLAC risk categories determined discharge timing, and outcomes included decision accuracy, fairness disparities across age, gender, race, and insurance subgroups, 30-day readmission, one-year post-discharge mortality, and hospitalization costs. Results showed that physicians who do not solely rely on the CADILLAC improved decision accuracy and reduced disparities in fairness across some demographic groups, though effects varied by probability-distance threshold and physician discrimination ability. These findings demonstrate that physician-algorithm concordance can shape how both fairness and

accuracy observed in static evaluations translate into clinical decisions, highlighting the importance of human-in-the-loop strategies when implementing risk prediction tools in time-sensitive care like AMI.

Tuesday, July 28, 2026, 9:30 AM - 10:45 AM

Tuesday Plenary

Ballroom B

Decision Intelligence for Clinical Operations: Designed for Impact

Plenary Session

Plenary

1 - Decision Intelligence for Clinical Operations: Designed for Impact

Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Achieving impact in clinical operations requires more than well-designed models. This plenary discusses how decision intelligence efforts move from problem scoping through governance to deployment, with impact as the primary objective. Using a real workload assignment tool from medical physics, the talk highlights practical lessons on designing, governing, and implementing models so they meaningfully support clinical workflows and operational decision-making.

Tuesday, July 28, 2026, 11:00 AM - 12:15 PM

TB01

301A

Dealing with the Data Deluge in Healthcare

Invited Session

AI in Healthcare

Chair: Ritu Agarwal, Johns Hopkins Carey Business School, Baltimore, MD, United States

1 - Patients use of Mhealth Wearable Devices and Their Healthcare Utilization

Indranil Bardhan, The University of Texas at Austin, Bee Cave, TX, United States, TJ Guo, Gari Clifford, Mujeeb Basit, DuWayne Willett, DuWayne Willett, Herman Taylor

This pilot study evaluates the feasibility of a novel mHealth prototype to analyze cardiovascular disease outcomes among patients by empowering them to embrace behaviors that can improve cardiovascular health through utilization of innovative digital health interventions. The research demonstrates the feasibility of integrating physiological data from wearable devices to predict patient engagement and future care utilization

2 - Stigma Reduces AI Aversion in Health Information-Seeking

Behnaz Bojd, University of California Irvine, Irvine, CA, United States, Aravinda Garimella, Haonan Yin

Access to health information is critical for navigating challenging circumstances, and information-seeking is often the first step toward pursuing diagnosis, treatment, and professional help. However, stigma presents a significant barrier to information-seeking. Recent advances in generative artificial intelligence have increased the popularity of AI chatbots as conversational information sources. This study examines how stigma shapes individuals' preferences for information-seeking from AI chatbots compared to human experts. Through a series of online randomized controlled experiments and a field experiment, we find consistent evidence that stigma increases the preference for AI chatbots relative to human experts, effectively reducing algorithm aversion. We identify two mechanisms underlying this effect: judgment concern and misinformation concern. Under high stigma, individuals are more concerned about social judgment, making chatbots a more appealing, non-

judgmental alternative. Simultaneously, stigma reduces concern about misinformation, further lowering resistance to AI chatbots. This study extends the literature on algorithm aversion by introducing stigma as a task-specific determinant of AI preference for information-seeking. It demonstrates a robust methodological approach that enables the causal identification of the effect of stigma on willingness to seek health-related information, capturing both stated and revealed preferences. For organizations seeking to improve access to critical information in stigmatized domains, the findings indicate that AI chatbots can expand information access but also lead to reduced user vigilance against misinformation. By judiciously integrating AI chatbots within broader support systems, organizations can utilize their benefits to improve access to information while mitigating risks, enabling responsible adoption of AI in sensitive domains.

3 - From Cgm Data Streams to Actionable Insights: An Agentic AI System for Event-Driven Diabetes Management

Junjie Luo , John Hopkins University, Baltimore, MD, United States, Ritu Agarwal, Gordon Gao

Continuous glucose monitors (CGMs) generate dense, high-frequency physiological data streams, yet the gap between raw glucose readings and actionable clinical guidance remains wide. Bridging this gap requires not only accurate prediction, but also contextualization, interpretation, and timely communication.
We present an agentic AI system that transforms CGM data streams into personalized, event-contextualized health insights. The system operates in two phases. First, we collect paired event-CGM data from patients via Samsung Galaxy Watch and the WellDoc platform, capturing key life events such as meals, exercise, medication, and stress, and train an event-driven prediction model conditioned on these events. Second, an LLM-based agent integrates model predictions with contextual reasoning to generate and deliver personalized messages to patients and care providers, including forecasts, explanations, and behavioral recommendations. Message quality is evaluated using both an LLM-as-a-judge framework and clinical expert assessment.
This work presents a deployable, end-to-end agentic AI pipeline that converts CGM data into interpretable, event-driven clinical communication. We discuss system design, ongoing deployment, and implications for AI-augmented chronic disease management.

TB02

301B

Advanced Modeling and Analytics in Healthcare

Invited Session

Applied Probability

Chair: Xiaoquan Gao, Singapore Management University, Singapore, Singapore

1 - Adaptive Dose Optimization Across Multiple Populations in Basket Trials

Yuwei Zhou, Indiana University, Bloomington, IN, United States, Linyun He, Ozge Yapar

We study dose optimization in basket trials, where multiple patient subpopulations are evaluated simultaneously within a single study. Our research is motivated by the rise of tumor-agnostic therapies, which target molecular alterations irrespective of tumor site and are typically evaluated in basket trials spanning subpopulations defined by cancer type, biomarkers, or histology. For these targeted agents, dose-efficacy relationships are often non-monotonic, rendering conventional dose-finding designs inadequate. The FDA guidance emphasizing dose optimization beyond the maximum tolerated dose paradigm has heightened interest in identifying doses that balance efficacy and safety within each subpopulation. Recent literature typically relies on Bayesian hierarchical borrowing and utility-based efficacy-safety tradeoffs, which may perform poorly under heterogeneous dose-response patterns. Frequentist methods with explicit safety guarantees and minimal structural assumptions remain limited.

We propose a frequentist adaptive design to identify, for each subpopulation, the most effective dose subject to an explicit safety constraint. Our design leverages two commonly observed structural properties: unimodal dose-efficacy curves and non-decreasing dose-toxicity curves. Under these assumptions, dose selection reduces to a joint boundary-finding problem: locating the efficacy peak and the toxicity-threshold crossover. Using a large

deviations framework, we show that “critical” dose levels near these boundaries dominate the probability of false selection (PFS) asymptotically. This insight yields an allocation rule that concentrates sampling on the most informative dose-subpopulation pairs, balancing decay rates of PFS under a finite budget. Preliminary simulations demonstrate substantial improvements in probability of correct selection and simple regret relative to uniform allocation.

2 - Algorithms for Substance use Disorder Treatment Selection

Eric Pulick, University of Wisconsin - Madison, Madison, WI, United States, Yonatan Mintz

This work studies the sequential decision-making problem of personalized treatment selection for substance use disorders via a digital therapeutics treatment platform. Our modeling centers on a time-varying construct that we refer to as the individual's capacity for engagement with treatment; we model each individual as possessing uniquely parameterized capacity dynamics and propose a model for how capacity determines the individual's adherence behavior. We first provide a theoretical result for parameter learning within this framework. We next propose a model based reinforcement learning algorithm based on optimism (i.e., upper confidence bounds) and prove sublinear regret of this algorithm for the treatment selection problem. To support these results, we provide computational experiments that compare the performance of our proposed algorithm to a collection of benchmark methods on synthetic patient cohorts.

3 - Optimal Push, Pull, and Failure Funding for Global Health

Chenxi Xu, Duke University, Durham, NC, United States, David Ridley, Peng Sun

Malaria and tuberculosis each cause over half a million deaths annually, yet commercial incentives to develop treatments for these and other diseases concentrated in low-income countries remain weak. Governments and nonprofits address this gap through push (e.g., grants) and pull (e.g., prizes) mechanisms. We propose a third approach: the funder pays only if the firm fails, reimbursing part of its testing costs. This failure insurance is optimal when markets are large enough to reward success but too small to justify initial investment. We model the problem as an infinite-dimensional optimization problem with adverse selection and moral hazard constraints, and use duality theory to characterize optimal funding mechanisms. Failure insurance is preferred for tuberculosis if testing costs are below \$1 billion. For most tropical diseases, including malaria, the optimal policy is pull funding with supplemental push support. These results challenge current push-heavy practice and offer broader insights for global health and innovation policy.

4 - Dynamic Routing Under Delayed Individual Information and Long-Term Impact

Xiaoquan Gao, Singapore Management University, Singapore, Singapore, Pengyi Shi

In many service systems, routing decisions must be made before detailed individual-level information is available. Delays in clinical evaluation or risk assessment create a tradeoff between committing early based on imperfect information and adjusting later at a cost, especially when routing decisions have persistent downstream congestion effects. We develop a dynamic routing framework that captures delayed information and long-term impact. Without information delay, the optimal policy admits an exact two-step structure: advance planning based on aggregate system information followed by priority-based individual assignment. Motivated by this benchmark, we approximate the delayed-information problem using a fluid model, yielding a tractable, structure-preserving two-step policy. Our analysis characterizes when adjustment is necessary and how large it should be. We identify regimes in which the value of individual-level information is small relative to adjustment costs and congestion persistence, under which adjustment is unnecessary and the two-step policy is asymptotically optimal. A jail diversion case study calibrated with ten years of data supports our theoretical insights and illustrates when flexibility can destabilize system performance.

TB03

302A

Nontraditional Hospital and Pharmacy Operations

Invited Session

Healthcare Operations Management

Chair: Maria Ibanez, Kellogg School of Management at Northwestern University, Evanston, IL, United States

1 - Holes in the "Donut of Truth": the Impacts of Technology Reliance on Emergency Department Resilience

Mitchell Tang, Columbia University, New York, NY, United States

As use of large language models (LLMs) has grown, so too have concerns of overreliance and deskilling. Now that professionals routinely use LLMs to help draft reports, they may see their critical thinking and writing skills deteriorate. In education, these concerns are starker. Middle school students, with LLMs readily at their fingertips during these formative years, may never build these foundational skills in the first place.

We won't be able to understand the ramifications of this trend for many years to come. However, while LLMs are new, the general phenomenon of technology reliance is not. "Technological crutches" that substitute for historically human skills can be found throughout our lives, from calculators that obviate the need for hand calculation to back-up cameras that simplify parking.

One notable example in healthcare is use of imaging in medical diagnosis. Younger clinicians fondly (and somewhat mockingly) refer to computed tomography (CT) scanners as "donuts of truth", while senior clinicians criticize what they perceive to be a growing overreliance on the technology in place of physical exams and diagnostic judgement. Just as with LLMs, there are concerns that clinicians trained today are not asked to develop these fundamental clinical skills, and that this may have important downstream consequences on the quality of general medical decision making and system fragility in times when imaging is less available. This paper aims to investigate the implications of the growing reliance of imaging on emergency departments and the broader phenomenon of skill development in the midst of "technological crutches".

2 - Dynamic Production Sequencing in Compounding Pharmacies

Kraig Delana, IE Business School, Madrid, Spain, Christopher Chen, Xiaoshan Peng

When patients require personalized medications or medications that are otherwise unavailable, e.g., due to patients' allergies or drug shortages, retail pharmacies must rely on compounding pharmacies that produce medications to order. Inspired by discussions with the management team of a compounding pharmacy, this project aims to improve timely access to compounded medications through the development of a dynamic production control policy designed to minimize the time between the prescription request and when the medication is ready. We propose and evaluate a novel and theory-driven heuristic policy based on a fluid model of production dynamics paired with a State-Dependent Riccati Equation controller, lightly modified for use as a value function approximation technique, which serves as the basis of an easily implementable index policy. Numerical studies suggest that for large systems this heuristic is within 11% of a lower bound on waiting cost per unit time within a representative fluid setting demonstrating strong theoretical performance. Within a data-driven simulation with more than 1000 unique medications where arrival rates are dynamic and must be estimated, a hybrid policy combining an age-limit with the proposed heuristic can reduce the average cost of waiting time by 32.5% compared to the status-quo oldest request sequencing policy without increasing worst case delays. The proposed heuristic thus contributes a novel and easily implementable approach to production sequencing within compounding pharmacies and a generalizable approach to challenging multi-product production management problems where optimal policies are intractable.

3 - Developing a System Dynamics Model for the Implementation of Hospital-Based Gene Therapy Production

Bora Zwart, Erasmus School of Health Policy & Management, Erasmus University Rotterdam, Rotterdam, Netherlands, Maura Leusder, Sandra Sülz, Maarten IJzerman

Traditional pharmaceutical production is often too slow, too costly, and too rigid to meet the urgent and highly individualized needs for patients with rare diseases. This mismatch highlights the need for alternative production models that can deliver personalized therapies in a timely and operationally feasible manner. Recent advances in RNA manufacturing and formulation technologies enable small-scale, point-of-care production of RNA-based therapies within hospital pharmacies, extending the traditional concept of compounding toward more advanced

nanomedicines. While hospital-based production offers the potential to improve responsiveness and access, its implementation introduces significant operational complexity.

To understand how such an approach could be integrated into existing hospital pharmacy operations, we develop a system dynamics model that integrates patient flow from prescription to administration with important pharmacy processes. The model captures critical feedback loops, capacity constraints, and process delays that affect system performance and is used to examine the impact on timeliness, capacity, and patient access.

By explicitly capturing nonlinear dynamics and feedback effects, the model supports identification of operational bottlenecks and leverage points for system design and capacity planning. The study demonstrates how system dynamics modeling can inform decision-making in nontraditional hospital and pharmacy operations involving personalized gene therapy production.

4 - When Changing the Cost of Data Creation Undermines Its Value

Celia Stafford, Harvard Business School, Boston, MA, United States, Rebecca Karp

Healthcare organizations increasingly intervene to reduce documentation burden on providers and increase data quality, both in the form of increased accuracy and completeness. But optimizing production does not optimize the system. This study examines how changing the cost of *writing* clinical notes raises the cost of *reading* them, and how clinicians respond by restructuring the cognitive work of care.

Drawing on 50 interviews and 350 hours of field observations conducted before and after the implementation of an ambient AI based note writing tool in a large health system, we document a systematic shift in how clinicians exercise discretion over information-gathering. As AI-generated notes grow longer and less structured, the interpretive cost of utilizing narrative documentation to make decisions rises. Clinicians respond by reallocating effort away from reading notes and toward structured quantitative data sources such as labs and imaging. The medical record does not disappear; it gets bypassed.

This reallocation is self-reinforcing. As reading declines, the perceived return to writing carefully also falls, further degrading note quality and accelerating the shift toward structured data. The result is a coordinated but unplanned reorganization of how clinical information flows through the system.

The findings contribute a behavioral operations mechanism: when technology changes the cost of one task in a workflow, workers exercise discretion to minimize cognitive load, shifting bottlenecks rather than eliminating them. For healthcare operations, this suggests that documentation technologies alter not only efficiency, but *which data clinicians trust and act on,* with direct implications for care coordination and clinical decision-making.

TB04

302B

Data-Driven Decision Making for Healthcare Operations

Invited Session

Healthcare Operations Management

Chair: Joel Goh, NUS Business School, Singapore, Singapore

Co-Chair: Yvonne Huijun Zhu, Pennsylvania State University, University Park, PA, United States

1 - The Power of One Change: Bayesian Sequential Experimentation Under Limited Flexibility

Yvonne Huijun Zhu, Smeal College of Business, Pennsylvania State University, University Park, PA, United States, Yifan Feng, Tu Ni

Sequential experimentation can improve learning efficiency by adapting data collection to accumulated evidence. In many operational settings, however, frequent reconfiguration is impractical due to logistical,

regulatory, or implementation costs. We study the design of Bayesian sequential experiments under limited flexibility, where a decision maker solves a binary hypothesis testing problem by sampling from a collection of available experiments but may change experiments at most N times. Performance is measured by expected sampling cost plus an error penalty C for an incorrect final decision. We introduce a simple policy, *Nested Sequential Probability Ratio Test* (Nested SPRT), which organizes experiment changes through nested belief thresholds. We show that allowing even a single change dramatically improves efficiency: the optimality gap relative to the fully flexible benchmark drops from Θ(logC) with no flexibility to O(loglogC). More generally, we establish that each extra change contributes one further layer of iterated-logarithmic improvement, which reveals sharply diminishing returns to flexibility. A single well-timed change can capture most of the informational benefits of full adaptivity while remaining operationally simple. These results provide practical insights into designing near-static sequential experiments in contexts where adjustment costs are nontrivial.

2 - Crisis at the Core: Examining the Ripple Effects of Critical Incidents on Emergency Department Physician Work Performance and Work Style

Lesley Meng, Yale School of Management, Yale University, New Haven, CT, United States, Huifeng Su, Rohit Sangal, Edieal Pinker

We examine how disruptive critical incidents (CIs)—acute, high-intensity patient arrivals requiring resuscitation—affect emergency department (ED) physician productivity and decision-making quality. Prior research on workplace disruptions has focused on settings with sequential task management and limited decision complexity, finding that such events diminish worker productivity. EDs, however, are characterized by parallel task management, high decision complexity, and strong provider discretion, making it unclear whether prior findings generalize. Using four years of electronic medical record and administrative data from a large U.S. hospital, we employ quasi-experimental regression approaches to estimate within-shift impacts of CIs on physician work behavior. Contrary to prior work, we find no significant decline in shift-level productivity following CI exposure. Physicians maintain throughput by managing more patients concurrently, extending working hours, increasing reliance on diagnostic testing, and discharging more patients. These compensatory strategies sustain patient flow but introduce decision-quality trade-offs. While most post-CI discharges show no elevated revisit rates, patients with conditions for which diagnostic testing provides limited diagnostic value experience significantly higher 72-hour ED revisit rates. Our findings suggest that in complex service environments with high worker ownership and discretion, productivity losses from disruptions may be masked by adaptive behaviors that shift costs onto decision quality. System-level interventions are needed to support physicians after CI exposure—specifically, shielding them from cases demanding high cognitive effort when diagnostic testing offers limited decision support.

3 - Context Switching and Relational Human Capital: Evidence from Kidney Transplantation

Yiwen Jin, Haskayne School of Business, University of Calgary, Calgary, AB, Canada, Jiayi Liu

When does switching between organizational contexts degrade individual performance? We study this question in kidney transplant surgery, where the answer is a matter of life and death. Using national registry data on over 200,000 transplants in 13 years, we exploit within-surgeon–center variation to estimate the effect of context switching between centers on patient outcomes. We find that switching to a different center than one's most recent case increases 1-year graft failure or death rate by 0.49 percentage points. This penalty is concentrated in same-day switches and vanishes with even a one-day gap. We further show that such context switching cost is moderated by some other organizational moderators.

4 - Structural Estimation of Kidney Transplant Candidates Quality of Life Scores: Improving National Kidney Allocation Policy Under Endogenous Patient Choice and Geographical Sharing

Yuwei Zhou, Indiana University, Bloomington, IN, United States, Baris Ata, John Friedewald, Yue Hu, Ali Randa

We develop a data-driven framework for assessing the impact of changes to the deceased donor kidney allocation policy taking into account the transplant candidates' (endogenous) organ acceptance behavior and sharing among different locations under UNOS' geographically tiered allocation structure. To be specific, we develop a dynamic structural model of the transplant candidates' accept/reject decisions for organ offers. Using United Network of Organ Sharing (UNOS) data on transplant candidates, donors, organ offers, and follow up

results on transplant outcomes, we estimate the transplant candidates' quality of life scores on dialysis versus post transplant. Our estimates yield similar results on average to what is typically assumed in the medical literature. However, they differ significantly when patient and donor characteristics are considered. We then consider policy proposals that have been considered by UNOS. This counterfactual analysis helps assess the (unintended) consequences of such policies.

TB05

302C

Smarter Healthcare Decisions: Data-Driven Analytics for Policy and Practice

Invited Session

Public Health Policy

Chair: Xiao Zang, University of Minnesota, Minneapolis, MN, United States

1 - Decision Support in a Changing Drug Landscape: Analytics for Substance Use Policy

Mohammad Jalali, Center for Health Technology Assessment, Harvard Medical School, Boston, MA, United States, Huiru Dong, Eshaan Sheth, Erin Stringfellow

Decision support for substance use policy has become increasingly challenging as the drug landscape, treatment practices, and patient populations evolve more rapidly than traditional evidence can keep pace. In this talk, we present how data-driven analytics can help inform healthcare policy and clinical decision-making in the context of the overdose crisis. Using examples from national simulation modeling and population-level analyses, we illustrate how diverse data sources can be integrated to evaluate intervention strategies, estimate long-term population impacts, and identify evidence gaps that are difficult to address through conventional studies alone. We then discuss how these approaches are evolving to incorporate richer longitudinal data and modern causal inference methods to evaluate emerging treatment strategies in the fentanyl era. These examples highlight how simulation, real-world data, and advanced analytics can generate evidence that supports more informed and adaptive healthcare decisions in rapidly changing public health environments.

2 - Monitoring Cannabis use Disorder Symptoms over TIME Through Reddit Discussions

Jiyang He, University of Southern California, Los Angeles, CA, United States, Shinyi Wu, Carolyn F. Wong, Bridgid Conn, Kaia Sheean, Shayra Robles, Isaac Bernstein

Background: Traditional longitudinal surveys of cannabis use provide high-quality evidence but are costly and slow to update. In contrast, online communities may capture more timely behavioral signals at scale. This study examines how discussions on Reddit reflect temporal changes in cannabis use disorder (CUD) related symptoms and impairments, and whether these patterns are consistent with previously reported trends.

Methods: We analyzed posts from the largest recreational cannabis use community on Reddit between October 2009 and December 2025. Using a structured annotation framework based on DSM-5 CUD criteria and functional impairments, we estimated the yearly proportion of posts reporting each symptom. We further examined state-linked cannabis communities (California, Florida, Colorado, Oregon, and Pennsylvania) to explore geographic variation. In addition, we compared temporal patterns observed in Reddit with those from a 10-year longitudinal study of young adults in Los Angeles.

Results: Descriptive analyses characterize changes in the prevalence and composition of CUD-related symptoms over time in online discussions, as well as variation across state-linked communities. Comparisons with longitudinal cohort data provide an initial assessment of consistency in temporal patterns across data sources.

Conclusions: Social media data may offer a complementary approach for tracking behavioral trends related to cannabis use. These analyses provide an initial step toward understanding how online discussions can inform population-level monitoring and support future work on early detection.

3 - A Microsimulation Modeling Platform to Inform Strategies for Hcv Micro-Elimination in Correctional Settings

Zongbo Li, Stanford University, Stanford, CA, United States, Ziping Ye, Rebecca Bilden, Nate Furukawa, Amanda Honeycutt, Kimberley Lucas, Donna Kalauokalani, Amy Krawiec, Teresa Puente, Joshua Salomon, Marissa Reitsma

Correctional settings are essential venues for Hepatitis C virus (HCV) elimination efforts. To illustrate the potential for modeling to inform micro-elimination strategies in prisons, we developed a microsimulation model of HCV transmission dynamics, testing, and treatment applicable to correctional populations. The model simulates weekly population flow (admissions and releases), HCV transmission, testing (for antibody and RNA), and direct-acting antiviral (DAA) treatment initiation after entry or subsequently during incarceration. Model parameterization and calibration was informed by electronic health record data covering all persons incarcerated in the California state prison system. As an illustrative application, we quantified the contribution of universal opt-out testing and treatment on entry to declines in HCV RNA-positive prevalence and number of new infections. Our data-driven modeling platform can inform HCV elimination strategies in correctional settings through retrospective counterfactual and prospective scenario analyses. The platform can be extended to evaluate long-term health outcomes, cost-effectiveness, and budget impacts of micro-elimination strategies, including integrated substance use disorder treatment and harm reduction programs, to address key questions identified by correctional health system partners.

4 - Optimizing Geographic Naloxone Distribution: a Simulation-Based, Metamodel-Assisted Approach

Heejin Choi, University of Minnesota, Minneapolis, MN, United States, Ryan Kuntz, Zongbo Li, Quishi Chen, Xiao Zang

Naloxone prevents opioid overdose deaths (OODs), yet decision-makers lack practical tools to guide geographic allocation of limited naloxone to maximize impact. We propose a simulation-based framework to optimize geographic naloxone allocation.

We developed a framework that combines an existing simulation model (PROFOUND) with a metamodel to optimize naloxone distribution across geographic subregions to minimize OODs under a fixed annual number of naloxone kits. The PROFOUND model is an individual-based microsimulation model used to evaluate naloxone distribution strategies. In the framework, we generated numerous naloxone allocation schemes across subregions and projected OODs using the simulation model. We trained a metamodel using simulation results to approximate the relationship between naloxone allocation and projected OODs, using an artificial neural network. Optimization algorithms were applied under a fixed total number of naloxone kits to identify allocations minimizing OODs, and optimized solutions were validated using the original simulation model. We applied this framework in Rhode Island under an annual constraint of 50,000 naloxone kits.

The optimized allocation reduced projected OODs from 358.5 under the status quo distribution to 340.9 in 2026 under the same total number of naloxone kits. The optimized strategy resulted in a different geographic distribution of naloxone, with Providence remaining the largest recipient but received fewer kits, while smaller and more underserved communities received increased allocations.

The proposed framework can help reduce OODs by approximately 5% without increasing the number of naloxone kits distributed. This finding highlights the potential of evidence-based optimization to improve the public health impact of existing resources.

5 - Bayesian Calibration for Multi-Setting Hiv Modeling of 49 Ending the Hiv Epidemic Priority Counties

Azzam Al-Rahbi, University of Minnesota, Minneapolis, MN, United States, Tri Van, Xiao Zang, Bohdan Nosyk, Qiyou Wu, Benjamin Enns, Elvis Cai

Background: Heterogeneity in HIV microepidemics across US cities necessitates locally-oriented, combination implementation strategies to prioritize resources. We calibrated a dynamic, compartmental HIV transmission model to establish city-level status quo projections through 2045, holding service levels constant while

accounting for population growth and demographic shifts across 32 US cities representing all 48 ‘Ending the HIV Epidemic’ (EHE) counties.

Methods: We adapted and extended a previously published model to replicate HIV transmission, disease progression, and clinical care in each city. The updated model incorporated structural modifications, including streamlined health states and age stratification. Seventeen epidemiological calibration targets from 2017–2022 were used. Bayesian calibration with Metropolis–Hastings Gibbs sampling was applied to estimate posterior parameter distributions and generate best-fitting projections. We projected HIV incidence rates and incidence rate ratios by race/ethnicity from 2025–2045.

Results: Calibrated models showed good agreement with observed targets, with most observations falling within 95% posterior predictive intervals. External validation indicated that most projected incidence estimates were consistent with published new diagnoses. Under constant 2022 service levels, projected HIV incidence rates generally declined or remained stable across most cities from 2025–2045, though substantial heterogeneity persisted across jurisdictions and racial/ethnic groups. Black/African American and Hispanic/Latinx populations consistently experienced higher incidence rates than White/Other populations.

Discussion: Contemporary HIV epidemic trends across US cities are more heterogeneous than previously projected, and racial/ethnic disparities are likely to persist under current service levels. City-specific status quo projections can support future evaluations of targeted HIV prevention and treatment strategies.

TB06

303

Healthcare Decision-Making with Modeling and Simulation

Invited Session

Medical Decision Making

Chair: Yifang Yan, The Pennsylvania State University, University Park, PA, United States

Co-Chair: Qiushi Chen, The Pennsylvania State University, University Park, PA, United States

1 - Determining Performance Thresholds for Biomarker-Guided Immunotherapy Selection in Metastatic Melanoma

Jacob Jameson, Harvard University, Cambridge, MA, United States, Soroush Saghafian, Giuseppe Tarantino, Alexander Gusev, David Liu

Artificial intelligence has accelerated biomarker discovery in oncology, yet newly discovered biomarkers rarely enter clinical practice. A central reason is that discovery focuses on predictive performance metrics rather than clinical and economic value. We present a framework that maps the full sensitivity–specificity trade-off space directly onto survival, toxicity, and quality-adjusted outcomes.

We illustrate the framework in first-line immunotherapy for metastatic melanoma, where patients face a binary escalation decision: single-agent anti-PD-1 (nivolumab) or combination anti-PD-1 plus anti-CTLA-4 (nivolumab + ipilimumab). We develop a patient-level microsimulation calibrated to 10-year outcomes from CheckMate 067, assigning latent response types and estimating counterfactual survival and toxicity under a potential outcomes framework.

A perfect biomarker achieves 5-year overall survival 2.5 percentage points above universal monotherapy while reducing adverse events by 46% relative to universal combination — establishing an upper bound on achievable utility. Sensitivity and specificity operate with largely independent effects: high specificity reduces needless toxicity in likely responders; high sensitivity ensures resistant patients receive combination therapy before progression. Applying the framework to two published genomic biomarkers reveals their distinct positions in this trade-off space. A genomic heterogeneity and ploidy model (AUC 0.75) prioritizes toxicity reduction; a TMB plus interferon-gamma model (AUC 0.79) prioritizes survival. Both outperform universal monotherapy while substantially reducing toxicity relative to universal combination.

Biomarker adoption fails not because models lack predictive power, but because AUC cannot answer the questions oncologists, patients, and payers actually face. This framework answers those questions directly and is transportable to any treatment-selection setting with asymmetric consequences.

2 - Physician Practice Variation and Operational Instability in the Emergency Department

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, David Rea, Margret Bjarnadottir, David Anderson

We study how attending physician heterogeneity in the emergency department affects disposition decisions, triage-to-disposition time (TTD), and downstream throughput stability. Using ED encounter data linked to attending physicians, we estimate machine learning models for disposition choice and for admission and discharge TTD. A calibrated simulation framework then models these operational constraints dynamically across several scenarios: actual, baseline, neutralized (physician effects removed), stressed (concentrating top-effect physicians). The estimates reveal substantial physician-level variation in both admission propensity and TTD. Provider profiling further suggests more years in practice are associated with shorter TTD for both admissions and discharges. In counterfactual analysis, concentrating high-effect physicians materially increases operational volatility rather than simply shifting mean throughput. Relative to the neutralized scenario, the stressed scenario increases the daily admission interquartile-range ratio by 1.44 and the high-admit-day rate by 31.4 percentage points, while admit-side TTD tail risk rises from 3.4% to 13.0%. These findings show that physician heterogeneity can act as an operational shock, implying that variance-aware scheduling is important for maintaining throughput stability.

3 - Accelerating Simulation Model Calibration with Bayesian Optimization

Yifang Yan, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen

Surrogate models are often used to approximate complex but computationally expensive simulations, offering significant advantages in various application settings. Bayesian optimization utilizes surrogate models, such as Gaussian processes (GP), to efficiently search for optimal solution and performance of the simulation model. In this study, we evaluated whether structured surrogate models could improve calibration efficiency for disease simulation models with multiple calibration targets. Within Bayesian optimization, we applied input warping to address non-stationarity and considered three GP-based surrogate strategies: a standard GP baseline, a higher-order GP (HOGP) model for the full structured output, and a multi-latent HOGP model for separate output families. We compared Bayesian optimization with differential evolution and simulated annealing. Performance was evaluated by mean squared percentage error (MSPE) between simulated and target outputs. We conducted computational experiments in a case study of calibrating a simulation model for colorectal cancer with 9 calibration parameters and 36 synthetic calibration targets. Compared with differential evolution and simulated annealing, Bayesian optimization required substantially fewer simulator evaluations to achieve acceptable calibration error. Within Bayesian optimization, structured surrogate models outperformed the standard GP baseline, improving both calibration accuracy and search efficiency. Relative to the standard GP baseline, higher-order GP surrogates improved both early-budget and end-of-run calibration accuracy. These results suggest that leveraging output structure can make calibration both more accurate and more efficient, highlighting structured surrogate models as a promising approach for calibrating disease simulation models with multi-output targets.

4 - Machine Learning Based Analysis of Breakthrough Pneumococcal Disease: Population-Level Evidence for Vaccine Policy in Australia

Melike Yildirim, Wayne State University, Detroit, MI, United States, Shayan Sharifi, Stephen Pelton, Inci Yildirim, Teena Chopra

Background:
Pneumococcal conjugate (PCV) and polysaccharide (PPSV) vaccines have reduced invasive pneumococcal disease (IPD) in Australia. We evaluated breakthrough IPD from 2009 to 2023 following PCV13 and PPSV23 immunization to inform strategies for next-generation vaccines.

Methods:
We analyzed laboratory-confirmed IPD cases from the National Notifiable Diseases Surveillance System to estimate the proportion of breakthrough IPD, defined as vaccine-type infection despite receipt of the minimum age-specific dose threshold. Temporal trends were assessed using logistic regression, and subgroup differences using chi-square or Fisher's tests. Machine-learning models (Random Forest, Gradient Boosting) with five-fold cross-validation identified predictors of breakthrough IPD.

Results:
A total of 25,643 IPD cases were reported between 2009 and 2023, during which annual incidence declined by approximately 40%. Among vaccinated individuals meeting dose thresholds, 29% of IPD cases among PCV13 recipients and 48% among PPSV23 recipients were vaccine-type breakthrough IPD, with serotype 3 accounting for the largest share. Across age groups, vaccinated individuals consistently showed lower proportions of vaccine-serotype IPD compared with non- or partially vaccinated cases. Breakthrough proportions were highest for pneumonia and in the Northern Territory. Older adults and Indigenous Australians experienced the greatest residual disease burden. Machine-learning models (accuracy ≈92.5%) identified age, clinical presentation, and vaccination-related variables as key predictors.

Conclusion:
Breakthrough IPD is most common among older adults and Indigenous populations. Serotype 3 contributes disproportionately to vaccine-type disease. Reducing the remaining burden will likely require improved coverage in high-risk populations and expanded use of higher-valent conjugate vaccines (e.g., PCV15/PCV20).

TB07

304

Opioid Use Disorder: Data-Driven Approaches to Prevention, Treatment, and Policy

Invited Session

Health Applications Society

Chair: Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States

1 - Evaluating Diversion and Treatment Policies for Opioid use Disorder

Veronica White, FAMU-FSU College of Engineering, Florida A&M University, Tallahassee, FL, United States, Laura Albert

The United States (US) opioid crisis contributed to 81,806 fatalities in 2022. It has strained hospitals, treatment facilities, and law enforcement agencies due to the enormous resources and procedures needed to respond to the crisis. As a result, many individuals who use opioids never receive or finish the treatment they need and instead have many interactions with hospitals or the criminal justice system. This paper introduces a discrete event simulation model that evaluates three opioid use disorder treatment policies: arrest diversion, reentry case management, and overdose diversion. Publicly available data from 2011 to 2019 in Dane County, Wisconsin, was used to forecast opioid-related outcomes through 2032. Through analyzing a variety of policy-mix implementations, the study offers a versatile framework for evaluating policies at various implementation levels. The results demonstrate that treatment policies that create new pathways and programming by utilizing treatment services and successfully divert at least 20% of eligible individuals can lead to more opioid-resilient communities. The benefits increase when more policies are enacted and/or offered to more individuals, with the largest impact from overdose diversion, followed by reentry case management, and the smallest impact from arrest diversion. The statistically significant 10-year cumulative total reduction in societal costs from 2023 through 2032 ranges from $39 M (USD) to $584 M (USD), excluding implementation costs of policies. To reverse the opioid crisis within a community, treatment policies may need to be combined with other strategies, such as harm reduction, supply reduction, and use prevention.

2 - Impact of Prescription Drug Monitoring Programs on Postpartum Opioid Prescribing Practices

Sai Tadi, The University of Texas at San Antonio, San Antonino, TX, United States, CeCe Cheng , Alexander Testa, John Byrne, Carolina Vivas-Valencia

Opioid prescriptions are commonly provided after childbirth, with over half of patients filling one within the first week of delivery. While effective for managing acute pain during recovery, higher doses or prolonged use raise risk of persistent opioid use, opioid use disorder, and overdose. This has led to federal and state initiatives to improve prescribing practices, including Prescription Drug Monitoring Programs (PDMPs) track controlled-substance prescriptions statewide. However, limited evidence exists regarding their impact in postpartum period. This study evaluates the association between PDMP and postpartum opioid prescribing from 2007-2015, prior to 2016 CDC opioid prescribing guidelines, and assesses differences between low-and high-risk postpartum populations. Low-risk was defined as opioid-naïve patients who receive ≤2 opioid prescriptions and ≤5 total days' supply within 42 days postpartum. High-risk was defined as patients with prior opioid exposure, documented opioid use disorder, or ≥3 opioid prescriptions, >5 total days' supply, or >700 cumulative morphine milligram equivalent (MME). We used a multi-year national claims database to identify females ages 12–59 with documented deliveries. We applied a difference-in-difference regression to compare changes in postpartum opioid prescribing based on PDMPs implementation timing, adjusting for age and medical comorbidities. PDMP implementation was not associated with statistically significant changes in MME per day, pills dispensed, days' supply, or total prescriptions during the postpartum period. Implementation during 2007-2015 was not associated with changes in postpartum opioid prescribing outcomes in either low or high-risk postpartum populations. Suggesting PDMPs alone have a limited impact on postpartum opioid prescribing patterns.

3 - Spatiotemporal Analysis of Opioid Overdose Hotspots in Texas

Crystal Holland, The University of Texas at San Antonio, San Antonio, TX, United States, Carolina Vivas-Valencia

Although opioid-involved overdoses have declined nationally over the past year, Texas continues to face a growing opioid burden, driven by rising overdose rates among Hispanics (40% of the state's population), and by increasing penetration of synthetic opioids such as fentanyl. Yet no prior study has evaluated the statewide spatial clustering of opioid overdoses or examined how area-level socioeconomic characteristics relate to overdose rates. In this study, we explore the spatiotemporal patterns of non-fatal opioid overdoses in Texas and how these geographic distributions change over time. Using outpatient data from Texas from 2018 to 2023, we employed Global Moran's I to evaluate overall spatial autocorrelation, followed by Local Indicators of Spatial Association (LISA) to identify statistically significant clusters. Additionally, we applied a geographically weighted Poisson regression (GWPR) model to evaluate whether area-level socioeconomic characteristics and demographic composition, are associated with overdose rates across Texas ZIP codes. Global Moran's I indicated statistically significant positive spatial autocorrelation in all study years, increasing from 0.12 in 2018 to 0.23 in 2023, with the sharpest rise between 2020 and 2021 (0.13 to 0.19). LISA maps revealed expanding High–High clusters around major metropolitan regions, particularly after 2020. High–High ZIP codes nearly doubled from 153 in 2018 to 300 in 2023. Persistent clustering showed sustained urban concentration, with Houston accounting for 26 ZIP codes for at least three consecutive years, followed by Austin (23) and Dallas (17). These localized patterns underscore the need for tailored interventions that reflect regional socioeconomic differences and evolving overdose dynamics.

4 - Integrating Fairness and Data Envelopment Analysis for Opioid Settlement Allocation

Qiushi Chen, The Pennsylvania State University, University Park, PA, United States, Paul Griffin

Multiple large opioid settlement agreements have been reached between the pharmaceutical industry and governments across the nation to mitigate the ongoing opioid crisis in the United States. While each state has begun to distribute the settlement funds following simple strategies such as per capita or an ad-hoc formula, it remains an open question how to develop a systematic approach to determine the settlement funds distribution across counties (or other political subdivisions) in a transparent, evidence-based, and fair manner. In this study, we develop a novel analytical framework that applies the fairness definition to the efficiency measures from the data envelopment analysis (DEA) to maximize the fairness of the allocation decisions. We formulate the decision problem in a bi-level optimization model, with the upper level problem representing the allocation decisions by the central planner (e.g., state government) and the lower level problem representing the county's assessment of

the allocation efficiency. We propose an efficient computational solution approach and demonstrate its application in a case study based on real-world empirical data in the context of opioid settlement allocation in Pennsylvania.

TB08

305A

Dementia Caregiving: Prospects and Potential for Smart AI-driven tools for Reducing Caregiving Burden

Invited Session

Healthcare Quality and Safety

Chair: Arunkumar Pennathur

1 - Strengthening the Dementia Caregiving Workforce: Lessons from the Geriatric Workforce Enhancement Program

Guillermina Solis, The University of Texas at El Paso, El Paso, TX, United States

The Geriatric Workforce Enhancement Program (GWEP)in the Borderland provides education and training on Alzheimer's disease and related dementia (ADRD) and established to health professionals, support staff, and informal caregivers. The U.S.-Mexico Border and Southern New Mexico region is characterized by limited access to evidence-based dementia education, early identification resources, trained providers, and coordinated systems of care. Significant barriers phased by the communities, include workforce shortages, scarce specialty services, and a lack of culturally and linguistically appropriate care services. Families in underserved and bilingual communities, often receive fragmented care that leads to late-stage dementia diagnoses and complications related to this. Additionally, persons experiencing homelessness, clients of non-profit social service programs, rural older adults, and caregivers frequently rely on informal information due to the absence of accessible, trusted resources.

The GWEP in the in the Borderland program integrates simulation-based digital learning with in-person programs to engage and strengthen workforce capacity and directly reach populations. We provide digitally accessible, interactive training for healthcare providers, first responders, nonprofit staff, and community health workers while simultaneously conducting on-site workshops, bilingual outreach, community listening sessions, and coordinated case discussions. The model ensures scalability through online programs while maintaining trust-building, face-to-face engagement where it is most needed. The presentation highlights the urgent dementia care challenges and to demonstrate how our GWEP addresses these gaps through culturally responsive, bilingual workforce training, community partnerships, and integration of the Age-Friendly Health System framework to improve early identification, education, and coordinated care for individuals living with ADRD and their caregivers.

2 - Ai Technologies Supporting Informal Dementia Caregivers: A Review of Applications, Barriers, and Future Opportunities

Juliana Guevara Mendoza, The University of Texas at El Paso, El Paso, TX, United States, Aditi Kaushal, Arunkumar Pennathur

Informal caregivers provide up to 80% of long-term care for persons with dementia, yet face significant challenges including emotional strain, information overload, and complex decision-making demands. Artificial intelligence (AI) technologies offer promising avenues to alleviate caregiver burden and improve care outcomes, but the evidence base remains fragmented. Our review synthesizes 50 peer-reviewed studies from Europe, North America, and Asia to provide a comprehensive taxonomy of AI technologies designed for informal dementia caregivers and evaluate their applications, demonstrated benefits, implementation barriers, and future opportunities.

Findings reveal diverse AI modalities, including conversational agents, sensor-based anomaly detection, large language models, reinforcement learning frameworks, and integrated IoT systems, supporting personalized

information delivery, emotional support, safety monitoring, and decision-making assistance. Studies report positive impacts on caregiver burden, stress, depression, and self-efficacy. However, widespread adoption is hindered by technological literacy gaps, privacy concerns, cultural mismatches, usability challenges, and limited real-world deployment. Methodological limitations including small sample sizes, heterogeneous outcome measures, and lack of longitudinal data constrain generalizability.

The review identifies critical gaps: insufficient cultural tailoring for diverse populations, absence of standardized evaluation frameworks, limited integration with existing healthcare systems, and scarce evidence on cost-effectiveness and scalability. Future directions emphasize personalized, culturally sensitive AI interventions developed through interdisciplinary co-design, explainable AI architectures to build caregiver trust, and scalable digital ecosystems that bridge the digital divide for people who are underserved by technology and living in rural settings. These findings inform healthcare operations researchers, technology developers, and policymakers seeking to optimize AI-driven interventions for effective dementia caregiver support.

3 - What Dementia Caregiving Can Learn from Community Health Worker Models

Catalina Alzate, University of Illinois Urbana-Champaign, Champaign, IL, United States

This presentation showcases an interdisciplinary research collaboration examining current and emerging uses of AI in community healthcare, with a central focus on health equity. Researchers from the University of Illinois Urbana-Champaign partnered with case managers and program coordinators at the Champaign-Urbana Public Health District to conduct participatory and speculative design research sessions to explore how AI is, or could be, integrated into community health contexts, with implications for dementia caregiving.

The project explores three key areas in technology design. First, how healthcare providers can develop AI literacy and competency, and gain a clear understanding of the ecosystems in which AI technologies are created and how these systems influence societal structures. Second, why AI implementation must be specific and context-driven. Rather than advocating for universal adoption, careful attention should be paid to use cases, decision-making processes, and actors involved in the use of these systems. Third, we explore how technology refusal is a valid alternative to AI adoption, when the foreseen risks outweigh the benefits.

The presentation leverages visuals, infographics, and research narratives to communicate these insights, offering practical tools for researchers and healthcare providers to analyze their own AI contexts. By emphasizing collective inquiry and participatory engagement, the work shows how AI can be implemented in ways that align with the lived realities of communities. Ultimately, this approach places people's agency at the center of healthcare, advancing equity while ensuring technology supports care in context-sensitive ways.

TB09

305B

Healthcare Simulation with Incorporation of Human Behaviors

Invited Session

Simulation Society

Chair: Nan Kong, Purdue University, West Lafayette, IN, United States

1 - A Dynamic Prediction Model for Opioid Overdose Among Medicare Beneficiaries with Incorporation of Evolving Information on Patient-Prescriber Ties and Geographic Context

Haopeng Zhou, Purdue University, West Lafayette, IN, United States, Paras Sandeep Punater, Nan Kong, Serena Jingchuan Guo, David R. Foster

Background: Opioid overdose remains a major public health concern in the United States. Timely identification of Medicare beneficiaries at elevated risk of overdose could inform targeted prevention efforts. However, most existing prediction models rely on static patient and prescriber characteristics and do not capture time-varying overdose risk or incorporate evolving prescription ties and shared geographic context. This study aims to

develop a dynamic prediction model for opioid overdose that updates risk over time using individual characteristics, evolving patient-prescriber ties, and geographic information.

Methods: A prognostic study of Medicare beneficiaries with an opioid prescription fill during 2014 - 2016 was conducted. Each recorded prescription fill was defined as the index event. Individual and contextual characteristics extracted from multiple predefined lookback windows prior to the index event are used to predict the risk of opioid overdose in the subsequent predefined outcome window. A heterogeneous graph neural network was developed in which patients and prescribers were represented as distinct node types connected by prescription ties. Node features and graph structure varied across index dates, capturing temporal changes in patient and prescriber characteristics, newly formed or discontinued prescription ties, and shared geographic context.

Expected Results: Model development and validation are ongoing. Explicit modeling of time-varying overdose risk using patient-prescriber ties and geographic context is expected to improve prediction compared with models based only on individual-level features or static representations.

Conclusions: A dynamic framework for opioid overdose prediction among Medicare beneficiaries is proposed that may improve risk stratification and support targeted prevention efforts.

2 - Digital Twin-Enabled Staffing Decision Support Platform Under Heterogeneous Service Demand of Nursing Homes Residents

Nan Kong, Purdue University, West Lafayette, IN, United States, Xuxue Sun, Hongdao Meng, Chris Masterson, Mingyang Li

Nursing homes (NHs) are a critical healthcare setting that provides 24/7 formal care and personal assistance to frail older adults. Adequate staffing is essential to ensure high-quality care and desirable resident outcomes. However, existing healthcare staffing literature often assumes homogeneous and/or static patient service demands, which may not reflect real-world complexities. This work proposes a digital twin-enabled staffing decision support platform for NHs that explicitly captures heterogeneous, temporally evolving, and diverse service demands of residents. The proposed framework integrates advanced statistical modeling, digital twin simulation, and optimization techniques to support staffing decisions. Specifically, latent variable models and statistical learning algorithms are developed to characterize the heterogeneous length of stay of NH residents by accounting for multiple discharge dispositions, as well as stochastic fluctuations in functional status over time during their stay. The NH residents resource utilization classification system is further refined to estimate the daily staffing time required by different types of caregivers for each resident group. For a heterogeneous population of NH residents with diverse individual characteristics, the developed prediction models are integrated with digital simulation to assess overall service demand at the facility level over time. Based on predicted demands, a two-phase decision analytics module is proposed to determine optimal resource planning decisions across multiple time scales, including annual bed capacity planning at the strategic level, bi-monthly nurse recruitment at the tactical level, and daily staffing levels and work schedules at the operational level. A case study is presented to demonstrate the efficacy proposed predictive optimization approach.

3 - Characterizing Trajectories of Co-Occurring Conditions in Individuals with Autism Spectrum Disorder

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States, Sai Tadi, Herta Montoya

Co-occurring conditions in autism spectrum disorder (ASD) often show atypical symptoms, complicating detection and long-term management. In 2022, about 1 in 32 U.S. children were identified with ASD by age eight. ASD involves deficits in social communication and restricted, repetitive behaviors. Individuals with ASD also show elevated rates of co-occurring conditions, including obesity, anxiety and mood disorders, seizure disorders, and gastrointestinal problems, compared with the general population. Nevertheless, prior studies consistently report a high comorbid burden, with many individuals presenting multiple co-occurring conditions. Using longitudinal claims data, this study characterizes temporal trajectories of psychiatric and medical comorbidities following ASD diagnosis. We developed an individual-level microsimulation model to estimate

personalized progression of co-occurring conditions based on observed demographics and clinical characteristics. Transition probabilities were derived from empirically estimated hazard functions, allowing simulation of condition onset, recurrence, and progression over time. The cohort included 323,385 individuals with ASD aged 2-21 years. We identified that developmental delays (50%), ADHD (45%), and speech/language disorders (43%) were the most prevalent conditions. Mood disorders were also observed, with depression affecting 15% of individuals. Medical comorbidities were also prominent, particularly gastrointestinal and immune disorders (35%), followed by sleep disorders (16%) and seizure disorders (9%). These preliminary findings demonstrate the substantial and heterogeneous comorbidity burden in ASD and illustrate the utility of structured longitudinal simulation for understanding individualized disease trajectories and informing long-term clinical management.

4 - Cost-Effectiveness Analysis of a Pharmacist-Led Anticoagulation Management Program for Patients with Atrial Fibrillation Receiving Direct Oral Anticoagulants

You Zhou, Purdue University, West Lafayette, IN, United States, Nan Kong, Brandon Bellows

Pharmacist-led programs have been proposed to improve health outcomes in patients with atrial fibrillation receiving direct oral anticoagulants, i.e., usual care, yet their cost-effectiveness remains uncertain, particularly under heterogeneous patient risks and adherence patterns.

We developed an individual-level simulation to evaluate the cost-effectiveness of implementing a representative pharmacist-led program compared with usual care. The model captures patient heterogeneity by simulating individualized risks of stroke and bleeding, which are estimated using validated risk scores. Medication adherence is measured by proportion of days covered (PDC). We modeled adherence dynamics by discretizing PDC into finite states and estimating state transition matrices for each patient subgroup based on the mean and standard deviation of observed monthly PDC trajectories. Adverse events, including bleeding complications, were explicitly modeled to induce treatment interruptions, thereby dynamically influencing both adherence patterns and subsequent clinical risks. Costs and quality-adjusted life years (QALYs) were accumulated over a lifetime horizon from a healthcare perspective.

In the primary analysis, a representative pharmacist-led program resulted in modest improvements in adherence and clinical outcomes but incurred additional costs, leading to an incremental cost-effectiveness ratio exceeding commonly accepted willingness-to-pay thresholds. Sensitivity analyses indicated that results were particularly influenced by the magnitude of adherence improvement and the baseline risk distribution. Subgroup analyses showed substantial heterogeneity, with patients characterized by elevated stroke risk and suboptimal adherence trajectory (e.g., early discontinuation) experiencing greater net benefits.

While the pharmacist-led program was not cost-effective at the population level, targeted implementation in selected high-risk subgroups may yield more favorable economic value.

TB10

306A

Visions for AI in Healthcare (Talk and Panel): Research Agendas, Opportunities, and Challenges

Invited Session

AI in Healthcare

Chair: Mehmet Ayvaci, UT Dallas, Richardson, TX, United States

1 - Designing and Evaluating AI in High-Stakes Healthcare: Bridging Models and Human Judgment

Mehmet Ayvaci, UT Dallas, Richardson, TX, United States

Advances in AI have led to substantial improvements in predictive performance in healthcare, yet these gains do not consistently translate into better decisions or improved patient outcomes. While increasing attention has been

paid to system design, workflow integration, and downstream impacts such as patient safety, less is understood about how AI interacts with human judgment in high-stakes clinical environments.

In this talk, I outline a research agenda focused on the gap between model performance and decision quality. Drawing on a portfolio of studies in medical decision-making, I highlight systematic discrepancies between how AI systems are evaluated and how they are used in practice. Factors such as communication, trust, cognitive biases, and user perception can lead to both over-reliance and under-utilization of AI, limiting its effectiveness even in well-designed systems.

I argue that evaluation frameworks centered on predictive accuracy or aggregate system outcomes are insufficient without understanding how decisions are formed and acted upon. This perspective motivates several research directions, including the design of human-AI decision processes, the development of evaluation metrics aligned with decision quality, and the study of behavioral and organizational factors that influence adoption and effectiveness.

By focusing on the interaction between AI and human judgment, this agenda aims to advance a more complete understanding of how AI can improve decisions in healthcare, not just predictions, systems, or outcomes.

2 - The Medical AI Revolution: a Healthcare OM Research Agenda

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

Medical AI is entering a new phase. For years, the field delivered technical advances, from expert systems to autonomous diagnostic tools, yet adoption in routine care remained slow. The rapid spread of ambient AI scribes has started to change that. For many clinicians, medical AI now means ambient tools, and their uptake suggests that workflow fit and incentive alignment can matter more for scale than model performance alone. Still, ambient AI is only a start. Better documentation does not by itself lead to better decisions, actions, or outcomes.

In this talk, I draw on a portfolio of recent work on medical AI to outline an operations management research agenda for the next phase of the medical AI revolution. I argue that OM has an important role in designing patient pathways, allocating decision rights between clinicians and AI, identifying when AI should act as a gatekeeper and when it should serve as a second opinion, and building systems that move from note automation to care transformation. The next phase centers on guided and agentic systems that carry out bounded, low-risk tasks under clinician supervision.

This agenda also raises broader questions about how to design AI-enabled care systems that are safe, effective, equitable, and affordable. In healthcare, AI succeeds through operations design as much as model development. The opportunity for OM scholars is to help shape medical AI systems that are scalable, trustworthy, and clinically meaningful across clinical settings and health systems.

3 - Ai in Healthcare: from Isolated Decisions to End-to-End Systems

Michael Lingzhi Li, Harvard Business School, Boston, MA, United States

Recent advances in AI in healthcare and life sciences have largely focused on narrowly scoped decisions, where models are designed and evaluated around a single task or prediction. These approaches have delivered strong performance in settings such as diagnosis, risk stratification, and clinical decision support, as well as in components of clinical trial design and execution. However, these domains are not organized around isolated decisions. Actions are embedded within complex, interconnected systems that evolve over time.

As a result, methods that perform well in static, task-level evaluations may fail to generate value at the system level. They may introduce unintended downstream consequences, interact poorly with existing processes, or induce decisions that are locally optimal but dynamically suboptimal when viewed over time. A decision that appears effective in the present can degrade performance across subsequent stages of care, trial execution, or drug development.

This gap motivates a shift from task-centric modeling toward a systems-level perspective on AI. I draw on examples from my research in healthcare delivery and clinical trials to illustrate how incorporating temporal dependencies, operational constraints, and human–algorithm interactions fundamentally changes both the design and evaluation of AI systems.

4 - Ai in Healthcare &Ndash; from Patient Safety to Patient Engagement

Rema Padman, Carnegie Mellon University, Pittsburgh, PA, United States

The US Institute of Medicine's landmark report in 1999, titled "To Err is Human: Building a Safer Health System", asserted that "the problem is not bad people in healthcare--it is that good people are working in bad systems that need to be made safer". It dramatically raised awareness of patient safety, knowledge and engagement issues in healthcare delivery to patients, researchers, practitioners and policy makers worldwide. Medical errors, such as surgical, medication, and diagnostic errors, cause serious harm to patients in every healthcare delivery setting, with aging populations being at particularly high risk for well-documented reasons such as multimorbidity and polypharmacy. Health IT and data analytics have been consistently recommended by successive studies to help detect and mitigate *preventable* errors and improve the safety of patients and the systems that provide care. However, 25 years later, the World Health Organization's Global Patient Safety Report 2024 indicates that almost one in ten patients is harmed in healthcare, resulting in three million deaths or more worldwide each year, with more than half of them being preventable. This discussion will highlight some of the major challenges in making healthcare systems safer and current research directions that combine AI, healthcare informatics, analytics and operations in meeting these challenges.

TB11

306B

Designing Healthcare Systems with the User in Mind

Invited Session

Healthcare Technology and Bioinformatics

Chair: Samuel Olatunji, The University of Tennessee, Knoxville, Knoxville, TN, United States

1 - Designing Healthcare Visualizations and Communication Systems with Older Adults in Mind

Renato Ferreira Leitao Azevedo, University of Georgia, Athens, GA, United States

Clinical test results are typically presented as numerical tables or graphs in both digital and clinical settings. Prior research shows that older adults often struggle to interpret results such as cholesterol, diabetes, and blood pressure readings, for example, failing to distinguish between normal and abnormal cholesterol levels or incorrectly classifying blood pressure readings as normal, elevated, or hypertensive. These difficulties are further compounded by the fact that many visualizations are designed primarily for healthcare providers rather than patients. As patients increasingly gain early access to their test results, misinterpretations may arise before they have the opportunity to discuss findings with their providers, potentially delaying care or compromising health decision-making. Poorly designed result visualizations can also place additional strain on the already limited availability of healthcare providers. Together, these challenges highlight the need for more intuitive, clinically aligned health data visualizations that accommodate varying levels of health, graphical and digital literacy. In this presentation, we present findings from both an in-person study and an online feasibility study examining older adults' preferences and responses to different communication modalities for receiving clinical test results. The study aims to (1) identify which communication medium produces the most favorable outcomes by evaluating whether these systems should function purely as informational tools or as relational agents emulating healthcare interactions, and (2) assess their effects on comprehension, emotional responses, and risk perception relative to traditional text-based formats. These findings inform the design of user-centered healthcare solutions that enhance health communication for aging populations.

2 - Engaging Older Adults with Disabilities for Participatory Design of An Assistive Robot

Harshal Mahajan, University of Illinois Urbana Champaign, Champaign, IL, United States

Participatory design methods aim at engaging target end users of technologies early in the design process. Employing these processes ensures that the design addresses needs that are deemed important by the targeted users. We will present two case studies where we implemented a systematic design process to evaluate and improve the Stretch assistive robot (Hello Robot Inc.). Older adults with disabilities interacted with Stretch in a home-like lab setting and in a naturalistic home environment. There was a broad variation in the mobility and cognitive status of the thirteen older adult participants (age 60-97). The sample included ambulatory individuals, mobility device users, and quadriplegia and self-reported cognitive impairment. Lack of familiarity, discomfort with, and anxiety around the robot could impact attitudes of the target users. Hence, in these studies, we systematically captured user perceptions and gathered feedback by gradually increasing exposure to the robot. This approach provided a unique opportunity to identify changes in perception and experience with exposure to the robot and implications for deploying such robots to support older adults with a range of abilities. We also evaluated comfort, perceptions, concerns, and design ideas for using Stretch for in-home activities.

3 - *CANCELED Designing Health Care Systems for Humans: a Sociotechnical Approach to Safer, Smarter Care

Abigail Wooldridge, University of Illinois Urbana-Champaign, Urbana, IL, United States

While the rapid proliferation of advanced technologies, artificial intelligence and novel medical treatments can lead to a focus on the technological aspects of health care, humans are central as patients, care partners/caregivers, clinicians and other healthcare workers. Further, clinical work occurs in complex, dynamic sociotechnical systems shaped by interactions between people, technology, and environments, with processes unfolding across time and space. This presentation will draw on human factors/ergonomics research across multiple health care settings (e.g., surgery, maternal care, care transitions within hospitals, resuscitations) to demonstrate how human factors/ergonomics can fit with other areas of industrial and systems engineering to examine health care as an interconnected system centered on users – teams of clinicians, patients, care partners/care givers and others – to improve outcomes for all. This presentation will showcase practical applications of human-centered principles and processes for sociotechnical systems design, including analysis, design and evaluation efforts that integrate augmented reality and optimize team cognition in high-stakes settings such as trauma. Our research reveals that high-performing tools and technology may fail to have positive, intended impacts when they fail to fit with workflows and meet users' needs. The presentation serves as a call to design healthcare technology and systems *with* users —not only for them—situating human needs, interactions, and work contexts as key to system innovation.

TB12

306C

Deployment and Governance for AI in Healthcare: Key Challenges

Panel Session

Healthcare Practice

Co-Chair: Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, NY, United States

1 - Moderator Panelist

Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, NY, United States

2 - Panelist

Bethany Daily, Massachusetts General Hospital, Boston, MA, United States

3 - Panelist

Andreea Bodnari, ALIGNMT AI, New York, NY, United States

Panelist

4 - Panelist

Eric Poon, Duke University Health System, Durham, NC, United States

TB13

307

Analytical Advances in Healthcare Operations

Invited Session

Location Analysis

Chair: Xiaoyu Wu, Fuqua Business School, Duke University, Durham, NC, United States

1 - To Abandon OR Not? Understanding the Impact of Participant Retention on Clinical Trial Outcomes

Xueze Song, The University of Alabama, TUSCALOOSA, AL, United States, Mili Mehrotra, Ujjal Kumar Mukherjee, Tharanga Rajapakshe

The process of drug development is replete with significant risks and uncertainties, with as many as 12% of clinical trials terminated early. One leading factor of trial termination is poor recruitment or retention. In this work, we study the impact of strategies to improve participant retention on clinical trial sponsor's decision to terminate the clinical trial. Specifically, we address the following research questions: How does the adoption of incentives and effort affect the sponsor's decision to terminate the clinical trial? How does the impact of such strategies vary across different clinical trials? How can the sponsor adjust their retention strategies to minimize the likelihood of termination? To answer these questions, we formulate a sponsor's decision to terminate a clinical trial as an optimal stopping problem with partial information on the drug's efficacy. We then leverage the data from the clinical study registry from the U.S. and empirically estimate the effect of various clinical trial characteristics on the likelihood of termination. Our analysis sheds light on the key factors that drive clinical trial termination and can inform future drug development efforts.

2 - Mitigating Shortages of Generic Drugs: the Role of Reliability Certification

Xiaoyu Wu, Fuqua Business School, Duke University, Durham, NC, United States, Robert Swinney, Can Zhang

Generic drugs play a vital role in keeping costs down in the U.S. healthcare system. However, generic drugs have experienced severe shortages due to supply disruptions, caused by a number of factors, including a lack of information available to buyers about manufacturers' reliability, leading manufacturers to underinvest in supply reliability. FDA has recently sought to address generic drug shortages, and one proposal is to implement a certification system that assesses and discloses the reliability of generic drug manufacturers. In this paper, we investigate the effectiveness of such a certification system.

We propose a game theoretic model that captures the effects of generic drug manufacturer reliability certification on manufacturers, buyers, a group purchasing organization, and patients. We find that certification may motivate manufacturers to improve their reliability; however, this may increase patients' shortage cost once the manufacturers adjust their wholesale prices and the buyers adjust their sourcing strategies in response. Nevertheless, we also identify conditions under which certification can benefit all stakeholders compared to the status quo with no certification. Finally, we find that combining certification with subsidies can further reduce the shortage cost when the subsidy level is sufficiently high, but it may backfire if the subsidy level is too low.

Our findings offer insights for policymakers and nonprofit organizations in designing reliability certifications to mitigate generic drug shortages, highlighting when and why certification can benefit various stakeholders in the supply chain, while cautioning against unintended consequences arising from its impact on manufacturers' price competition and buyers' sourcing strategies.

3 - AI-Augmented Gatekeeping: Combining Human and Machine Intelligence in Screening Programs

Jiatao Ding, Nanyang Technological University, Singapore, Singapore, Michael Freeman, Tinglong Dai

Health screening programs increasingly deploy AI to prioritize patients and triage follow-up, but the central operational challenge is how to combine AI and clinician review under limited capacity. We develop an analytical model of an AI-augmented screening program in which each case may be automatically cleared,

automatically referred for downstream diagnostic workup, or escalated to human review. The program jointly chooses AI capability, the operating threshold, and the routing policy for AI-negative and AI-positive cases, accounting for false negatives, false positives, review costs, referral costs, and limited review capacity. We show that optimal review is concentrated on borderline cases near the AI threshold rather than spread uniformly across patients. With scarce clinician capacity, the optimal workflow can be one-sided (i.e., either reviewing AI-positive cases to filter false alarms or reviewing AI-negative cases to rescue missed disease), before expanding to two-sided review as capacity grows. The optimal AI threshold generally differs from the AI-only benchmark because it determines not only sensitivity and specificity, but also which review protocol handles marginal cases. In a dynamic extension, current routing choices affect future clinician capability; preserving expertise can justify maintaining review on cases that appear unattractive in the short run. These results provide a framework for designing AI-enabled health screening workflows and coordinating threshold setting, staffing, and AI investment to balance safety, workload, and long-run human expertise.

4 - Clinical AI, Medical Reasoning, and Hospital Workforce

Qi Wang, Boston University, Boston, MA, United States, Anita Carson

This paper examines how Clinical and Operational AI adoption affects workforce composition in U.S. hospitals. Using position-level data across U.S. hospitals, we develop Diagnostic and Management Reasoning Intensity scores to investigate whether the workforce effects of AI vary with the reasoning demands of healthcare work.

Tuesday, July 28, 2026, 1:30 PM - 2:45 PM

TC01

301A

Transforming Healthcare with AI/ML: What's Next for OR?

Panel Session

AI in Healthcare

Co-Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, United States

1 - Moderator Panelist

Michael Lingzhi Li, Harvard Business School, Boston, MA, United States

A panel on how AI and machine learning are reshaping operations research in healthcare. Panelists will discuss recent impact, barriers to implementation, and emerging directions.

2 - Panelist

Tinglong Dai, Johns Hopkins University, Baltimore, MD, United States

This is a panel, and I don't expect to give an oral presentation.

3 - Panelist

Brian Denton, University of Michigan, Ann Arbor, MI, United States

4 - Panelist

Jun Li, Ross School of Business, University of Michigan, Ann Arbor, MI, United States

5 - Panelist

Julie Swann, NC State University, Raleigh, NC, United States

TC02

301B

Frontiers in Healthcare Operations and Analysis

Invited Session

Applied Probability

Chair: Alan Scheller-Wolf, Tepper School of Business, Pittsburgh, PA, United States

1 - Fast Non-Episodic Finite-Horizon RL with K-Step Lookahead Thresholding

Kyra Gan, Cornell Tech, Cornell University, New York, NY, United States, Jiamin Xu

Online reinforcement learning in non-episodic, finite-horizon MDPs remains underexplored and is challenged by the need to estimate returns to a fixed terminal time. Existing infinite-horizon methods, which often rely on discounted contraction, do not naturally account for this fixed-horizon structure. We introduce a modified Q-function: rather than targeting the full-horizon, we learn a K-step lookahead Q-function that truncates planning to the next K steps. To further improve sample efficiency, we introduce a thresholding mechanism: actions are selected only when their estimated K-step lookahead value exceeds a time-varying threshold. We provide an efficient tabular learning algorithm for this novel objective, proving it achieves fast finite-sample convergence: it achieves minimax optimal constant regret for $K=1$ and $\mathcal{O}(\max((K-1),C_{K-1})\sqrt{SAT\log(T)})$ regret for any $K \geq 2$.
We numerically evaluate the performance of our algorithm under the objective of maximizing reward. Our implementation adaptively increases K over time, balancing lookahead depth against estimation variance. Empirical results demonstrate superior cumulative rewards over state-of-the-art tabular RL methods across synthetic MDPs and RL environments: JumpRiverswim, FrozenLake and AnyTrading.

2 - A Structural Analysis of Organ Procurement Organization Decisions in the United States

Shubham Akshat, University of Houston, Houston, TX, United States, Ariel Rava, Nisarg Nigam, Emanuel Zur

An estimated more than 25,000 transplantable organs are not recovered every year from potential organ donors in the United States. These organs could have saved the lives of thousands of patients on the waiting list. The Organ procurement organizations (OPOs) play a central role in converting potential donors into successful transplants, yet many of their most consequential choices happen upstream of any single organ-offer — starting with whether to pursue a referral at all, and continuing with which organs to recover and place from a given donor. In this study, we develops forward-looking framework for OPO decision-making that links these multi-organ choices to financial incentives and operational costs. We use data from the Scientific Registry of Transplant Recipients (SRTR) and unique hand-collected data from between 2013 to 2023 via Freedom of Information Act (FOIA) requests to detail costs, revenue, and operations in the organ procurement system, providing a comprehensive view of performance as well as cost comparisons across organ types, and OPOs.

3 - Shift Effect in Emergency Departments: Empirical Evidence and Assessment of Impact

Opher Baron, Rotman School of Management, University of Toronto, Toronto, ON, Canada, Shany Azaria, Dmitry Krass

The *shift effect* in healthcare and other service systems arises when completing service requires multiple interactions with a provider and when provider priorities vary over the course of a work shift. In emergency departments (EDs), this typically occurs when a patient first receives a Physician Initial Assessment (PIA), triggering tests or imaging, followed later by a second assessment leading to disposition.

Physicians working fixed-length shifts often prefer to complete care for patients they have already seen before their shift ends. This preference may be driven by compensation structures, as billing is often higher when care is completed, as well as by continuity-of-care concerns, since handoffs are perceived to pose risks to quality. As a result, physicians tend to prioritize new patients early in their shifts and patients awaiting reassessment later in the shift. While individually rational, this behavior can increase both average and tail probabilities of waiting times, contributing to ED congestion.

Although the shift effect is widely acknowledged anecdotally by ED clinicians, rigorous analytical evidence remains limited. In this study, we use detailed operational data from a medium-sized Canadian hospital ED to empirically investigate the presence and impact of the shift effect. Consistent with Ontario's Pay-for-Results framework, we examine Time to Physician Initial Assessment (TPIA) and Emergency Department Length of Stay (EDLOS), focusing on both mean values and the 90th percentile. Using the SiMLQ-Emergency simulation

platform, we show that the shift effect can account for up to 35% of EDLOS and discuss operational strategies to mitigate its impact.

4 - Resource-Constrained Healthcare Optimization: Effect of Tenure and Learning

Alan Scheller-Wolf, Tepper School of Business, Pittsburgh, PA, United States, Canary Zhu

Healthcare systems worldwide face the critical challenge of optimally allocating limited resources across diverse patient populations while ensuring equitable access and outcomes. We consider a healthcare provider who manages a fixed-tenure outpatient program -- for example a 30-day program to reduce 30-day readmission rates for recently discharged heart patients. The provider must allocate limited treatment resources among a fixed number of heterogeneous patients in the program: Patients vary in health state and time until they complete the program.

We construct an MDP framework that captures the stochastic nature of patient health trajectories while incorporating realistic resource constraints. As exact solutions are only available for small instances, we leverage known asymptotically optimal relaxations for larger instances, as well as other simple heuristics. These enable us to solve realistic-sized problems and to evaluate how solutions change under different utility/cost/penalty structures.

Results reveal significant variations in allocation patterns and outcomes depending on whether patient tenure is taken into account when making treatment decisions. When caregiver learning effects are additionally taken into account (empirically, care has been found to become more efficient as caregivers become more familiar with their patients) these effects can become stronger.

Our results motivate consideration of how maximizing rewards in constrained healthcare environments can lead to provision of disparate levels of care to patients having the same conditions, and how different objective functions may amplify or dampen these effects.

TC03

302A

Innovations in Healthcare Operations

Invited Session

Healthcare Operations Management

Chair: Xiaoquan Gao, Singapore Management University, Singapore, Singapore

1 - Soft-Binned Distributionally Robust Multimodal Learning

Yu Ma, University of Wisconsin Madison, Madison, WI, United States

Multimodal machine learning systems are increasingly deployed in high-stakes domains, yet they remain vulnerable to distribution shifts that affect different modalities in heterogeneous ways. Existing robust learning approaches typically rely on either heuristic uncertainty modeling or consider all modalities jointly, downplaying the effects of modality-specific noise. We propose Soft-Binned DRO, a distributionally robust optimization framework that explicitly imposes per-modality marginal constraints. To ensure tractability in high dimensions, we introduce a kernel-based soft-binning scheme that yields a smooth, differentiable, and scalable surrogate. We develop a stabilized primal-dual algorithm and provide theoretical guarantees on its generalization bound and sublinear convergence rate. We further show that Soft-Binned DRO induces a structured relaxation of empirical risk minimization, bounding the adversarial influence to semantically meaningful directions. This explains how it avoids the pessimistic overfitting of standard sample-level DRO approaches. Empirically, across controlled simulations and a real-world clinical dataset, our framework delivers consistent robustness gains over strong baselines, validating modality-aware ambiguity sets as a principled foundation for robust multimodal learning.

2 - Enhancing Mental Health Service Accessibility on University Campuses: the Impact of Tragic Events

Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States, Xiaoyan (Amber) Liu, Wilson Lin

Student mental health has emerged as a pressing concern in higher education. In contrast to the growing demand among students for mental health services, universities have lagged significantly in prioritizing these needs. Particularly, following tragic events such as student deaths, impacted individuals often experience substantial emotional or psychological suffering, requiring appropriate crisis-response operational strategies to sustain access to mental health care. We examine how universities can enhance access to mental health counseling services on campus in light of the occurrence of tragic events. Leveraging a comprehensive dataset from the Counseling and Psychological Services center of a U.S. university, we address this issue by first estimating the impact of a tragic event on demand for mental health counseling services through a difference-in-differences analysis. We find that demand increases significantly, with individuals being 124% more likely to make an appointment in the immediate days following the tragic event. Among impacted individuals, those who belong to the proximate cohort of the victims are more likely to seek care. Second, we propose a multi-period optimization model that accounts for the anticipated demand surge and strategically balances the trade-off between meeting immediate and future demand, guiding budget allocation decisions for staffing mental health clinicians in a rolling-horizon approach. Our optimization model yields substantial gains compared to less sophisticated benchmarks used in practice, including a myopic approach that depletes the remaining budget immediately following a tragic event without accounting for future needs, and an empirically-informed approach that partially accounts for future needs.

3 - Dynamic Scheduling with Switching Costs: Theory and Applications to Healthcare Operations

Xiaoquan Gao, Singapore Management University, Singapore, Singapore, Yvonne Huijun Zhu, Joel Goh

In many service operations, frequent task switches can lead to setup delays and operational inefficiencies, particularly when transitions require equipment reconfiguration or context shifts across task types. We study this problem in a dynamic environment where appointments must be assigned in real time without knowing future task types. Starting from a static sequencing model, we derive explicit structural conditions analogous to the cμ and shortest weighted processing time rules under which mirror stacking schedules are optimal. We then generalize these insights to a stochastic dynamic setting, where analytical structure guides the design of a tractable online scheduling heuristic. Simulation experiments calibrated with data from a major hospital's imaging department show significant reductions in transition delays and improvements in daily throughput and patient waiting times.

4 - The Effects of Information on Abandonment and Congestion in Non-Stationary Priority Queues

Junqi Hu, Old Dominion University, Virginia Beach, VA, United States, Philipp Afeche, Rouba Ibrahim, Vahid Sarhangian

In observable priority service systems, e.g., the emergency department of a hospital, customers are typically left to infer their own information about their queue positions in the waiting room. Because such inferences are likely inaccurate, the system manager may choose to provide accurate information. This paper examines the impact of providing (queue position) information on key operational performance measures. We consider a two-class, multi-server priority queue with time-varying arrivals and state- and information-dependent abandonment rates. We compare two information levels: no information (abandonment rates depend on customers' perceived queue positions) and full information (abandonment rates depend on customers' true queue positions). For each level, we establish fluid approximations of the queue length process, and the existence, uniqueness, and asymptotic stability of a periodic fluid model equilibrium. We leverage these approximations for analytical comparison results of these information levels in terms of the abandonment rate, queue-length, and total cost of each class. These results yield the following guidelines: (1) Providing queue position information does not uniformly improve performance, but instead involves two trade-offs: (i) within each priority class, between queue length and abandonment; (ii) across priority classes, i.e., information may benefit one class while harming the other. These effects are most pronounced when the high-priority class alternates between under- and overloaded. (2) Therefore, the information policy that minimizes the total cost depends on the load, customer perceptions, and the relative costs of delay versus abandonment. (3) Effective information design may require class-specific disclosure or operational levers, e.g. staffing, to eliminate trade-offs.

TC04

302B

Operational Decision-Making in Healthcare with AI and Advanced Analytics

Invited Session

Healthcare Operations Management

Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States

Co-Chair: Xiaoyan (Amber) Liu, Santa Clara University, Santa Clara, CA, United States

1 - The Winner's Curse: Why Optimizing Structural Models May Overpromise and Underdeliver

Neha Sharma, The Wharton School, University of Pennsylvania, Philadelphia, PA, United States, Hamsa Bastani

Deploying data-driven policies—from regulatory design to dynamic pricing—entails substantial fixed costs and consumer-facing risk. As a result, adoption decisions hinge on credible ex ante estimates of policy gains. In a growing subset of prescriptive work, policies are selected by optimizing an objective built from estimated primitives and then evaluated using the same estimates. We show that this coupling can produce systematically optimistic projected gains: optimization selects actions that benefit from favorable estimation noise. We characterize this optimism bias to leading order, separating a selection (optimizer's curse) component from a curvature component arising from nonlinear policy evaluation. We further show that several popular uncertainty-aware or decision-focused variants—such as optimizing under an estimated parameter distribution or training models end-to-end for a downstream decision—do not generally eliminate the bias and can increase it by amplifying sensitivity to estimation error in decision-relevant directions. We then propose policy evaluation via analytic correction with cross-fitting and pessimistic optimization using a tuned lower-confidence-bound objective. Together, these tools improve the credibility of policy recommendations by reducing false positives while preserving the ability to identify genuinely profitable opportunities.

2 - Beyond Optimization: Fast and Robust Operating Room Scheduling via Reinforcement Learning

Xenia Konti, Duke University, Durham, NC, United States, Michael Zavlanos

Operating room (OR) scheduling is one of the most resource-intensive planning problems in hospital management, with OR costs accounting for a substantial share of total hospital expenditure. Each surgical day requires assigning surgical cases to ORs such that patients flow efficiently through a pipeline of shared, constrained resources. The schedule must balance competing objectives around resource utilization and timely case completion, while respecting a range of operational constraints. Inherent uncertainty in surgical durations further compounds the difficulty, making the problem computationally demanding for traditional optimization approaches. We propose a hierarchical reinforcement learning (RL) framework trained with Proximal Policy Optimization (PPO) to address this challenge. A strategic agent determines the number of ORs to activate, while a tactical agent sequentially assigns cases to ORs by interacting with a discrete-event simulator that enforces all resource constraints in real time. Transformer-based state representations capture dependencies across cases, surgeons, equipment, and OR availability. Unlike optimization-based approaches, which can require up to 30 minutes to produce a schedule, the trained policy generates solutions in seconds, while exhibiting greater robustness to duration uncertainty through exposure to diverse historical surgery days during training.

3 - Designing Effective Wellness Incentives Under Delay Discounting Heterogeneity

Hung Tran, Virginia Tech, Blacksburg, VA, United States, Sait Tunc

Unhealthy behaviors drive substantial preventable morbidity and spending, motivating widespread use of wellness programs that reward either participation (effort) or attainment (outcomes). Despite widespread adoption, many programs have a limited impact on the populations most in need. We study a missing behavioral mechanism in existing incentive-design models: unhealthy behavior is strongly associated with elevated delay discounting, suggesting that delayed, target-based rewards may be significantly less motivating for high-risk individuals. We develop a discrete-time, finite-horizon model of a planner designing participation and attainment incentives for a heterogeneous population that differs in health status and delay discounting. The model characterizes forward-looking compliance decisions, induced health trajectories, and optimal incentive design under planner objectives ranging from welfare-oriented to self-interested. Our results show how incentive timing and incentive type interact with delay discounting heterogeneity, clarifying when attainment-heavy designs can

systematically underserve high-risk groups and when earlier, compliance-based incentives are needed to improve reach and equity.

4 - End-to-End Uncertainty-Aware Optimization for Multi-Stage Infusion Center Scheduling

Manos Giannopoulos, Duke University, Durham, NC, United States, Yi Shen, Michael Zavlanos

Scheduling patients in infusion centers is inherently complex: a single visit spans multiple sequential stages — pre-screening, clinical evaluation, medication preparation, and infusion — each involving distinct resources, staff, and stochastic durations. Current practice schedules these stages in reverse order, anchoring on infusion slot availability and backfilling upstream steps, a myopic approach that propagates bottlenecks and inflates patient wait times. We present an end-to-end, uncertainty-aware framework for joint optimization across all stages. Our framework couples predictive models for stochastic duration estimation with a combinatorial optimization model, validated and stress-tested via discrete event simulation, and refined through a feedback loop on actualized schedules. Together, these components shift infusion center scheduling from reactive slot-filling to proactive, system-aware planning.

TC05

302C

Incentives, Operations, and Policies in Public Health

Invited Session

Public Health Policy

Chair: Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

1 - Dynamic Staffing for Hospitals in a Pandemic

Yvonne Huijun Zhu, Smeal College of Business, Pennsylvania State University, State College, PA, United States, Joel Goh, Yee Wei Lim, Louisa Jin

During a pandemic, surging infections and uncertain patient demand place severe strain on healthcare systems. This situation often leads to shortages of healthcare personnel, particularly nursing staff, thereby compromising patient care and outcomes. We develop a dynamic staffing model for healthcare providers. A key feature of the model is that healthcare workers themselves face infection risk and must undergo recovery before returning to duty. The model partitions staff into two teams: a hot team serving infectious patients and a clean team serving non-infectious patients. By anticipating future patient demand and accounting for workforce infection dynamics, we characterize optimal dynamic staffing policies across teams. Our framework provides actionable guidance for hospitals to allocate and schedule healthcare workers efficiently during future infectious disease outbreaks.

2 - Telehealth Initiation and Continuity of Care in Outpatient Psychotherapy: Evidence from a Nationwide Claims Database

Yi Tang, Old Dominion University, Norfolk, VA, United States, Shenghan Li, Kingshuk Sinha

Mental health has become a growing public health crisis in the United States, yet significant barriers — including provider shortages, geographic distance, and stigma — continue to limit access to consistent care. The rapid expansion of telehealth, accelerated by the COVID-19 pandemic, has introduced a new modality for delivering psychotherapy, though its implications for patients' sustained engagement remain poorly understood. Continuity of care is critical in psychotherapy, where therapeutic alliance and treatment consistency are central to recovery. Using a nationwide claims dataset, this study examines how telehealth utilization at both the initial and follow-up visits affects treatment continuity and overall service utilization among patients receiving outpatient psychotherapy. Findings from this study have implications for policymakers, insurers, and clinicians navigating the integration of telehealth into behavioral health care delivery.

3 - Countering Information Avoidance: Optimal Incentives for Diagnostic Testing During a Pandemic

Zack Zhu, University of British Columbia, Vancouver, BC, Canada, Steven Shechter, Woonghee Huh

Many individuals choose to avoid medical diagnostic tests, even when the results can have significant impacts on follow-up decisions. In the case of highly contagious diseases, test results can affect quarantine decisions,

which can greatly affect the rate of further disease spread. Studies have reported that a substantial portion of the population may have avoided COVID-19 tests, even when they were freely available. We develop a decision-analytic model that explicitly incorporates the psychological impacts of learning one's disease status, leading to structural results explaining when individuals will prefer to actively avoid the information of diagnostic tests. We validate our model-based results against independent estimates of the prevalence of COVID-19 testing avoidance and then use our model to evaluate governmental interventions for addressing this behavior. Specifically, we perform a cost-effectiveness analysis of different compensation levels for taking tests and find that in addition to offering free at-home COVID-19 test kits, governments may want to consider actually paying individuals to test.

TC06

303

Data-driven Inverse Optimization

Invited Session

Medical Decision Making

Chair: Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States

Co-Chair: Farzin Ahmadi, Johns Hopkins University, Baltimore, MD, United States

1 - Inverse Optimization for Learning Noisy Decisions

Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States, Farzin Ahmadi, Fardin Ganjkhanloo

Inverse optimization recovers unknown parameters of an optimization problem from observed decisions assumed to be optimal or near-optimal. The field has produced powerful estimation methods, yet practical deployment calls for scalable formulations and identifiability of the solutions. This talk presents a unified framework that advances both fronts. We introduce Inverse Learning, which shifts the inferential target from the unknown parameters to the latent optimal decisions, yielding a formulation whose size is independent of the number of observations. We show that inverse optimization is in general non-identifiable, even with normalization and multiple observations, the parameter set compatible with data is generically multi-dimensional, and regularization does not resolve this. We then show the necessary and sufficient conditions for identifiability. To address the tension between observational fidelity and constraint adherence, we formalize the Observation-Constraint Tradeoff and develop Goal-Integrated Inverse Learning models that enable structured navigation of this spectrum with guaranteed monotonicity. Our results show higher parameter recovery rates and computational speedups. We apply the framework to personalized dietary recommendations using NHANES data and a proof-of-concept demonstrating improved glycemic control in a prospective feasibility study.

2 - Generative Inverse Optimization for Healthcare Decision-Making

Danielle Ripsman, Johns Hopkins University, Baltimore, MD, United States, Kimia Ghobadi

In this talk, we propose a framework for learning difficult decision-making problems in healthcare, using the applications of radiation therapy and the diet problem. In radiation therapy, modeling new treatment plans based on historical databases remains an open question. While Generative AI (GAI) is a potential avenue, its opaqueness and instability hinder adoption. Similarly, the diet problem requires "nudging" users toward healthy, palatable choices, which is a gap that GAI has failed to fill reliably.
Building on inverse optimization (IO), we develop Generative IO (GIO), which allows for the generation of constraints outside of a rigid IO structure. Unlike GAI, GIO maintains the mathematical interpretability of IO while offering a less rigid framework than traditional models. Using this expanded philosophy, we demonstrate how these complex healthcare problems fit into the GIO paradigm, ultimately fostering better cohesion between expert practitioners and optimization tools.

3 - Non-Asymptotic Guarantees for Inverse Optimization: A Finite-Sample Framework via Concentration Inequalities

Farzin Ahmadi, Towson University, Towson, MD, United States, Kimia Ghobadi

Inverse optimization (IO) provides a principled framework for inferring the latent objectives and parameters of an optimization model from observed decisions — a task of fundamental importance in healthcare, where clinician choices, patient behaviors, and treatment protocols encode complex, often implicit, preferences. Classical IO methods assume noiseless, perfectly rational observations, an assumption that fails systematically in clinical data. While recent statistical extensions accommodate noisy observations, they rest on asymptotic consistency arguments that offer little actionable guidance when, as is typical in healthcare, only tens or hundreds of observations are available.

This paper develops a comprehensive non-asymptotic theory for parameter estimation in IO. We establish finite-sample concentration bounds on estimated optimal solutions and recovered parameters under sub-Gaussian noise, characterizing precisely how estimation error decays with sample size, problem dimension, and confidence level. Building on these bounds, we introduce IO-GBF (Inverse Optimization with Geometric Bound Feasibility), a minimax estimation framework that jointly optimizes over the KKT-feasible set of the forward problem and a statistical confidence ball centered at the sample mean. This geometry-aware formulation admits a tractable dual reformulation and yields provably tighter worst-case bounds than existing approaches. For linear forward problems, we develop an exact polyhedral algorithm enumerating all candidate parameter estimates, and derive a sample complexity model — solved via bisection — prescribing minimum data requirements for a target estimation guarantee.

Numerical experiments in a healthcare resource-allocation setting demonstrate substantial accuracy improvements over standard IO benchmarks, particularly in small-sample regimes, with direct implications for clinical decision support and protocol learning.

4 - Multi-Objective Constrained Optimization Inference via Partial Outcome Observation

Emmett Springer, Johns Hopkins University, Baltimore, MD, United States, Kimia Ghobadi

Despite advances in operations research over the last several decades, the healthcare industry remains slow to adopt potentially relevant methodologies for optimizing decision-making. Many problems involve optimizing decisions for a broadly defined objective which cannot easily be modeled as a single objective function with known parameters. For example, healthcare practitioners may want to maximize the efficiency of care delivery, which entails providing quality, effective treatment to patients who need it while avoiding providing costly treatments to patients who will not benefit from them. The parameters driving a decision's constraints and outcomes are often unknown (i.e. how the treatment options affect outcomes under different clinical conditions) and must be inferred from data. In contrast with current approaches, we focus on settings where decisions, while not assumed to be optimal, contain valuable information about hidden constraints and/or objectives not directly observable or measurable by decision outcomes. Furthermore, we consider unobserved parameters that cannot be directly observed, and rather are parameters for the constructed cost function. This setting reflects the complexity of expert decision-making in healthcare. We define a framework for Multi-objective Constrained Optimization Inference via Partial Outcome Observation, central to which is the goal of learning jointly from past decisions and from partially-observed decision outcomes. Our approach draws inspiration from inverse decision modeling, contextual optimization, and off-policy evaluation.

5 - Maximum Likelihood Estimation and Confidence Regions for Inverse Optimization

Fardin Ganjkhanloo, Johns hopkins University, Washington, DC, United States, Farzin Ahmadi, Kimia Ghobadi

Inverse optimization recovers unknown parameters of an optimization problem from observed decisions, with growing applications in healthcare for inferring clinical preferences, treatment priorities, and resource allocation strategies from practice data. Existing methods recover point estimates by minimizing loss functions but provide no uncertainty quantification, a critical gap in clinical settings where decision-makers need to know how much to trust inferred parameters.

We develop a maximum likelihood estimation (MLE) framework for inverse convex optimization that enables both principled parameter recovery and rigorous confidence regions for the inferred parameters. We show that

existing loss-minimization approaches are special cases of our framework, corresponding to MLE under isotropic Gaussian noise, while the general formulation accommodates anisotropic and non-Gaussian noise models. Our confidence region construction decomposes inference into two steps: discrete model selection that identifies which active constraint manifold contains the true optimum via profile likelihood, and continuous within-manifold estimation. Numerical experiments validate coverage properties across diverse problem geometries. By equipping inverse optimization with uncertainty quantification, this framework supports more reliable and interpretable inference of clinical decision-making models from noisy observational data.

TC07

304

Using Decision Science to Guide Healthcare and Public Health Interventions

Invited Session

Health Applications Society

Chair: Isabelle Rao, University of Toronto, Toronto, ON, Canada

1 - Modeling Health and Economic Outcomes of Providing Stable Housing to Homeless Adults with Opioid use Disorder

Isabelle Rao, University of Toronto, Toronto, ON, Canada, Margaret Brandeau

The number of people experiencing homelessness (PEH) in the United States is increasing, and drug overdose, primarily driven by opioids, is the leading cause of death in this population. Addressing opioid use disorder (OUD) among PEH is a critical public health priority.

We develop a continuous time compartmental model to evaluate the health and economic impacts of providing stable housing to PEH with OUD, without requiring entry into OUD treatment. We simulate a cohort of 1,000 PEH with OUD, comparing outcomes under the housing intervention to those under the status quo (no housing provision) to determine the incremental cost-effectiveness ratio (ICER).

Over five years, the housing intervention reduces total deaths from 191 to 140, and reduced total overdoses from 554 to 494. The intervention increases lifetime quality-adjusted life years (QALYs) by 3.59 per person at an incremental cost of $26,200 per QALY gained compared to the status quo. Across a wide range of sensitivity analyses that could make the housing intervention appear less favorable, the ICER remains below $90,000 per QALY, indicating that the intervention is cost-effective.

Our findings show that investing in stable housing for this marginalized population, even with no requirement to enter OUD treatment, would not only save lives and improve health outcomes but would also be cost-effective. Efforts are urgently needed to improve the health of PEH with OUD. It is essential to understand the effectiveness and cost-effectiveness of housing provision for this marginalized population, as housing status is a key social determinant of health.

2 - Optimal Timing and Choice of Vascular Access Placement in Advanced Ckd: a Simulation-Informed Markov Decision Process Framework

Yiwen Cao, University of Southern California, Los Angeles, CA, United States, Sze-chuan Suen

Vascular access planning for patients with advanced chronic kidney disease (CKD) is a sequential decision problem under uncertainty, requiring clinicians to balance the risks of premature intervention against the consequences of delayed access, particularly catheter dependence at dialysis initiation. Existing clinical guidelines rely largely on static recommendations and do not explicitly account for patient heterogeneity or the stochastic nature of disease progression and access outcomes. In this study, we develop a data-driven decision framework that integrates simulation-based disease modeling with a Markov decision process (MDP) to determine the optimal timing and type of vascular access placement in pre-hemodialysis patients.

We first estimate stage-specific hazard functions for CKD progression and mortality using longitudinal trajectories generated from a validated microsimulation model. These estimates are translated into a continuous transition kernel and embedded within a finite-horizon MDP that captures both the disease stage and vascular access states. This formulation enables direct optimization of clinical decisions without repeated simulation of various policies.

Our results indicate that the optimal policy exhibits a threshold structure across disease stage, age, and frailty. Delayed intervention is preferred in early CKD stages, whereas immediate access placement becomes optimal in advanced stages and older ages. At dialysis initiation, arteriovenous graft (AVG) placement is consistently preferred in older ages due to its rapid usability under limited planning horizons. Frailty and access site availability substantially influence decision timing.

These findings demonstrate the value of embedding simulation-derived dynamics within a structured MDP framework and provide actionable insights for personalized vascular access planning.

3 - Health and Economic Impacts of Stable Housing Provision for Older Adults: a Model-Based Analysis

Varun Gande, Stanford University, Stanford, CA, United States, Jeremy Goldhaber-Fiebert, Margaret Brandeau

The population of people experiencing homelessness (PEH) in the United States is aging, with 20% aged 55 years or older, and this group experiences substantial morbidity, mortality, and high acute health care costs. This study assessed the health outcomes and cost-effectiveness of providing stable housing to older PEH compared with no housing provision. Using a model-based cost-effectiveness analysis, we simulated two cohorts of 1,000 PEH: older adults (55 years or older) and younger adults (18–54 years). The intervention consisted of providing stable housing without requiring engagement in opioid use disorder treatment. Outcomes included five-year overdoses and deaths, lifetime per-person discounted quality-adjusted life-years (QALYs) and costs, and incremental cost-effectiveness ratios (ICERs) relative to the status quo.

For older PEH, the housing intervention reduced five-year deaths from 218 to 184 per 1,000 individuals, increased lifetime QALYs from 5.64 to 7.52, and increased per-person costs from $308,598 to $460,673, resulting in an ICER of $80,700 per QALY gained. For younger PEH, deaths declined from 72 to 62 per 1,000 individuals, lifetime QALYs increased from 9.60 to 12.13, and costs rose from $391,976 to $530,492, yielding an ICER of $54,800 per QALY gained. Sensitivity analyses showed that the housing intervention for older PEH consistently cost less than $100,000 per QALY gained.

Providing stable housing for PEH saves lives, improves health outcomes, and is cost-effective across age groups. As the homeless population continues to age, expanding permanent supportive housing for older PEH represents an urgent national priority with strong humanitarian, clinical, and economic justification.

4 - Incentives for Local Vaccine Manufacturing In Africa: Cournot Competition with Correlated Yield Uncertainty

Olumurejiwa Fatunde-Iloeje, Georgetown University, Washington, DC, United States, Isabelle Rao, Gonzalo Romero, Prashant Yadav

Countries without local vaccine manufacturing capacity are more vulnerable to supply disruption, as demonstrated by delayed and insufficient access to health products during public health catastrophes such as the COVID-19 pandemic. Sub-Saharan Africa is particularly vulnerable; in 2023, the region produced less than 1% of the vaccines it procured. As such, there is significant scope for expanding local production. Policymakers have set a goal of increasing Africa's share of locally-produced vaccines to 60% by 2040.

Vaccine production includes early-stage activities that convert raw materials to bulk antigen and late-stage activities (often referred to collectively as "fill-and-finish") that convert the bulk product into a finished good. A manufacturer's local production strategy can 1) focus solely on late-stage production or 2) feature integrated production, i.e., both early-stage and late-stage activities. In addition to having distinct cost and risk

implications, these two strategies differ in their levels of sourcing and process yield uncertainty. Yield uncertainty adds complexity to vaccine supply chains, and in some cases reverses established results prescribing how governments should intervene through contracts or other mechanisms.

In this paper, we generalize the model of Cournot competition under yield uncertainty from Deo and Corbett (2009) to account for correlated yield uncertainty among integrated and late-stage manufacturers due to market concentration among upstream suppliers. We show that highly correlated yield uncertainty reduces the number of manufacturers that enter the market, the expected quantity produced, and expected social welfare. Our results inform the design of practical market interventions for this important problem.

TC08

305A

Modeling Approaches to Address Health Disparities

Invited Session

Health Opportunities for All

Chair: Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States

1 - Probabilistic Network Modeling of Early Detection and Control for Cardiovascular Disease Risk in Type 2 Diabetes

Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States, Yukti Kathuria, William J. Gallagher, Elizabeth B. Selden, Kristen Miller

Patients newly diagnosed with type 2 diabetes (T2D) are at increased risk of developing cardiovascular disease (CVD), the leading cause of morbidity and mortality in this population. Early detection and attainment of glycemic control within the first year after diagnosis reduce the risk of CVD after the T2D diagnosis. Yet gaps persist regarding how to: (i) operationalize early detection of T2D using Electronic Health Records (EHR) data, and (ii) estimate the impact of diagnostic delay on subsequent CVD risk considering longitudinal patient-level characteristics. We propose a novel probabilistic network model to analyze the interdependencies between early detection of T2D (represented by the abnormality of the diagnostic lab values), indicators of post-diagnosis glycemic control and CVD diagnosis. Using a temporally structured discrete Bayesian network learned from longitudinal individual-level EHR data of patients (N=9,450) who received primary care services in ten primary care clinics between 2017 and 2023, we quantify the probabilistic dependencies between demographics, diagnostic delay surrogates, post-diagnosis glycemic control and occurrence of CVD after the T2D diagnosis. Results highlight that age at diagnosis is the strongest determinant of long-term CVD risk, while early detection and early glycemic control contribute modest but measurable changes in predicted risk across clinically defined scenarios. Scenario-based probabilistic inference further demonstrates how combinations of diagnostic timing and early glycemic trajectories influence downstream cardiovascular outcomes. The proposed framework achieves predictive performance comparable to conventional machine learning models while providing interpretable representations of temporal clinical relationships, supporting its potential utility for exploratory analysis and clinical decision support.

2 - Mechanistic Modeling Social Conditions in Disease Simulations for Disease Prediction and Intervention Analysis

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States, Amir Khosheghbal

Epidemic models typically simulate the spread of an infectious disease as a function of disease risk behaviors. However, research has shown that multilevel factors such as poverty, housing instability, food insecurity, substance use disorder, neighborhood context, etc., can influence risk behaviors and effectiveness of intervention response, and are thus typically referred to as social determinants of health (SDH). There is increasing awareness for the need to incorporate SDH into epidemic simulation models to improve epidemic forecasting and for joint analysis of structural (SDH-related) and behavioral (disease-related risk behaviors) interventions.

However, the multivariate joint associations between SDH and behaviors, or the causal influence of SDH on intervention effects, needed for such modeling are not available and are difficult to estimate through typical approach of controlled studies. Taking sexually transmitted infections as a case study, we present preliminary work on developing algorithms that integrate copula-theory with graphical models to learn associations or casual structures. These methodologies can help provide a framework for evaluating feature associations or intervention effects when data, specifically SDH data, comes from disparate datasets.

3 - Understanding of Drivers' and Public Transit Geographic Accessibility to Buprenorphine for Opioid Use Disorder Treatment Using Spatial Econometrics

Carolina Vivas-Valencia, The University of Texas at San Antonio, San Antonio, TX, United States, Sara Abu-Aridah, Lisa Losada-Rojas

Opioid use disorder remains a US public health concern requiring better medication access. This study quantifies geographic accessibility to buprenorphine prescribers in Texas using travel-time metrics for driving and public transit across five major metropolitan areas. Using a Texas outpatient dataset (2018-2023) and buprenorphine prescriber data from SAMHSA, we included records of opioid overdoses for both patient reasons of visit as the principal diagnosis. Contextual factors, including age, gender, race and ethnicity, and the Social Vulnerability Index (SVI), were analyzed. Accessibility was assessed using a binary indicator based on travel time, followed by paired t-tests to compare travel times between driving and public transit. Logistic regression was used to evaluate the influence of contextual factors on accessibility, followed by bivariate Moran's I and a spatial lag regression model to examine the relationship between travel time to the nearest prescriber and overdose counts at the zip code level. Our findings reveal significant variations in travel times and accessibility across transportation modes, demographics, and contextual factors. Specifically, females and older adults (65+) had lower odds of accessibility, while areas with high SVI had higher accessibility rates, when using CDC's SVI measure. Geographic accessibility maps highlight the unique challenges rural populations face in reaching prescribers. This study shows the critical need for targeted interventions to enhance accessibility to buprenorphine treatment, particularly for vulnerable populations, and suggests that transportation mode plays a vital role in accessing lifesaving treatment for opioid use disorder in Texas.

4 - Doing More with Less: Automation, AI, and Scope Optimization in Rural Value-Based Care

Jenn Block, DJB Consulting, Columbia, TN, United States

Rural primary care is facing a workforce crisis that no hiring bonus can fully solve. As hospitals and health systems continue to outbid independent and small-group practices for clinical talent, rural providers are left navigating an impossible equation: deliver value-based care outcomes with the staff they have or fall short of the quality benchmarks that determine their financial survival.

This session reframes the staffing challenge as a design problem — and offers a practical playbook for rural PCPs ready to think differently. Drawing on real-world experience scaling value-based care programs across rural Tennessee and East Texas, Dr. Jenn Block will explore how forward-thinking practices are closing the gap through two parallel strategies: maximizing clinical scope within existing teams, and deploying AI and automation to eliminate the repetitive, time-consuming tasks that dwindle staff capacity without adding clinical value.

Attendees will leave with actionable frameworks for identifying automation opportunities in their own workflows, strategies for expanding scope without expanding headcount, and a clear-eyed perspective on where AI can — and cannot — replace human judgment in rural care delivery.

KEY TAKEAWAYS

• How rural practices can audit clinical workflows to identify high-value automation targets

• Scope optimization strategies that increase capacity without adding FTEs

• Practical AI tools already being used in value-based care settings — and how to evaluate them

• The human elements that must remain human: where technology ends and relationship-driven care begins

TC09

305B

Simulation-Optimization for Intelligent Healthcare Systems

Invited Session

Simulation Society

Chair: Xueping Li, University of Tennessee, Knoxville, Knoxville, TN, United States

1 - AI-Enhanced Multi-Method Simulation for Intensive Care Unit Digital Twinning

Xiang Zhong, Univ. of Florida, Gainesville, FL, United States, Siddharth Vipankumar Abrol, Grace Yao Hou, Amos Lal, Ognjen Gajic, Yue Dong

The growing prevalence of critical illness presents an urgent public health challenge, exacerbated by the aging population and associated comorbidities. This has led to an increased demand for intensive care, alongside rising complexity in managing critically ill patients and ongoing strain on the healthcare system with limited monetary and personnel resources. This research aims to develop novel multi-method simulation, hybridizing discrete event and agent-based models with AI-informed agents for supporting operational decision making in the intensive care unit (ICU) setting. We develop an AI pipeline to cluster, model, and predict clinical trajectory phenotypes a critically ill patient should follow. We then integrate these models in the simulation environment, allowing for runtime evaluation of AI models to drive simulated agent behaviour. The data-driven patient trajectory modelling and prediction enables us to capture the interplay between patients and clinicians, thereby providing a refined assessment of clinician workload and ICU resource utilization. A case study is developed to demonstrate the effectiveness of the proposed approach, and we envision the model supports digital twin deployment when granted live access to patient and system data.

2 - Bidirectionally Coupled Sd&Ndash;Abm with AI-Persona Agents for Clinically Grounded Opioid Policy Simulation

Gerald Jones, University of Tennessee-Knoxville, Knoxville, TN, United States, Emma Wilhoit, Thomas Berg, Xueping Li

Opioid use disorder (OUD) is driven by nonlinear interactions among substance use, treatment access, relapse, and overdose mortality. Policy evaluation in this domain requires models that are both empirically calibrated at the population level and capable of representing heterogeneous individual behavior.

This talk presents a bidirectionally coupled system dynamics–agent-based (SD–ABM) framework for opioid policy analysis, calibrated using heuristic optimization and extended with AI-driven behavioral agents.

At the population level, a system dynamics model is calibrated to state-level data on opioid prevalence, treatment participation, overdose incidence, and mortality. Genetic algorithms, Monte Carlo sampling, and simulated annealing are used to estimate high-dimensional transition rates, producing robust convergence across overdose and treatment trajectories.

The model introduces a two-way SD–ABM coupling. Agents use environmental context derived from system-level population stocks and calibrated transition probabilities to inform decisions. In turn, aggregated agent actions dynamically modify SD flow rates during simulation, enabling cross-scale behavioral feedback.

To evaluate behavioral realism, an AI-persona agent architecture is compared against a stochastic greedy heuristic baseline. Results show that AI-persona agents produce behavioral patterns more closely aligned with clinical expectations described in DSM-5, SAMHSA, and NIDA literature, particularly in relapse, usage, and treatment dynamics and treatment engagement.

This hybrid SD–ABM–AI framework demonstrates how heuristic calibration, cross-scale feedback, and

interpretable AI agents can be integrated to support clinically grounded, policy-relevant simulation in complex healthcare systems.

3 - Robust Return-to-Play Decisions for Concussion under Parameter Uncertainty

Dongjoon Lee, Georgia Institute of Technology, Atlanta, GA, United States, Eunhye Song, Gian-Gabriel Garcia

Determining an appropriate symptom-free waiting period (SFWP) for safe return-to-play (RTP) decisions is challenging because concussion recovery is difficult to observe directly. We study this problem using a Hidden Markov Model (HMM) simulator that represents latent concussion severity states and observed multimodal assessments. The goal is to identify SFWPs such that the probability of post-RTP risk (e.g., recurrent concussion) is below a threshold. However, because the simulator depends on the estimated parameters, their estimation error can substantially affect the decision.

To address this, we first estimate the sampling distribution of the HMM parameters by combining bootstrap and the expectation-minimization algorithm. Then, we identify the set of robust SFWPs whose probabilities of acceptable post-RTP risks are within a threshold under the sampling distribution.

This problem is formulated as nested superlevel set estimation; the inner problem at each SFWP finds the set of parameters that produce acceptable post-RTP risk probabilities, while the outer level identifies the set of robust SFWPs. We propose a Bayesian sequential sampling algorithm, which models the probability of post-RTP risk as a function of SFWP and HMM parameters with a Gaussian process and samples the next SFWP-parameter pair by maximizing the expected changes in the set estimators. We further accelerate the procedure by exploiting the monotonicity between SFWPs and post-RTP risk probability. We prove that the algorithm terminates in finite time with a correct classification guarantee almost surely.

This framework applies to a broader class of healthcare problems to find policies robust to parameter uncertainty.

4 - On-Device Decentralized Federated Learning for Intelligent Healthcare Systems: Current Landscape and Future Directions

Xueping Li, University of Tennessee, Knoxville, Knoxville, TN, United States, Anam Nawaz Khan

The rapid proliferation of intelligent edge devices in clinical environments, from bedside monitors to wearable sensors and point-of-care diagnostic tools, is reshaping how AI can be delivered at the point of care. Yet the promise of real-time, privacy-preserving clinical intelligence remains constrained by a fundamental tension: how do we enable collaborative learning across distributed healthcare settings without centralizing sensitive patient data?This presentation examines the current landscape and future directions of on-device decentralized federated learning (DFL) as a framework for intelligent healthcare systems. We discuss how moving AI computation directly onto edge devices, eliminating reliance on central servers, addresses critical challenges in clinical deployments including data privacy, regulatory compliance, communication bottlenecks, and equitable access across resource-constrained settings such as rural clinics and underserved communities. We further explore emerging opportunities at the intersection of on-device AI and healthcare: adaptive model personalization across heterogeneous patient populations, continuous learning from real-world clinical data streams, and resilient multi-site deployment without compromising HIPAA compliance. Looking ahead, we outline a research agenda for simulation-driven optimization of decentralized clinical AI systems, identifying open challenges in device heterogeneity, model robustness, and scalable edge deployment that must be addressed to realize the full potential of intelligent, distributed healthcare AI.

5 - Critical Care Predicted Occupancy Model (Ccpod) - a Tool to Inform Decision Making to Avoid Critical Care Bed Alerts

Weihang (Susana) Su, University of Toronto, Toronto, ON, Canada, Kali Barrett, Terra Lerasts, Michael Carter

Critical care bed shortages impose major operational pressures in large academic hospitals, often resulting in surgical cancellations, delayed transfers, and repeated capacity alerts. At the University Health Network (UHN), critical care bed alerts are issued when intensive care demand exceeds staffed bed availability, requiring rapid escalation and prioritization of admissions. Current processes are often reactive, with limited ability to anticipate near-term occupancy pressures before they become operationally critical.

This study presents the Critical Care Predicted Occupancy Model (CCPOD), a decision-support framework that integrates predictive analytics with discrete-event simulation to support more proactive ICU capacity management. Using retrospective patient-level and operational data, the framework is designed to forecast short-term ICU occupancy, identify periods of elevated capacity risk, and provide advance insight into potential alert conditions under uncertainty.

The model captures major system dynamics relevant to critical care capacity planning, including multiple admission streams, heterogeneous lengths of stay, discharge processes, and operational constraints. Beyond forecasting future occupancy levels, the framework is intended to evaluate the timing and potential impact of operational responses such as staffing adjustments, admission prioritization, and surgical scheduling changes.

Model performance will be assessed through retrospective validation based on its ability to reproduce historical occupancy patterns and anticipate bed alert periods with meaningful lead time. By combining predictive modeling with simulation-based scenario analysis, CCPOD aims to shift ICU bed alert management from reactive response toward proactive, data-informed decision-making. This work contributes to healthcare operations management by demonstrating how integrated predictive-simulation approaches can strengthen hospital resilience under capacity strain.

TC10

306A

Falling Through the Cracks? Preventing Gaps in Care Using Data

Contributed Session

Data Analytics

Chair: Richard Dixon, Virginia Commonwealth University, 301 Virginia St, Richmond, VA, 23219, United States

1 - Directed Conformal Prediction: Decision-Controlled Precision in ICU Resource Allocation

Chandra Dronavajjala, University of Connecticut, Harford, CT, United States

A physician evaluating a liver disease therapy must assess effects on bilirubin, albumin, alkaline phosphatase, and four other endpoints simultaneously. She needs the uncertainty region to be tightest on bilirubin her primary marker of disease progression even if that means wider bounds on secondary endpoints. Current methods cannot accommodate this. Bonferroni correction divides the error budget equally across outcomes. Ellipsoidal methods adapt to correlations but allocate precision according to the covariance structure, not clinical priorities. In neither case can the physician direct where precision concentrates.

We propose Directed Conformal Regions (DCR), a framework built on conformal prediction a distribution-free inference methodology that provides finite-sample coverage guarantees without parametric assumptions. DCR gives clinicians explicit control over which outcomes receive the tightest bounds. A hierarchy constraint encodes clinical priority primary, secondary, and safety endpoints and a single steering parameter governs how aggressively precision concentrates on the primary endpoint at the cost of wider bounds elsewhere. Joint coverage is guaranteed at any steering level without retuning.

We validate DCR on the ACTG-175 HIV antiretroviral trial (3 outcomes), the NHEFS smoking cessation study (6 outcomes), and the Mayo Clinic PBC liver disease trial (7 outcomes). On PBC, directing precision toward bilirubin cuts its prediction interval to half of Bonferroni's width and one-quarter tighter than the best ellipsoidal alternative, while maintaining 93% joint coverage. Across all datasets, switching which endpoint is designated

primary successfully redirects precision to that endpoint. We characterize when directed precision is most beneficial and provide guidance for multi-outcome treatment evaluation.

2 - Overcoming 'Signal Washing': a Precision Public Health Framework to Identify Subpopulation Variation in Smoking Cessation

Diana Lopez-Soto, NORTH DAKOTA STATE UNIVERSITY, Fargo, ND, United States, Luis Alcala Becerra

Tobacco use costs the U.S. over $600 billion annually and claims more than 480,000 lives. In North Dakota, 13.3% of adults smoke, incurring $379 million in annual costs. Despite programs like NDQuits, cessation remains a challenge; while over two-thirds of the 49 million U.S. adults who smoke in 2022 attempted to stop, only 8.8% succeeded. Although various models have been developed to identify cessation influential factors, traditional predictive approaches often overlook individual variability, leading to general interventions that diminish efficiency for specific subpopulations.

This study implements a precision public health framework to move beyond generalized approaches. By applying unsupervised clustering to the NDQuits telephonic dataset, we captured subpopulation-level variation in cessation outcomes. We then analyzed success drivers for each group using logical rules and SHAP analysis. Key predictors; including age, urgency to smoke, and daily cigarette consumption, varied significantly across clusters.

Our Clustered XGBoost model achieved 5% improvement in AUC/AUPRC over global baselines. In precision health context, this shows that global models suffer from “signal washing” where majority trends drown out the distinct characteristics of smaller subgroups. By segmenting the population, we reveal hidden data patterns that global models cannot detect, providing a granular foundation for personalized intervention strategies that enhance public health efficacy.

3 - Modeling Follow-up Care Disengagement in Childhood Cancer Survivors with Zero-Inflated Survival Methods

Rachel Bennett, University of Oklahoma, Oklahoma City, OK, United States, Amanda Janitz, David Noyd, Talayeh Razzaghi

Long-term follow-up care is essential for childhood cancer survivors, yet many experience extended gaps in care or do not return after treatment. These patterns create a challenging zero-inflated time-to-event setting in which some survivors never engage in follow-up, while others disengage over time. In this study, we examine two-year gaps in survivorship care among childhood cancer survivors in Oklahoma using a zero-inflated survival modeling framework designed to capture both processes. The proposed approach combines a machine learning component to identify survivors unlikely to engage in follow-up with survival modeling methods to estimate time to disengagement among those who do. By accounting for both non-engagement and later dropout, the framework better reflects the range of follow-up care trajectories among survivors.

4 - Understanding the Needs of Dementia Caregivers

Lina Zhou, The University of North Carolina at Charlotte, Charlotte, NC, United States, Bikramjit Roy, Brendan Zhou, Ethan Zhang

With global dementia prevalence rising every three seconds and economic burdens projected to reach $2.8 trillion by 2030, care models are shifting from residential facilities to home-based, person-centered care. This transition places immense pressure on family caregivers who often lack formal training and face high rates of burnout, emotional distress, and financial crisis. While social media has emerged as a vital resource for peer support, the specific needs expressed within these digital communities remain understudied.

This study utilizes inductive coding to analyze caregiver discourse on Reddit. Through multiple rounds of sample data annotation, we developed a refined codebook to characterize support-seeking behavior. Our analysis of annotated data shows that female persons living with dementia were more commonly supported than their

male counterparts. Informational support was the most frequently sought category. Practical needs ranged from seeking effective communication strategies to finding adaptive equipment.

Qualitative findings highlight a profound sense of isolation and a reliance on personal family care, with minimal mention of assisted living or professional services. These findings provide a nuanced understanding of the real-time challenges faced by caregivers, offering critical insights for developing targeted digital interventions, educational resources, and policy frameworks to improve the quality of home-based dementia care.

5 - Formal Feature Attribution as a Causal Calculator For Drug-Drug Interaction Risk in Geriatric Polypharmacy: Evidence from a State-Level All-Payer Claims Database

Richard Dixon, Virginia Commonwealth University, Richmond, VA, United States

Background: Polypharmacy (≥5 concurrent medications) affects more than 40% of adults aged ≥65 years and drives adverse drug events (ADEs) responsible for an estimated 100,000 hospitalizations annually. Pairwise DDI databases flag co-prescribed pairs in isolation, missing synergistic multi-drug effects where each constituent pair appears benign.

Methods: Using Virginia's All-Payer Claims Database (APCD; 6,929,576 patients; 2016–2019), we constructed a polypharmacy cohort (n = 1,182 geriatric cases; 65–114 years; 5:1 matched controls). A 30-day causality window isolated proximal drug exposures. A per-event-density ensemble (CatBoost/XGBoost/XGBoost-RF; Optuna-tuned; training: 2016–2018; holdout: 2019) fed Formal Feature Attribution (FFA) extended to multi-feature interaction testing of drug pairs and triplets, with Intervention Rate (IR) scores quantifying expected risk reduction per deprescribing action. FP-Growth network visualization contextualized causal findings within drug co-occurrence topology.

Results: All 8 age bands achieved PR-AUC ≥ 0.908 (peak: 0.997 at 75–84 band). FFA identified 115 synergistic drug pairs and 5,021 high-risk triplets invisible to pairwise DDI databases. Levofloxacin acted as a pharmacological hub (CYP1A2 inhibition), appearing in 4 of the top 5 synergistic pairs: acetaminophen + levofloxacin (IE = 16.3) and levofloxacin + lorazepam (IE = 11.9). Top IR-ranked deprescribing targets: simvastatin (IR 7.0×10^{-4}), furosemide (IR 2.0×10^{-4}), alprazolam (IR 1.0×10^{-4}). A U-shaped Z-code monitoring effect identified a fragmented-care phenotype (Q4; OR ≈ 0.94 vs. unmonitored) distinct from protectively monitored patients (Q2; OR = 0.25).

Conclusions: FFA multi-feature interaction analysis constitutes a causal calculator for polypharmacy ADE risk, detecting synergistic combinations invisible to pairwise DDI databases and yielding IR-ranked deprescribing priorities from standard claims data.

TC11

306B

AI for Health Informatics

Invited Session

Healthcare Technology and Bioinformatics

Chair: Karen Hicklin, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY, 10029, United States

Co-Chair: Jaber Qezelbash-Chamak, University of Florida, Gainesville, FL, United States

1 - Artificial Intelligence Based Conversational Assistant: Development and Validation for Emergency Triage and Postpartum Support in High-Risk Mothers

Gokalp Cosgun, Wayne State University, Detroit, MI, United States, Zhongjie Shi, Sidhartha Tan, Melike Yildirim

Doulas can reduce maternal mortality, breastfeeding problems, postpartum depression, and newborn problems. Postpartum morbidity and mortality disproportionately affect underserved populations, including adolescent mothers in urban centers like Detroit, where access to doula care and traditional support systems is limited by socioeconomic barriers. While digital chatbots offer a scalable solution, their clinical utility is contingent upon proven accuracy in triage, the capacity for empathetic engagement, support for behavioral change, and reliable safety protocols.

Our objective was to develop an AI-based conversational assistant (AI-CA) to support postpartum mothers. We hypothesized that the AI-CA would identify maternal or infant symptoms requiring immediate medical attention. Our goal was to provide a platform offering understanding, empathy, reliability, and security.

Using established postpartum reference materials, we constructed a curated question–answer dataset covering postpartum recovery, newborn care, and maternal mental health. Each example was labeled by topic and clinical urgency (routine, caution, emergency). A large language model was fine-tuned to generate domain-specific responses. In parallel, two severity-classification models were developed to independently assess urgency in user inputs and generated outputs. When emergency-level conditions were detected, the system replaced the response with a standardized emergency message.

Preliminary evaluation demonstrated that the model generated semantically aligned guidance and that the safety layer identified high-risk scenarios, including self-harm statements and serious postpartum symptoms, during testing.

These findings support the feasibility of a layered AI framework that integrates generative support with safety-aware triage. Such systems may help extend evidence-based postpartum guidance and promote timely care-seeking among high-risk mothers.

2 - Kanbalance: Kolmogorov&Mdash;Arnold Network Mitigates Class Imbalance

Jaber Qezelbash-Chamak, University of Florida, Gainesville, FL, United States, Karen Hicklin, Minhee Kim

Addressing class imbalance is especially important in high-stakes biomedical machine learning, where underrepresentation may arise not only from clinical rarity but also from structural inequities in healthcare access, such as limited affordability of care or geographic barriers to specialty services. We introduce KANBalance, a method that combines Kolmogorov—Arnold Networks (KANs) with Focal Loss to improve learning under skewed class distributions. By replacing fixed neural activations with trainable univariate spline functions, KANs enable localized feature transformations that better capture subtle patterns in minority subgroups, while Focal Loss emphasizes hard-to-classify and underrepresented samples. We evaluate KANBalance on two imbalanced medical imaging benchmarks: pediatric chest X-rays for pneumonia detection and brain MRIs for tumor classification, with subgroup-aware analyses to assess performance under representation gaps. Across both tasks, KANBalance outperforms common resampling baselines, including SMOTE and ADASYN, in accuracy, F1 score, and AUC. Ablation studies further show that the combination of spline-based adaptation and minority-focused loss is key to these gains, highlighting KANBalance as a practical approach for more equitable minority detection in real-world healthcare data.

3 - Comparing the Algorithmic Fidelity of Large Language Models in Predicting Human Decision Making: a Case Study of Vaccination Choice

Shaochong Xu, Johns Hopkins University, Baltimore, MD, United States, Liyue Zhang, Amelia Jamison, Samee Saiyed, Abe Hou, Hongru Du, Lauren Gardner

This study explores the algorithmic fidelity of four well-established LLM architectures when attempting to replicate complex health-decisions like vaccination. The analysis leverages individual survey responses and real-world online news content within prescribed model designs that vary in their relative inputs - namely, basic demographic information, revealed attitudes based on survey data, and personalized media diets. We evaluate the predictive accuracy and sensitivity for each model across all four LLMs, and reveal differing biases inherent to each LLM design. We further explore the role of media exposure on predicted vaccination choice within a counterfactual scenario analysis, by varying the ratio of authoritative to low credibility content. The results indicate that the LLM architectures vary substantially in their prior assumptions of human behavior, as well as their sensitivity to curated media exposure. Specifically, three of the four LLMs illustrate a clear pro-science alignment, as evidenced by their predicted pro-vaccination behavior. These findings carry critical implications for future efforts that strive to build LLMs as reliable tools for modeling individual decisions and population-level health outcomes in complex environments.

TC12

306C

Practical AI #1: Building and Deploying Digital Twins in Healthcare

Invited Session

Healthcare Practice

Chair: Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, NY, United States

1 - Discrete Event Simulation for Infusion Scheduling at the Duke Cancer Institute

Manos Giannopoulos, Duke University, Durham, NC, United States, Yi Shen, Michael Zavlanos

Discrete event simulation offers a principled way to model the stochastic, multi-resource dynamics of infusion center operations without disrupting live clinical workflows. We present a Discrete Event Simulator built for the Duke Cancer Institute that captures the full outpatient flow across Pre-screening, Clinical Evaluation, Medication Preparation, and Infusion. The Simulator operates in two modes: a data-driven mode that samples stage durations from historical distributions to produce accurate statistics for any given schedule, and a scenario-testing mode that allows schedulers to manually introduce delays, reallocate resources, or simulate counterfactual events to stress-test schedule robustness. Together, these modes transform the Simulator into a practical decision-support tool for schedulers, enabling both rigorous schedule evaluation and rapid response to unexpected disruptions.

2 - Application of Discrete-Event Simulation for Capacity Allocation Decisions in the Inpatient Setting at A Hospital Specialized in Cancer Care

Anup Mokashi, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Effective management of resources such as inpatient beds in a high-census urban hospital environment can present several challenges to hospital administrators and clinical leaders. When making capacity allocation changes, it is often necessary to weigh the impact of proposed changes on patient care needs, overall patient experience, clinical requirements for various hospital departments, resource characteristics, and staffing considerations. It is also vital to assess the impact of variation in lengths of stay, specialized care needs, as well as the variability in admission-and discharge patterns in order to accurately capture the demand placed by distinct patient cohorts on limited hospital resources.

We present a discrete-event simulation model that has served as a strategic decision support tool for inpatient bed management at Memorial Sloan Kettering Cancer Center (MSKCC) for several years. We describe some key assumptions and considerations made during model development process based on the unique requirements at MSKCC. We demonstrate an example scenario to showcase how the model is used in practice by hospital leadership, to assist with periodic capacity allocation changes.

3 - Digital Twin to Improve Operations at a Mothers Milk Bank

Natalia Summerville, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Human milk banks operate as highly specialized, safety-critical supply chains, yet their complexity is not widely understood outside of those directly involved. From the moment milk arrives to when it is processed, stored, and prepared for shipment, operations must balance strict handling requirements, batch constraints, and fluctuating demand from neonatal intensive care units (NICUs).

This presentation describes building and implementing a discrete event simulation model, as a practical tool to help milk bank stakeholders evaluate strategic scenarios and improve throughput and distribution. The design work centered on structured working sessions with operational teams, using the model to explore "what-if" scenarios related to scheduling, resource allocation, and prioritization.

A key challenge—and focus of this work—was translating real-world operations into a usable model. Data were often incomplete or inconsistent, and much of the process knowledge resided with frontline staff rather than in formal documentation. Iterative engagement with the milk bank team, including walkthroughs and validation sessions, was essential to accurately reflect workflows and build trust in the model outputs.

These facilitated sessions enabled stakeholders to better understand system constraints, identify bottlenecks, and align on strategies to improve reliability and equitable distribution to NICUs.

This presentation highlights both the operational complexity of milk banks and the value of using simulation as a structured, collaborative tool to support informed decision-making in a domain that is often overlooked but critically important.

4 - Obsessed with Failure: Driving the Hro Journey Through Data Analytics from Standardized Mock Code Training Across the Veterans Health Administration

Debra Mosley, Veterans Health Administration: SimLEARN, Orlando, FL, United States, Payton Muehlefeld, Tracey Robilotto

The Veterans Health Administration (VHA) is the largest integrated healthcare system in the United States providing care at 1,380 healthcare facilities. Over the past decade, survival from in-hospital cardiac arrest (IHCA) has remained unfavorable at 18-20%. Because real code-events are a low volume high risk event, simulation provides an excellent opportunity to practice and hone skills in a safe environment. In 2020, the VHA's Resuscitation Education and Innovation (REdI) team embarked on a journey to develop and scale a standardized approach to mock code training and data collection. Because the VHA facilities range in complexity, developing a standardized program that would meet varied needs was challenging. Built on a foundation of HRO principles, this hybrid training approach brings multiple VHA facilities together to collaborate in a virtual classroom and then a face-to-face follow-up at the individual facility. What sets this training program apart from others, is the inclusion of a modified Healthcare Failure Mode and Effect Analysis (HFMEA) Hazard Scoring Matrix to evaluate the systems of care. Utilizing this tool, clinicians can rate contributive factors for potential risk and provide recommendations for mitigation. Currently, more than 50% of VHA facilities have implemented the program and data collected from their local training events is being entered into a national database. Data from more than 500 Mock Code events have been collected and analyzed. This presentation will highlight the top five national latent safety risks and how this is being operationalized for quality improvement on both the facility level and enterprise wide.

TC13

307

Digital Health and Analytics for Care Delivery and Community Health

Invited Session

eBusiness

Chair: Xiaojin Liu, Virginia Commonwealth University, Richmond, VA, United States

1 - Consistent Care: Counterfactual AI for Data-Driven Treatment Recommendations

Deepika Gopukumar, Saint Louis University, Wildwood, MO, United States, Nirup Menon, Sriram Somanchi, Martin Schoen

Optimizing patient outcomes requires not only effective treatment selection but also consistency in recommendations among patients with similar predicted benefits. Conventional approaches to treatment assignment prioritize average outcome maximization, often leading to inconsistent decisions across comparable individuals. This study proposes a counterfactual AI framework to estimate individualized treatment effects using both real-world and simulated patient data. Simulated data are used to validate the methodology under controlled conditions, ensuring robustness. We construct policy trees to generate actionable treatment recommendations, where an unconstrained policy tree optimizes expected outcomes based on individual benefit estimates. To evaluate consistency, we introduce a normalized individual-level consistency metric that captures variation in recommendations among similar patients. Results from both simulated and real-world datasets demonstrate that the proposed approach achieves high consistency without compromising outcome optimization. This framework provides a scalable and interpretable solution for designing clinical decision policies that balance effectiveness with consistent, reliable, and ethically aligned treatment recommendations.

2 - Evaluating the Impact of Telemedicine Services on Community Health: a County-Level Analysis

Xiaojin Liu, Virginia Commonwealth University, Richmond, VA, United States, Susan Goldstein, Kingshuk Sinha, Karen Soderberg

We propose and empirically analyze a framework for understanding the impact of telemedicine services on health outcomes at scale across communities in the United States. Our analysis of county-level panel data reveals that increased availability of telemedicine services improves community health that is central to the wellbeing of a community. We evaluate the benefits of telemedicine services in aggregate – i.e., overall health outcomes – and along specific dimensions of reductions in premature death rate, low birthweight rate, preventable hospital stays, smoking rate, and the need for diabetes monitoring. In addition, we find that greater availability of telemedicine services is associated with decreased COVID-19 related mortality. Heterogeneity analyses further show that lower socio-economic status reduces the effectiveness of telemedicine services while superior digital infrastructure and greater innovation capacity enhance its effectiveness. Demographics by way of higher proportions of Black and female community members enhance the effectiveness of telemedicine services whereas a higher proportion of older adults reduces its effectiveness. Taken together, the proposed framework and the findings underscore the potential of telemedicine services to improve community health while recognizing the moderating effects of contextual factors. This study makes a significant contribution towards advancing the literature on seamless coordination of digital service innovations to drive community-wide health benefits.

3 - Introducing a Novel Machine Learning Approach to Identify, Model, and Predict Breast Cancer Risk

Kasra Dolatkhahi, University of Memphis, Memphis, TN, United States, Pavankumar Mulgund

Early identification of breast cancer risk remains a critical challenge in preventive healthcare, particularly given the multifaceted nature of clinical, environmental, and behavioral risk factors. While recent advances in machine learning have improved predictive capabilities, many existing models rely on generalized datasets. To address this limitation, this study adopts a Design Science Research approach to develop and evaluate a predictive artifact for breast cancer risk assessment that integrates domain expertise, structured modeling, and machine learning techniques.

First, potential breast cancer risk factors were identified through systematic literature review and qualitative content analysis. Second, these factors were refined and validated through expert consultation with oncology and breast surgery specialists using a Delphi-based consensus approach, ensuring contextual relevance. Third, ISM was employed to construct a conceptual model that captures the interrelationships among the identified risk factors. Finally, the predictive artifact was operationalized using machine learning algorithms and evaluated using a dataset comprising medical records from 5,208 patients, resulting in 26 salient risk factors. Comparative evaluation of machine learning models demonstrated strong predictive performance, with the best-performing model achieving an accuracy of 96.12%. The findings indicate that integrating localized clinical knowledge with data-driven predictive modeling enhances the accuracy and contextual relevance of breast cancer risk prediction.

This study contributes to design science by demonstrating how context-sensitive predictive artifacts can be systematically designed through the integration of expert knowledge, structured modeling techniques, and machine learning methods. The proposed approach improves early risk assessment and supports more targeted prevention and intervention strategies in clinical practices.

Tuesday, July 28, 2026, 3:15 PM - 4:30 PM

TD01

301A

Machine Learning for Clinical Decision Support

Invited Session

AI in Healthcare

Chair: Nick Street, University of Iowa, Iowa City, IA, United States

Co-Chair: Anjolaoluwa Popoola, Tippie College of Business, University of Iowa, Iowa City, IA, United States

1 - Predicting Nephrotoxic Acute Kidney Injury in Hospitalized Adults Using Machine Learning

Yaohua Wang, The University of Iowa, Iowa City, IA, United States, Nick Street, Mary Vaughan-Sarrazin, Diana Jalal, Benjamin Griffin

Acute kidney injury (AKI) is a frequent and high-cost complication among hospitalized patients, often developing rapidly and associated with increased mortality, prolonged length of stay, and downstream chronic kidney disease. Early identification of patients at risk for imminent AKI remains a critical challenge for clinical decision support systems, particularly given the irregular, time-varying nature of inpatient electronic health record (EHR) data.

We develop a sequential deep learning framework to predict 24 to 48-hour AKI onset among hospitalized adults using routinely collected EHR data. The proposed model employs a Bidirectional Gated Recurrent Unit (Bi-GRU) architecture to capture both short-term fluctuations and longer temporal dependencies in laboratory values, vital signs, medications, and comorbidities. Unlike static models that rely on snapshot features, our approach leverages dynamic trajectories to generate continuously updated risk predictions throughout hospitalization.

Model performance is evaluated using patient-level cross-validation with discrimination (AUC, PR-AUC), calibration, and early-warning lead-time analyses. We compare the Bi-GRU against logistic regression baselines to assess the value of sequential modeling. Results demonstrate improved predictive accuracy and earlier risk detection using the Bi-GRU framework, highlighting the importance of temporal modeling in AKI prediction.

This work advances machine learning–driven clinical decision support by providing an operationally deployable framework for early AKI risk stratification, with potential applications in alert systems, nephrotoxin stewardship, and resource allocation within inpatient care settings.

2 - Early Prediction of Late Onset Preeclampsia Using Machine Learning Methods

Anjolaoluwa Popoola, Tippie College of Business, University of Iowa, Iowa City, IA, United States, Nick Street

Early prediction of preeclampsia from routinely collected electronic health record (EHR) data remains challenging due to its heterogeneous nature and presentation in patients, class imbalance, and irregular longitudinal observations. We present a methodological comparison of cross-sectional snapshot models and recurrent neural network architectures for the early prediction of late-onset preeclampsia (generally diagnosed at 34 weeks gestational age and above) using EHR data available as early as 13 weeks gestation from 20,396 pregnancies at a large Midwestern academic medical center. We evaluate XGBoost snapshot models at four clinically relevant gestational age cutoffs (13, 16, 20, and 24 weeks) and bidirectional Gated Recurrent Unit (GRU) networks trained on cutoff-matched sequences truncated at 13, 16, 20, and 24 weeks to enable early

prediction. Our results suggest that XGBoost achieves higher precision while GRU achieves higher recall, showing complementary strengths between these architectures. These findings indicate that early prediction of late-onset preeclampsia is feasible with routine prenatal data and suggest that more consistent data collection could improve predictive performance.

3 - Assessing Mental Health Treatment Pathways Prior to Suicidal Ideation Among Medicaid-Enrolled Youth

Amaya McNealey, Georgia Institute of Technology, Atlanta, GA, United States, Violet Cai, Daniel Kim, Jovan Julien, Nicoleta Serban

Suicidal ideation and suicide attempts among adolescents have risen sharply in the United States, yet many youths engage with the healthcare system well before a crisis occurs. Understanding how mental health treatment pathways differ prior to these outcomes may reveal opportunities for earlier intervention, especially for Medicaid-enrolled adolescents who face elevated risk and structural barriers to care.

We conduct a population-based matched case–control study using multi-state Medicaid claims data from 2016 to 2022. Adolescents diagnosed with suicidal ideation or suicide attempt are matched to clinically similar controls without suicidal diagnoses using an exact deterministic algorithm incorporating mental health diagnostic profiles, demographic factors, and healthcare access characteristics. For each individual, we analyze mental health service use during the twelve months preceding the index date, including outpatient psychosocial services, psychotropic medication use, emergency department visits, and transitions across care settings.

We compare treatment intensity, continuity of care, and reliance on emergency services between matched cohorts. We hypothesize that adolescents who later experience suicidal ideation or attempts demonstrate lower engagement in outpatient psychosocial care, higher dependence on emergency-based services, and more fragmented treatment pathways despite comparable baseline mental health conditions. By identifying differences in pre-crisis care trajectories among clinically similar youths, this study aims to support data-driven strategies to close mental health care gaps and strengthen early intervention within Medicaid systems.

4 - Rashomon Sets of Falling Trees

Zachery Boner, Duke University, Durham, NC, United States, Varun Babbar, Margo Seltzer, Cynthia Rudin

Many real-world decisions require prioritizing high-risk cases, such as clinicians prioritizing high-risk patients before lower-risk ones. Falling rule lists (FRLs), which are ordered if--then rules with monotonically decreasing risks, provide an interpretable framework for such tasks; however, their single-path structure yields a highly restricted model class. We introduce falling trees, a new family of interpretable models that enforces the same monotonic risk constraint while permitting tree-structured branching. We present GraviTree, a novel dynamic-programming-with-bounds algorithm for learning the Rashomon set of falling trees under depth and branching constraints, together with bounds that use the falling constraint to provably reduce the search space. Our formulation can interpolate between rule lists and full decision trees, enabling user-desired model expressivity. Across clinical and public-risk datasets, falling trees match or outperform FRLs and other interpretable baselines, often producing lower-sparsity decisions for high-risk instances. Our results show that falling trees strike a practical balance between interpretability, expressiveness, and risk prioritization for high-stakes settings.

TD02

301B

Applied Probability in Healthcare Operations

Invited Session

Applied Probability

Chair: Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

1 - Measuring the Heterogeneous Effect of Emergency Department Boarding on Inpatient Length of Stay: Toward Efficient and Equitable Inpatient Bed Assignment

Jing Dong, Columbia University, New York, NY, United States, Alyssa Liu, Yosef Berlyand, Martin Copenhaver

Emergency Department (ED) boarding, the practice of holding admitted patients in the ED while awaiting an inpatient bed, is a common challenge in hospitals worldwide. It is a major contributor to ED crowding and raises concerns about the quality of care, such as increased morbidity and mortality. When inpatient bed capacity is limited, hospitals must decide how to prioritize boarding patients for inpatient bed assignment in a way that is both efficient and equitable. In this study, we investigate the heterogeneous causal impact of boarding on inpatient length of stay (LOS), a key metric influencing downstream capacity. Using methods from double machine learning, instrumental variables, and heterogeneous treatment effect modeling, we identify patient subgroups for whom prolonged boarding has different effects on LOS. We then integrate these individualized effect estimates with queueing models and policy learning techniques to design inpatient bed assignment policies that reduce overall boarding time. A key feature of our approach is the explicit integration of fairness considerations into operational policy design. First, we construct policies that enforce group fairness. Second, we propose a class of policies that navigates the tradeoff between system efficiency and individual fairness. We validate our proposed policies using a high-fidelity, trace-driven simulation model calibrated with real hospital data. Results show that our policies can reduce the average boarding times by over 75\% compared to the current bed assignment policy while maintaining strong fairness performance.

2 - A Brownian Inventory Model for Last-Mile Health Distribution in Liberia

Yufan Chen, Booth School of Business, University of Chicago, Chicago, IL, United States, Baris Ata, Robert Montgomery

Liberia is among several countries that have successfully introduced Community Health programs to combat common childhood illnesses such as malaria, diarrhea, and pneumonia. These community-focused delivery systems, while effective in offering services at the point of need, present inventory balancing challenges in a system prone to frequent medication shortages. When CHAs lack the necessary medications, patients are forced to travel to health facilities (HFs) to seek care. Travel distances vary between locations, but often result in lost wages. Furthermore, the dispersal of medications between the HFs and CHAs increases the probability that patients arriving at a HF will encounter a medication stockout at the central location.

To address this challenge, we formulate this last-mile distribution problem as a finite-horizon, two-echelon stochastic control model. The system dynamics are driven by a multidimensional Reflected Brownian Motion (RBM) featuring coupled boundary conditions that capture the flow of referred patients.

To overcome the curse of dimensionality in this problem, we employ the Deep Splitting Method, a state-of-the-art approach that leverages deep neural networks to approximate the solution to the high-dimensional Hamilton–Jacobi–Bellman (HJB) equation. By calibrating the model with empirical malaria data from Lofa County, Liberia, we demonstrate the method's computational feasibility for large-scale networks. Furthermore, our analysis quantifies the critical trade-off between minimizing system costs and maximizing health equity, providing actionable insights for improving humanitarian supply chains.

3 - Ai as A Scalable Mental Health Resource: Evidence on Guided use Among Adolescents

Jingyuan Hu, University of California, Irvine, Irvine, CA, United States, Junyou Chen, Yican Liu

This study examines whether structured guidance on artificial intelligence (AI) use can improve adolescents' mental health outcomes. AI tools are increasingly used by young people for emotional support and problem solving, yet their effectiveness as an informal mental health input remains unclear. In the absence of guidance, such use may be inconsistent or counterproductive. We conduct a field experiment in which students are exposed to different approaches to AI use. We track changes in AI engagement, psychological well-being, and academic performance over time. The study evaluates whether improving the quality of AI interaction enhances its value as a low-cost, scalable mental health resource. More broadly, it informs how digital tools can support early-stage mental health care and expand access in resource-constrained settings.

4 - Data to Dose: Personalized Dosing with Data Generation and Expert Feedback

Hansraj Satyam Verma, Carnegie Mellon University, Pittsburgh, PA, United States, Holly Wiberg, Shixiang Zhu, Sridhar Tayur

Personalized treatment optimization is often constrained by data scarcity, particularly in settings involving heterogeneous patient profiles where fully labeled clinical outcomes are limited. Standard data-driven approaches struggle in these regimes, limiting their ability to provide individualized dosing recommendations. We introduce GenEx, a hybrid framework that integrates preference-based Bayesian optimization with a generative model refined through expert feedback. Expert-provided pairwise dose rankings guide updates to both a surrogate model and a generative component, which proposes synthetic data points in uncertain regions of the patient-dose space when the generator is sufficiently aligned with expert preferences. A trust gate controls synthetic injection, mitigating the accumulation of bias from generated samples. Theoretical analysis establishes sublinear regret bounds and bias-aware confidence intervals that account for the interaction between expert feedback and synthetic augmentation. Computational experiments on simulated and retrospective clinical data demonstrate that GenEx achieves faster convergence and stronger dosing utility relative to preference-only and synthetic-only baselines. These results suggest that integrating expert preferences with generative augmentation can meaningfully improve personalized decision-making in data-limited clinical contexts.

TD03

302A

Modeling Drug Development and Clinical Trials

Invited Session

Healthcare Operations Management

Chair: Ozge Yapar, Indiana University Bloomington, Bloomington, IN, United States

1 - Clinical Trials and Conditional Approval Schemes: Implications for Healthcare Systems and Companies

Ozge Yapar, Indiana University, Bloomington, IN, United States, Andres Alban

Conditional approval (CA) schemes used by healthcare payers allow new drugs to receive reimbursement based on preliminary evidence, with continued reimbursement contingent on further data collection. The goal of CA schemes is to balance early patient access to new treatments with informed reimbursement decisions. Examples include the UK's Cancer Drugs Fund and Innovative Medicines Fund. We study how the introduction of CA schemes changes pharmaceutical companies' incentives in clinical trial design, particularly their choices of sample size and submission timing, and analyze whether these schemes improve or distort evidence production and reimbursement outcomes. We use a Bayesian decision-theoretic model in which a company chooses the trial's total sample size and timing of the submission based on anticipated outcomes of regulatory approval and cooperative bargaining with a payer. Negotiations determine reimbursement decision and pricing and are modeled with a Nash cooperative bargaining model in which the company maximizes profits, and the payer maximizes population health value. Our results suggest that CA can improve outcomes for company, payers, and patients, but its effectiveness depends critically on how the scheme interacts with the company's trial design incentives. We show that the company might strategically reduce the amount of data collected for initial submission while increasing the total sample size of the trial. Thus, while conditional approval expands total evidence generated for a treatment, it also might induce earlier submissions supported by less data. A case study calibrated to publicly available data illustrate and quantify these effects.

2 - Information Design for Dose-Finding Trials

Amin Khademi, Clemson University, Clemson, SC, United States, Ningyuan Chen

To enhance enrollment rates in early-stage dose-finding clinical trials, we propose an information design approach, where the clinical investigator (CI) commits to an information releasing mechanism (IRM) based on the treatment's uncertain efficacy and toxicity to encourage patients to participate in the trial. The optimal IRM has a threshold structure, involving completely revealing/pooling the state (utility) when the realization is in the respective region. We analyze the comparative statics under the optimal IRM. Because general IRMs may be

difficult to implement in practice, motivated by response-adaptive clinical trials where information from past patients will help drive decisions for future patients, we consider a practical IRM, where the CI decides on the number of patients to recruit, whose efficacy/toxicity outcomes will be used as information for future patients. For example, the proposed IRM may be the number of patients in the first batch in dose-escalation methods in Phase I, expansion cohort at the end of Phase I, or the number of patients assigned to maximum tolerated dose in Phase II. We show this practical IRM can achieve at most 50% of the value of the optimal abstract IRM.

3 - Conditional and Lenient Drug Approvals

Chenxi Xu, Duke University, Durham, NC, United States, Giuseppe Lopomo, David Ridley, Peng Sun

Firms are often allowed to sell new drugs based on preliminary efficacy evidence, with the final approval contingent on confirmatory testing. We characterize optimal approval policies for a government agency that maximizes consumer welfare while considering the firm's incentives. We assume that firms have private information about testing costs and the government's payoff depends on expected efficacy. The optimal policy may include partial conditional approval for drugs with low expected efficacy, and leniency in granting final approval below conventional standards. Our calibration suggests these tools can generate hundreds of millions of dollars in annual social value.

TD04

302B

Empirical Studies of Provider Behavior

Invited Session

Healthcare Operations Management

Chair: David Rea, Lehigh University College of Business, Bethlehem, PA, United States

1 - Hidden Flexibility: How Workload Shapes Strategic Effort Allocation During ED Shifts

David Rea, Lehigh University College of Business, Bethlehem, PA, United States

tbd

2 - The Impact of Hospital Algorithm Adoption on Care Delivery, Outcomes, and Equity

Qi Wang, Boston University, Boston, MA, United States, Jiayi Liu, Anandhi Bharadwaj, Ruba Aljafari, Abhinav Goyal

Predictive algorithms are increasingly deployed in clinical decision support, promising efficiency but raising concerns about codifying historical biases. Using a difference-in-differences analysis of 308,137 inpatient admissions for acute coronary syndrome, we find that hospital algorithm adoption reallocates care toward high-severity patients—more diagnostic testing, more intensive treatment, and lower mortality—while care for low-severity patients is de-escalated with no change in outcomes. These survival gains, however, accrue exclusively to White patients; non-White high-severity patients see no mortality improvement and receive less diagnostic testing and intensive treatment following adoption. Efficiency-oriented algorithms can deepen health inequities, underscoring the need to manage these tools for equitable outcomes.

3 - Operational Overload: the Impact of Workload on High-Skilled Workforce Attrition

Blair Liu, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Diwas KC, Brad Staats

Worker attrition is critical and costly, disrupting operations across industries and leading to significant productivity losses. In healthcare, nurse attrition poses even greater challenges, exacerbated by persistent shortages and increasing burnout. Despite its importance, nurse attrition remains underexplored in operations management (OM) literature, particularly concerning how different workload dimensions influence voluntary attrition. This study aims to address this gap by investigating how various workload dimensions. Utilizing high-resolution data from a large U.S. hospital system, we analyze 26 months of operational, clinical, and HR records, capturing nurses' dimensions of workload leading up to voluntary attrition decisions. By addressing a critical issue in high-stakes environments like healthcare, this study enriches OM literature and provides practical insights for organizations seeking to retain talent in knowledge-intensive fields.

4 - Physician Variability and Throughput Instability in Emergency Department Operations

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, David Rea, Margret Bjarnadottir, David Anderson

We study how attending physician heterogeneity in the emergency department affects disposition decisions, triage-to-disposition time (TTD), and downstream throughput stability. Using ED encounter data linked to attending physicians, we estimate machine learning models for disposition choice and for admission and discharge TTD. A calibrated simulation framework then models these operational constraints dynamically across several scenarios: actual, baseline, neutralized (physician effects removed), stressed (concentrating top-effect physicians). The estimates reveal substantial physician-level variation in both admission propensity and TTD. Provider profiling further suggests more years in practice are associated with shorter TTD for both admissions and discharges. In counterfactual analysis, concentrating high-effect physicians materially increases operational volatility rather than simply shifting mean throughput. Relative to the neutralized scenario, the stressed scenario increases the daily admission interquartile-range ratio by 1.44 and the high-admit-day rate by 31.4 percentage points, while admit-side TTD tail risk rises from 3.4% to 13.0%. These findings show that physician heterogeneity can act as an operational shock, implying that variance-aware scheduling is important for maintaining throughput stability.

TD05

302C

Economic Evaluation and Modeling to Inform Public Health Policy Decisions

Invited Session

Public Health Policy

Chair: Xinmeng Zhao, Centers for Disease Control and Prevention, Atlanta, GA, United States

1 - Modeling the epidemiology dynamics of Respiratory Syncytial Virus infections in the United States under novel pharmaceutical-based interventions

Xinmeng Zhao, Centers for Disease Control and Prevention, Atlanta, GA, United States, Brian Gurbaxani, Michael Melgar, Jefferson Jones, Heidi Moline, Ismael R. Ortega-Sanchez, Diya Surie, Emily Carter, Ryan E. Wiegand, Phillip Salvatore

Respiratory syncytial virus (RSV) has historically been a leading cause of hospitalization among infants and a substantial cause of hospitalization among older adults in the US. The implementation of novel RSV immunization products and strategies requires a modeling framework that captures direct protection and indirect population-level effects. Existing RSV transmission models often include multiple immunity strata but assume unidirectional progression toward higher immunity levels without waning to lower immunity states.

To address this gap, we developed a deterministic, age- and risk-structured susceptible–exposed–infectious–recovered (SEIR) transmission model that incorporates bidirectional transitions across immunity strata. Individuals move to higher immunity levels upon recovery from infection, while immunity wanes over time, allowing return to lower protection levels across the life course. The model was calibrated to pre-pandemic weekly RSV hospitalization data in four high-burden age groups

The calibrated model reproduced pre-pandemic RSV hospitalization burden in the four calibrated age groups, with seasonal hospitalization totals falling within the posterior 95% Crl for all age-season combinations. Among infants aged 0–5 months, the reported total in 2018–19 was 1,633.2 per 100,000, compared with a simulated median of 1,747.7 (95% CI, 1,217.4–2,510.6). Calibration results further indicated that observed epidemic dynamics were consistent with a period of heightened post-infection protection lasting 250 days (95% Crl: 200-500), aligning with the persistence of elevated neutralizing antibodies observed in human challenge studies.

This platform provides a tool for assessing current and future RSV intervention guidelines including their direct and indirect effects, disease burden and economic implications.

2 - The Economic Burden of Congenital Cytomegalovirus in The United States

Han Xiao, Centers for Disease Control and Prevention, Atlanta, GA, United States, Andrew Leidner, Tatiana Lanzieri, Scott Grosse

Congenital cytomegalovirus (cCMV) infection is the most important infectious cause of hearing loss and neurodevelopmental delays in US children. Although most infected infants are asymptomatic at birth, some can develop long-term conditions, which can generate significant direct and indirect economic costs. Despite growing interest in prevention and treatment strategies, including newborn screening, vaccine development, and antivirals, the economic burden of cCMV in the United States has not been well–characterized in recent years.

We developed a cost-of-illness model to estimate the lifetime economic burden of cCMV for the 2024 birth cohort in the US from a societal perspective, accounting for direct medical costs and indirect costs including special education and productivity losses. Long-term clinical conditions included hearing loss, visual disorders, neurological disorders, and intellectual disabilities. To account for co-occurring conditions and avoid double-counting, we employed a maximum entropy framework to estimate joint-state probabilities consistent with pooled marginal prevalence estimates and epidemiologic constraints. One-way and probabilistic sensitivity analyses were conducted to assess parameter uncertainty.

We estimated at least 16,000 cases with cCMV infections occurred in infants born in 2024; approximately 2,000 cases developed long-term conditions. The total lifetime economic burden was estimated to be >$2 billion for the birth cohort (range: $1.5- $4 billion). Co-occurring conditions involving neurological and neurodevelopmental disorders exhibited the highest per-case costs, exceeding $2 million.

This study accounted for co-occurring conditions and updated inputs to produce more conservative cCMV burden estimates, which could provide foundation for future cost-effectiveness evaluations of cCMV prevention and intervention strategies.

3 - Evaluating Housing-Based Interventions for People Experiencing Homelessness and Opioid use Disorder in California: a Microsimulation Model

Yuewei Ling, Stanford University, Stanford, CA, United States, Sean Tsung, Margaret Brandeau, Marissa Reitsma

California’s intertwined opioid and housing crises require policies that address both health and social needs. People experiencing homelessness (PEH) face high rates of opioid use disorder (OUD), overdose, unmet health needs, and other barriers to recovery. The objective of our model-based study is to guide the development of integrated housing and addiction policies in California, which has the largest population of PEH in the US. Using data from the California Statewide Study of People Experiencing Homelessness, we developed a microsimulation model to estimate the lifetime health and economic consequences of a “Housing First” style intervention for PEH with OUD in California. The model tracked monthly transitions across three housing states (unsheltered, sheltered, housed) and three opioid use states (off treatment, on treatment, recovered). Model parameters were drawn from California data and published literature. Compared with the status quo, the housing intervention increased time spent in permanent housing by 1.26 years per person, averted 2 expected lifetime overdoses per 100 persons, and generated 0.39 additional quality-adjusted life-years (QALYs). Discounted lifetime costs increased by $10,663, yielding an incremental cost-effectiveness ratio of $27,220 per QALY gained. The intervention remained cost-effective when housing costs were doubled and when the PEH-associated health utility decrement was decreased by 50%. In probabilistic sensitivity analysis, the intervention was preferred in 97% of simulations at a willingness-to-pay threshold of $50,000 per QALY. These findings suggest that expanding stable housing for PEH with OUD can improve health at an acceptable cost, supporting housing as both a social and health policy.

4 - Phase-Specific and Biomarker-Stratified (Hbeag Status) Health Utilities in Chronic Hepatitis B: Implications for Health Policy and Economic Evaluation

Yichi Yin, University of Waterloo, School of Pharmacy, Kitchener, ON, Canada

Background:
Health utility estimates are essential inputs for decision-analytic models. Although utility values have been

reported in CHB populations, these estimates are not aligned with immunological disease phases and stratified by key biomarkers (e.g., HBeAg status), limiting their structural suitability for decision models.

Objective:
To generate phase-specific health utility estimates for CHB by systematically synthesizing published HRQoL data stratified by disease progression stage including HBeAg status and treatment exposure.

Methods:
We will conduct a systematic review of MEDLINE, Embase, and Cochrane Library for studies reporting preference-based health utility values in CHB individuals from 1989 onward. Eligible studies must report original utility data using validated instruments (e.g., EQ-5D, HUI) or direct measurement (e.g., Time Trade-Off, Standard Gamble). Utility estimates will be categorized according to clinically meaningful CHB health states, each further stratified by HBeAg status and treatment exposure.

Meta-analyses will be conducted within health-state–instrument using fixed- or random-effects models. A multi-level multivariate meta-regression will integrate utility values across instruments while accounting for clustering within studies and heterogeneity related to study design, liver disease severity, treatment status, and measurement method. Publication bias and statistical heterogeneity will be assessed, and sensitivity analyses will test robustness to missing variance estimates.

Results (Expected Contribution):
The study will generate standardized mean utility values aligned with clinically meaningful CHB health states. Cross-instrument harmonization will provide internally consistent utility inputs.

Conclusion:
Aligning health utilities with CHB disease progression and HBeAg status strengthens the methodological rigor of economic modeling and supports more accurate cost-effectiveness evaluations of emerging therapies.

TD06

303

Medical Decision Making Using Real-World Evidence

Invited Session

Medical Decision Making

Chair: Mert Sahinkoc, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States

Co-Chair: Jagpreet Chhatwal, Harvard Medical School, Mass General Hospital, Boston, MA, United States

1 - Machine Learning Based Analysis of Asthma Heterogeneity Using Multi-Domain Data

Gokalp Cosgun, Wayne State University, Detroit, MI, United States, Audrey Fu, Inas Khayal, Deborah Ellis, Melike Yildirim

Asthma is a heterogeneous chronic disease influenced by clinical, environmental, and social factors, with persistent disparities in control and outcomes across populations. This study aims to identify asthma subgroups and examine factors associated with uncontrolled asthma by combining unsupervised clustering with supervised prediction in a real-world cohort.

We analyzed Michigan-based adults with persistent asthma from the Electronic Health Records (EHR) dataset. A multidomain feature set spanning demographic, socioeconomic, environmental, comorbidity, clinical, and social determinants of health variables was constructed. Unsupervised k-means clustering was used to identify patient subgroups with similar profiles. In parallel, supervised machine learning models were developed to predict asthma control status, and model performance and feature importance were evaluated both in the full cohort and within clusters.

The clustering analysis identified subgroups with distinct contextual and clinical profiles that differed in asthma control rates. Supervised models, particularly tree-based approaches, showed strong overall discrimination. Feature contributions varied across clusters, with differences observed across multiple domains, including comorbidities, medication patterns, and healthcare utilization.

Together, these results illustrate how integrating clustering with supervised prediction can provide complementary perspectives on asthma heterogeneity in real-world data, and how subgroup structure can inform the interpretation of predictive modeling results without relying on a single population-level summary.

2 - Neural Index Policies for Multi-Action Restless Multi-Armed Bandits with Heterogeneous Budgets

Himadri Pandey, Georgia Institute of Technology, Atlanta, GA, United States, Kai Wang, Gian Garcia

Restless Multi-Armed Bandits (RMABs) provide a scalable framework for sequential decision-making under uncertainty. However, classical assumptions about RMABs on binary actions and global budgets restrict their applicability to real-world healthcare problems with multiple actions. To address this, we introduce a Neural Index Policy (NIP) for multi-action RMABs with heterogeneous budget constraints. Our approach learns to assign budget-aware indices to arm-action pairs using a neural network, and converts them into feasible allocations via a differentiable knapsack layer formulated as an entropy-regularized optimal transport (OT) problem. The resulting model unifies index prediction and constrained optimization in a single end-to-end differentiable framework, enabling gradient-based training directly on decision quality. The network is optimized to align its induced occupancy measure with the theoretical upper bound from a linear programming relaxation, bridging asymptotic RMAB theory with practical learning. We evaluate NIP on a simulated dataset and a pediatric post-surgical monitoring task. Empirically, NIP achieves near-optimal performance (within 5% of the oracle) while strictly satisfying heterogeneous budgets and scaling to hundreds of arms.

3 - Reducing Kidney Discards Through Risk-Aware Organ Acceptance Policies for Out-of-Sequence Offers

Yili Wang, University of Michigan, Ann Arbor, MI, United States, Amirhossein Moosavi, Mariel Lavieri, Gongjun Xu, Ji Zhu, Yee Lu, Danielle Haakinson

The out-of-sequence policy, which is used to allocate marginal kidneys, allows transplant centers to bypass higher-ranked candidates and allocate organs to patients lower on their center-specific waitlists. This creates a complex decision-making problem for clinicians, who must determine which patients—differing in health status and waitlist position—should receive such kidneys. The challenge is further complicated by variation in clinicians' risk tolerance—potentially influenced by their transplant center's historical performance—and differences in the time horizons used to evaluate outcomes. In this study, we develop a Conditional Value-at-Risk (CVaR)-incorporated Markov Decision Process (MDP) framework to support decision-making for patients and clinicians regarding whether to accept kidneys under the out-of-sequence policy. Our results show that higher levels of risk tolerance are required to justify accepting lower-quality kidneys, particularly for patients in poorer health and when decisions are evaluated over longer time horizons. Applying our framework to patients who died or were removed from the waitlist due to medical unsuitability between 2022 and 2024, we estimate that approximately 12% of those patients could have received an out-of-sequence transplant prior to death or removal.

4 -

5 - A Microsimulation-Based Policy Analysis for Risk-Stratified Liver Cancer Surveillance

Mert Sahinkoc, Massachusetts General Hospital, Harvard Medical School, Boston, MA, United States, Gizem Nemutlu, Hashem El-Serag, Mary Linton Peters, Ali Hajjar, Leigh Anne Dageforde, Eugene Koay, Fasiha Kanwal, Jagpreet Chhatwal

Current surveillance for liver cancer in patients with cirrhosis is implemented using a uniform semiannual ultrasound strategy, despite substantial heterogeneity in individual risk and real-world adherence. We developed a decision-analytic microsimulation model to evaluate risk-stratified surveillance policies informed by a validated, real-world–derived risk prediction model (Texas HCC Risk Index) that was developed using

multicenter U.S. cohorts across eight institutions and externally validated in a national Veterans Affairs cohort of over 20,000 patients.

The model simulates disease progression, competing mortality, and surveillance pathways in a heterogeneous population. Real-world evidence was incorporated to parameterize risk distributions, adherence patterns, and clinical pathways. We identified optimal decision thresholds that determine when to initiate, escalate, or de-escalate surveillance (no screening, ultrasound-based strategies, abbreviated MRI, or MRI) as a function of underlying liver cancer risk. Outcomes included, costs, stage at detection, and downstream diagnostic utilization.

The optimal policy transitions from no screening to ultrasound at an annual liver cancer incidence of 0.25%, and from ultrasound to MRI at 2.40%. When abbreviated MRI is available, transition from ultrasound to abbreviated MRI was at 0.70% and from abbreviated MRI to full MRI at 4.10%. The resulting policy improved health outcomes and was cost-effective compared with a uniform strategy. Risk-based policies increased detection of very early-stage cancers (42.0 vs 36.2 per 1,000 patients) while reducing total testing burden (6,798 vs 8,151 tests per 1,000) and downstream diagnostics.

This study demonstrates how real-world evidence can inform risk-based liver cancer screening policies that improve early detection while reducing unnecessary testing.

TD07

304

From Prediction to Action: Data-Driven Methods for Clinical Decision Optimization

Invited Session

Health Applications Society

Chair: Katherine Adams, University of Texas at San Antonio, San Antonio, TX, United States

1 - Optimizing Micro Randomized Trials with Nonstationary Power Constrained Multi-Armed Bandits

Yonatan Mintz, University of Wisconsin Madison, Madison, WI, United States, Fengxu Li

Multiarmed bandits (MABs) are a widely used framework for sequential decision-making under uncertainty, with significant applications in personalized healthcare. While traditional MAB methods typically assume stationary reward distributions, real-world settings are often nonstationary. For instance, in personalized interventions, treatment effectiveness may decline due to habituation and later recover, creating complex nonstationary behaviors that standard methods fail to capture. In addition to adapting to these evolving reward structures, a critical challenge in clinical trial contexts is preserving statistical power for reliable inference post hoc. Existing bandit algorithms generally optimize cumulative rewards but often neglect adequate exploration which is necessary for detecting meaningful treatment effects. In this talk we propose a novel posterior sampling algorithm explicitly designed to tackle this dual challenge. Our method adapts to nonstationary environments characterized by habituation and recovery dynamics, while simultaneously satisfying specified power constraints through controlled exploration. We provide theoretical guarantees via sublinear regret bounds and validate the approach on a real world data set from a sedentary behavior and physical activity intervention. In addition, we present results from a real world prospective trial we ran using our approach to promote safer driving.

2 - Predicting Pediatric Surgery Duration via Machine Learning

Arka Roy, The University of Texas at San Antonio, San Antonio, TX, United States, Katherine Adams, Yang Yang, Gian-Gabriel Garcia, Xin-Li Leong, Dai Chung, Vishal Ahuja

Accurate prediction of pediatric surgical case duration is critical for efficient operating room management. We developed machine learning models that combine clinical variables with novel non-clinical features capturing team familiarity, surgeon experience, and operational context, and evaluated them using a large pediatric surgical dataset. Among the models compared, LightGBM achieved the best performance and was the most computationally efficient, substantially reducing prediction error relative to existing scheduling estimates. Non-

clinical factors represented half of the most influential predictors, highlighting the importance of operational dynamics. Performance gains were greatest in otolaryngology, gastroenterology, orthopedics, and among children in early childhood. These results demonstrate that incorporating team and operational features into machine learning models can significantly improve pediatric surgical duration prediction, supporting more accurate scheduling and improved operating room efficiency.

3 - Data-Driven Decision Support for Anesthesiology Assignment

Katherine Adams, University of Texas at San Antonio, San Antonio, TX, United States, Arka Roy, Vishal Ahuja

Anesthesia staffing in high-concurrency academic environments is a complex, high-stakes administrative task that contributes significantly to provider burnout and perceived inequity. We present a data-driven decision-support framework designed to automate high-level staffing decisions – including supervision type and concurrency levels – while promoting longitudinal fairness across anesthesiologists.

We developed a data-recovery pipeline to extract and structure historical faculty-supervisee assignments from PDF documents at a large academic health center and affiliated county hospital system. This longitudinal dataset served as the foundation for supervised learning models that capture institutional staffing norms and encode implicit provider preferences through historical assignment frequencies. The resulting system generates a daily "staffing blueprint" specifying the type and number of supervisees (residents or CRNAs) for each attending anesthesiologist.

The pipeline successfully transformed unstructured records into a strategic database, uncovering that faculty frequently exceeded contracted room-coverage thresholds. Despite the inherent volatility of surgical schedules – where add-on cases represent approximately 40% of daily volume – the model produced consistent recommendations. Critically, this shifted the "board runner" (lead anesthesiologist) role from manual schedule generation to expert refinement within structured, data-informed parameters.

This framework directly targets two drivers of anesthesiologist burnout: excessive cognitive load and lack of perceived institutional support. By anchoring assignments in historical patterns, the system promotes fairness over time without sacrificing flexibility. We argue that integrating algorithmic decision-support into perioperative management preserves clinical judgment while ensuring operational safety and contractual compliance – offering a scalable solution for an increasingly strained anesthesia workforce.

4 - Surrogate-powered Causal Inference on Censored Survival Outcomes

Arielle Anderer, Cornell University SC Johnson College of Business, Ithaca, NY, United States, Jelena Bradic, Yaroslav Mukhin, Tereza Oprea, Christina Yu

Missing information is a central challenge in clinical trials with survival endpoints. Limited follow-up and right censoring reduce the effective sample size, weakening statistical power even in randomized experiments. We study how much of the efficiency lost to censoring can be recovered by incorporating intermediate post-treatment outcomes, such as disease progression or recurrence, alongside observed survival data when estimating treatment effects on long-term outcomes such as overall survival. Censoring creates an opportunity to leverage additional sources of information to improve efficiency. When survival outcomes are incomplete, intermediate post-treatment variables and baseline covariates provide complementary information about censored event times. Incorporating these signals allows recovery of a portion of the information lost to censoring, yielding more precise estimates of marginal causal effects. These gains do not rely on restrictive modeling assumptions such as proportional hazards or full mediation, highlighting the practical value of surrogate outcomes in censored clinical trial settings. We characterize the magnitude of these efficiency gains. The extent of recovery depends on (i) the balance of surrogate-defined subgroups, (ii) the degree to which the surrogate separates patients with different survival prospects, and (iii) the timing of the surrogate relative to censoring. These insights clarify when surrogate endpoints are most informative. The implications for trial design are direct: incorporating informative surrogate outcomes can help trial designers increase power, shorten follow-up horizons, or reduce required sample sizes. We quantify these gains using influence function–based variance decompositions, supported by simulations and a semi-synthetic application calibrated to oncology trials.

TD08

305A

Health Equity

Invited Session

Health Opportunities for All

Chair: Shannon Harris, Virginia Commonwealth University, Richmond, VA, United States

1 - Punishing the Symptom: Dismissal Policies and Disparate Outcomes for ADHD Patients in Primary Care

Christopher Kwaramba, Virginia Commonwealth University, Richmond, VA, United States, Christopher Kwaramba, Jeff Shockley, Shashank Rao, Madeline Straubel

Healthcare systems use formal dismissal policies to manage appointment nonattendance, terminating the patient-provider relationship after a threshold number of missed visits. The policies are facially neutral. Their effects are not. Across 782,891 appointments at 55 primary care clinics, patients with ADHD are dismissed at more than twice the rate of patients without ADHD (OR 2.135, adjusted), a disparity that survives controls for socioeconomic status, insurance, distance, and comorbidities. Prospective memory theory locates ADHD nonattendance at the execution stage of intention fulfillment, not at formation or retrieval. Patients with ADHD intend to attend and remember the appointment; they fail to act on the intention at the moment of execution. Dismissal policies treat this execution failure as equivalent to deliberate noncompliance, sanctioning patients for symptoms of the condition the system exists to treat. The disparate impact is compounded by a procedural problem: the policy does not account for the conditions it punishes. Reminders and patient education do not address this failure mode, because the failure is not in information transfer. The remedy is redesign of the policy itself.

2 - Hidden Figures in Cancer Care: How Quality Measures Obscure Racial Inequalities at End-of-Life

Inas Khayal, Wayne State University, Novi, MI, United States

Calls to reduce disparities in cancer care often rely on aggregate comparisons across racial and ethnic groups, which may obscure inequities driven by systematic differences in the timing of care. This study aimed to (1) identify racial/ethnic differences in the timing of end-of-life care delivery and (2) assess whether such differences are detectable using standard quality measures.

We conducted a retrospective decedent follow-back study of Medicare beneficiaries aged 65–99 with advanced cancer who died between April and December 2016. Among hospitals treating at least 11 White decedents and 11 decedents of color (including Black or African-American, Asian/Pacific Islander, Hispanic, and American Indian/Alaska Native), we calculated hospital-level differences between groups using: (1) aggregate measures of any palliative care or hospice use, and (2) longitudinal “signatures” capturing daily use of these services during the six months before death.

The cohort included 30,319 decedents across 217 hospitals; 7,852 (25.9%) were people of color (POC). The median hospital-level difference in aggregate measures was −5.35% (IQR 12.83) for palliative care, indicating higher use among POC, and 3.66% (IQR 12.45) for hospice care, indicating higher use among White patients. Longitudinal signatures revealed five distinct patterns of inequality. However, inequality classifications based on signatures aligned with aggregate measures in only 46.5% of hospitals for palliative care and 39.2% for hospice care.

Incorporating timing through longitudinal data uncovers racial and ethnic inequities in end-of-life cancer care that are frequently missed by traditional aggregate quality measures.

3 - Implicit Racial Bias in Healthcare Scheduling Delays

Shannon Harris, Virginia Commonwealth University, Richmond, VA, United States

We show that the traditional objective of minimizing patients' waiting time and provider overtime leads to scheduling the patients at higher risk of no-show farther into the future than other patients. Because no-show probabilities are often correlated with race, this results in racial disparities in terms of access to care.

4 - Equity-Efficiency Symbiosis

David Rea, Lehigh University College of Business, Bethlehem, PA, United States, Violet (Xinying) Chen

While previous research has focused on the potential tradeoff between equality-based fairness and efficiency, this research considers when equity-
based fairness works in conjunction with efficiency. Such scenarios occur frequently in healthcare.

TD09

305B

Practical AI #2: How US Hospitals Actually Use AI

Invited Session

Healthcare Practice

Chair: A. Cecilia Zenteno, Massachusetts General Hospital, Boston, MA, United States

Co-Chair: Martin Copenhaver, Johns Hopkins Hospital, Baltimore, MD, United States

1 - Mayo Clinic Practice Transformation with Artificial Intelligence (AI)

Yu-Li Huang, Mayo Clinic, Rochester, MN, United States

Transforming practice operations using artificial intelligence (AI) is indeed one of the hottest topics in healthcare, translating operations research and data science principles to applications in system integration. This presentation provides: 1. Facts regarding Mayo Clinic and Kern Center for the Science of Health Care Delivery; 2. Mission for AI works in process improvement space for practice transformation; 3. Projects demonstration on the impact of operational efficiency using operations research, data science, and system engineering principles; 4. The skills needed to succeed in healthcare environment, closing the gap between theory and practice.

2 - Implementation of a Short-Stay Patient Prediction Model in the Mgh Emergency Department

A. Cecilia Zenteno, Massachusetts General Hospital, Boston, MA, United States, Martin Copenhaver, Retsef Levi, Peter Dunn, Bethany Daily, Pratyush Bhattacharya, Thomas Kyle, Kyan Safavi

The MGH Emergency Department (ED) faces 60+ daily boarders (94% admissions) with a median boarding time of 12.3 hours. Among these, approximately 19 admissions per day are discharged in less than 72 hours (short-stay patients). Housing short-stay patients on inpatient floors reduces bed availability for higher-acuity cases and potentially increases their length of stay (LOS). To address this the Admission Pathways Team (APT) identifies potential short-stay patients to reroute them to a newly created, separate unit. However, manual identification requires chart review and clinical judgment - a time-consuming, non-trivial process. The ED Short Stay Tool supports this process by generating a score that correlates with the likelihood of a patient being short-stay. In this talk, we will present how the algorithm was originally conceived and how it has evolved with its deployment, as well as future directions.

3 - Deploying and Scaling An Algorithm-Enabled Type One Diabetes Care Platform

Jamie Kurtzig, Stanford University, Stanford, CA, United States, David Scheinker

Few AI-enabled clinical technologies transition beyond academic proof of concept. We describe a scalable, algorithm-driven remote patient monitoring platform for type 1 diabetes that has moved from single-center outcomes research to a clinic-agnostic, nationally deployable solution — and outline ongoing work to expand its capabilities.

The Teamwork, Targets, Technology, and Tight Control (4T) Study has helped newly diagnosed pediatric type 1 diabetes patients achieve a mean HbA1c of 6.58% and time in range of 68% at 1-year post-diagnosis. As part of

the 4T Study, patients receive remote patient monitoring (RPM) via the Timely Interventions for Diabetes Excellence (TIDE) platform. TIDE identifies patients at higher risk of deteriorating glucose management. TIDE has improved patient glycemic outcomes, reduced clinician screen time, and is financially sustainable at CMS reimbursement rates.

In partnership with the diabetes technology non-profit Tidepool, we developed Tidepool-TIDE, a clinic-agnostic, turnkey solution available to any clinic in the United States. Tidepool-TIDE provides EHR integration; APIs supporting data exports of patients flagged for clinical review; and hosts data for 85 CGMs, insulin pumps, and blood glucose meters. To facilitate the adoption of TIDE, we developed a quantitative framework for monitoring clinical workload, patient glucose management, and timeliness of care associated with the use of TIDE.

Ongoing efforts to improve TIDE include using LLMs to simplify dot-phrases, pre-defined message templates summarizing CGM data, and improving the readability of the messages written by clinicians to patients.

4 - Real-TIME AI Detection of Hospital-Acquired Conditions Within Native Clinical Workflows

Sai Samyukta Palle, Duke University School of Medicine, Durham, NC, United States, Mike Revoir, Marshall Nichols, Will Knechtle, Matt Gardner, Bradley Hintz, Suresh Balu, Dr. James J Jung

Background: Incomplete documentation of hospital-acquired conditions obscures the true clinical picture, creating gaps in early detection of clinical deterioration, care coordination, and appropriate revenue capture. When patient complexity is under-represented, the healthcare system fails to recognize true severity of illness, inadvertently leading to inaccurate acuity-based billing. To bridge this gap, we developed Clinical Condition Ascertainment & Detection Engine (CADEN), a real-time AI platform functioning as a clinical co-pilot, identifying high-risk conditions during active inpatient encounters and delivering evidence-based documentation support within native clinician workflows.

Methods: CADEN was deployed January 12, 2026, across Urology, Transplant, and Cardiothoracic, Vascular, Minimally Invasive, and General Surgery divisions, targeting six conditions: acute blood loss anemia (ABLA), acute kidney injury (AKI), hyperkalemia, hypokalemia, hypernatremia, and hyponatremia. The platform integrates into EHR note templates via dot phrases with near–real-time condition identification. Diagnosis capture rates were extracted from EHR billing records and compared between equivalent pre- and post-deployment periods (January 12 – March 31, 2025 vs. 2026), with comparable encounter volumes across both periods.

Results: Over 12 weeks, 257 clinical providers utilized CADEN across 1,278 encounters. Diagnosis capture rates increased across all six conditions: hyponatremia (+126%), hypokalemia (+118%), ABLA (+41%), hyperkalemia (+27%), hypernatremia (+20%), and AKI (+5%).

Conclusions: CADEN was associated with substantial improvements in diagnosis capture across all six targeted conditions, meaningfully enhancing clinical fidelity of the medical record. By ensuring documentation accurately reflects true patient complexity and acuity, the platform naturally optimizes Case Mix Index and positions reimbursement as a transparent reflection of high-intensity resources deployed.

TD10

306A

Early-Career Faculty Mentoring Session

Panel Session

Health Applications Society

Co-Chair: Qiushi Chen, The Pennsylvania State University, University Park, PA, United States

1 - Moderator Panelist

Qiushi Chen, The Pennsylvania State University, University Park, PA, United States

Join colleagues to discuss navigating the pre-tenure years, building a sustainable research program, balancing teaching and service, and growing your professional network.

2 - Panelist

Jennifer Lobo, University of Virginia, Charlottesville, VA, United States

3 - Panelist

Xiang Zhong, Univ. of Florida, Gainesville, FL, United States

TD11

306B

AI Applications in Healthcare Operations and Process Improvement

Contributed Session

AI

Chair: Hsinning Chen, 900 Willow Dr, 27514

1 - Design Principles for Agentic Artificial Intelligence in Healthcare Call Centers: a Path to Care Coordination Realized

Olga Biedova, College of Charleston, Charleston, SC, United States, Matthew Mullarkey, Iris Junglas, Matthew Cantonis

The goal of this research is to identify design principles for agentic artificial intelligence working in situ in a large physician-group healthcare system that improve patient access, satisfaction, and outcomes while maximizing capacity at the lowest cost. Using action design research methods we explore four design activities:

First, AI agents are shown to answer patient calls in milliseconds without additional human staffing, eliminating missed or dropped calls and automating routing to appropriate schedulers. These capabilities substantially improve operational efficiency and patient experience.

Second, AI agents evaluate all feasible scheduling combinations across providers, locations, and constraints to optimize appointments for individual patients while maximizing system-wide capacity utilization. This enables timelier care, improved access, and better clinical outcomes, while increasing overall throughput.

Third, AI agents learn from large volumes of historical scheduling data, including cancellations, no-shows, and incorrect bookings, to generate probabilistic recommendations for improved scheduling decisions. By incorporating critical attributes such as patient acuity, resource constraints, and visit type, these models can reduce inefficiencies and mismatches between demand and capacity.

Fourth, AI agents retain longitudinal knowledge of individual patients, transforming discrete call center transactions into continuous care coordination across visits, specialties, acuity levels, and the broader continuum of care. At this highest level of application, agentic AI enables dynamic adaptation of scheduling, capacity management, and operational workflows.

Together, these capabilities position agentic AI as a foundational tool for redesigning access and care coordination, allowing healthcare systems to deliver more personalized, efficient, and timely care while achieving the lowest cost.

2 - The Automation&Ndash;Alignment Paradox in U.S. Healthcare Operations: Why More AI and Automation Do Not Always Mean Better Decisions or Processes

Gulce Cuhaci Tasdelen, Istanbul Bilgi University, Stamford, CT, United States

The U.S. healthcare sector's transition to value-based care requires balancing multiple objectives across payers and providers, but the operational processes linking them remain mostly designed to optimize a single performance dimension. This paper argues that the primary source of the "automation–alignment paradox" is not

a lack of technological capability but misaligned processes between payers and providers. Deploying AI and automation tools on misaligned workflows often intensifies friction rather than reducing it. In Medicare Advantage, 53 million requests for prior authorization were processed in 2024 with 4.1 million denials, and 80.7% of appealed denials were overturned (KFF, 2026). Based on AMA survey results, 93% of physicians report prior authorization delays patient care, and several AI-powered systems generated up to 16 times higher denial rates than average (AMA, 2024; 2025). In clinical decision support systems, override rates reach 96%, indicating persistent misalignment between automated recommendations and clinical decisions (Graafsma et al., 2024). We analyze three mechanisms through which automation amplifies misalignment: (1) divergence between payer cost-optimization and provider clinical-optimization logics; (2) alert overload that weakens human oversight as overrides scale; and (3) trust erosion that reinforces friction between payer automation and provider defensive behavior. We propose a framework using overrides as structural misalignment indicators, integrating override analysis, automation–alignment trade-off modeling, and time-framed impact estimation. Scenario modeling of representative payer and provider cases, grounded in operational data from AMA, KFF, and CAQH, suggests that aligning payer–provider processes before scaling automation could reduce overrides by 20–30% and improve automation ROI by 25–40%.

3 - Improving the Performance of AI on Target Populations Using Custom Ensembles

Mehmet Ayvaci, UT Dallas, Richardson, TX, United States, Mehmet Ahsen

Ensuring that artificial intelligence (AI) models generalize across patient populations remains a central challenge in healthcare deployment. We propose an adaptive ensemble approach that enhances generalizability by dynamically reweighting outputs from multiple algorithms based on their alignment with a target population.

Using eleven models from a public data competition, we construct both an original ensemble (OE) and an adaptive ensemble (AE). The AE is evaluated on an external cohort of mammograms, with model weights recalibrated using a subset of the target population and performance assessed on held-out data. To ensure robustness, we implement repeated random train–validation splits. We compare AE and OE using standard performance metrics, including AUROC, sensitivity, and specificity, and employ Bayes factor analysis to provide a probabilistic framework for model comparison.

Our approach is designed to identify and upweight models that better match the characteristics of the target population, while downweighting less transferable models. This framework leverages algorithmic diversity to address distributional shifts that commonly arise in real-world healthcare settings.

This work highlights the potential of adaptive ensembling as a practical and scalable strategy for improving the reliability of AI systems across heterogeneous populations. By focusing on generalizability rather than performance in a single dataset, our approach contributes to the development of more robust and deployable AI solutions in healthcare operations.

4 - Ehr Navigation Sequences as Predictors of Resident Performance: a Frequent Pattern Mining Approach

Hsinning Chen, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Catherine Pratt, Ralph Quillin, Danny Wu

Background: Electronic health record (EHR) audit logs provide granular insights into clinical workflows and decision-making. This study employs sequential pattern mining and machine learning to investigate whether EHR navigation behavior can predict the clinical performance of graduate surgery residents.

Methods: We analyzed 1.59 million EHR events from 28 residents using an Epic system. Access logs were aggregated into daily navigation sequences, where characters represented discrete sections (e.g., Clinical Info, Reporting, Medications). Contiguous n-gram patterns (lengths 3–5) were extracted using a 10% minimum support threshold, generating 1,180 candidate features. Feature selection was conducted using Mann-Whitney U tests within a leave-one-out cross-validation (LOOCV) framework to identify performance-discriminating patterns while preventing data leakage. Six machine learning models were evaluated via LOOCV.

Results: Gradient Boosting with the top ten selected features achieved an AUC of 0.857 and an accuracy of 0.893. Specific patterns, such as EAGA (Reporting-Clinical Info-Demographics-Clinical Info) and GFIE (Demographics-Problems-Medications-Reporting), were significantly more frequent in lower-performing residents ($p<0.001$). These findings indicate that lower performance is characterized by repetitive, disorganized multi-section toggling. Conversely, high-performing residents demonstrated more purposeful, direct access. Notably, session length did not significantly differ by performance level, suggesting that navigation quality and workflow efficiency, rather than total time, distinguish clinical proficiency.

Conclusion: This study demonstrates that sequential EHR behavior is a predictive, behaviorally meaningful metric, offering a scalable, objective methodology for resident assessment in surgical training programs.

TD12

306C

AI for Healthcare Oversight and Operations

Invited Session

MSOM: Healthcare

Chair: Mohammad Zhalechian, Kelley School of Business, Bloomington, IN, United States

1 - Decoding Nurse Attrition: from Unstructured Note to Operations Decisions

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, Xinyang Zhou, Pengyi Shi

Nurse attrition is a major operational challenge in healthcare because departures reduce effective staffing, increase workload for remaining nurses, and can trigger reinforcing turnover cycles. At the same time, much of the information health systems collect about why nurses leave remains buried in unstructured exit-survey comments, limiting its use for operational decision making. This study develops an end-to-end framework that transforms nurse narratives into operationally meaningful retention insights. Using de-identified exit surveys covering 26,458 nurse exits across six U.S. provider systems from 2017 to 2024, we apply a locally hosted open-source large language model to classify free-text schedule-related comments into interpretable categories such as understaffing, excessive workload, inflexible scheduling, and work-life balance concerns. We then embed these measures in a discrete-time hazard model of exit timing and connect the results to a mean-field workforce dynamics model to study system-level stability. The results show that understaffing and excessive workload are the strongest predictors of exit risk, with especially large effects for early-tenure nurses. Counterfactual analysis suggests that targeted early-career support can materially improve retention and average tenure. The theoretical analysis further shows that when staffing falls below a critical threshold under limited hiring capacity, turnover dynamics can push the system toward a persistent low-staff equilibrium. More broadly, the study demonstrates how unstructured workforce voice can be converted into actionable signals for staffing policy and organizational resilience.

2 - Detecting Suspicious Pharmacies in Opioid Supply Chains

Iman Attari, University of Cincinnati, Cincinnati, OH, United States, Mohammad Zhalechian

In this research, we propose a semi-supervised graph learning framework designed to proactively detect suspicious pharmacies within the prescription opioid supply chain by leveraging extensive transactional data from the prescription opioid supply network in the US. Integrating this framework into the Drug Enforcement Administration's monitoring process can enhance early detection of opioid diversion, optimize resource allocation for investigations, and ultimately contribute to reducing the diversion of prescription opioids into vulnerable communities.

3 - Ai for the Management of Chronic Eye Disease

Arlen Dean, Washington University in St. Louis, St. Louis, MO, United States, Naichen Shi, Cheoljoon Jeong, Lei Cao, Lizhuo Zhou

We study how predictive analytics can support the management of chronic eye disease. Motivated by challenges in monitoring progression and making timely care decisions, we consider how longitudinal imaging and related

clinical information can be used to generate forecasts of disease progression. We discuss how such predictions may support clinical decision making, including patient monitoring, triage, and treatment planning, while also informing operational decisions such as follow-up timing, clinic scheduling, and the allocation of limited resources in eye care settings. More broadly, this work connects predictive modeling in ophthalmology with healthcare operations by examining how disease forecasts may improve both patient care and system performance. We also discuss practical challenges that arise when developing predictive tools for real-world implementation.

4 - Improving Inspection Resource Allocation to Control Organizational Misconduct

Yuwei Zhou, Indiana University, Bloomington, IN, United States, Finn Petersen, Rachna Shah

While machine learning (ML) has demonstrated significant potential in predicting organizational misconduct, the integration of these predictions into operational decision-making to better prevent the occurrence of misconduct remains a key challenge. This study explores how regulators can optimally allocate limited inspection resources to reduce misconduct in the form of restaurant hygiene violations based on predicted inspection outcomes. We develop a prescriptive resource allocation framework that combines empirical ML predictions with model-based planning to enhance the effectiveness of regulatory inspections.

We first use a gradient-boosted regression trees (GBRT) model to predict the likelihood of restaurant hygiene violations based on historical inspection data, time since the last inspection, and Yelp review information. We then formalize the inspection scheduling problem as a capacity-constrained interactive partially observable Markov decision process (I-POMDP), which models the decay of hygiene compliance over time as a Markov chain and incorporates the uncertainty of the ML predictions. We implement an index policy derived through Lagrangian relaxation to optimally allocate inspection resources.

Evaluating our approach using data from the Southern Nevada Health District (2012–2019), we find that a policy allocating resources based on predictions alone yields a 63% relative improvement over a regular inspection interval. Crucially, combining predictions with our model-based planning provides an additional 48% relative improvement over the prediction-only policy, demonstrating the superior performance of an integrated prescriptive framework.

TD13

307

Digital Frontiers in Healthcare

Invited Session

eBusiness

Chair: Yu Ma, University of Wisconsin Madison, Madison, WI, United States

1 - Physician-AI Collaboration in Patient-Facing Medical Communication: Aligning Clinician and Non-Clinician Evaluations

Mehmet Ayvaci, University of Texas-Dallas, Richardson, TX, United States, Dicle Yagmur Ozdemir, Alejandro Zentner

Patients (typically laypersons) increasingly encounter AI-generated responses to health questions through consumer platforms and clinical workflows, yet responses perceived as high quality by patients may be clinically inaccurate or misleading. This study examines how distinct physician-AI collaboration workflows affect alignment and divergence between clinicians' and patients' evaluations of response quality to patient health questions. In a cross-sectional survey experiment, 100 patients with self-reported relevant health experience and 96 clinicians recruited via Prolific Academic evaluated patient-facing medical responses generated under three workflows—AI-only, combined (physician + AI co-produced), and refined (AI-edited physician response)—against a physician-only baseline, using questions drawn from Reddit's r/AskDocs forum. Both groups rated AI-assisted responses higher in overall quality than physician-only responses, with patients rating them more favorably across all conditions. When clinicians assessed AI-only responses as clinically satisfactory, the evaluative gap between groups persisted (refined: $t = 3.10$, $P = .002$). When clinicians assessed AI-only

responses as clinically unsatisfactory, physician-AI collaboration narrowed that gap substantially—particularly under the refined workflow, which eliminated divergence entirely (refined: $t = -0.30$, $P = .77$; AI-only: $t = 4.05$, $P < .001$). Physician-AI collaboration mitigates the risk of AI-generated health misinformation not by uniformly improving clinical judgment, but by closing the gap between patient and clinician quality perceptions. Collaboration design, not merely AI involvement, is central to the responsible integration of AI into patient-facing medical communication.

2 - AI-Augmented Gatekeeping: Combining Human and Machine Intelligence in Screening Programs

Jiatao Ding, Nanyang Technological University, Singapore, Singapore, Michael Freeman, Tinglong Dai

Health screening programs increasingly deploy AI to prioritize patients and triage follow-up, but the central operational challenge is how to combine AI and clinician review under limited capacity. We develop an analytical model of an AI-augmented screening program in which each case may be automatically cleared, automatically referred for downstream diagnostic workup, or escalated to human review. The program jointly chooses AI capability, the operating threshold, and the routing policy for AI-negative and AI-positive cases, accounting for false negatives, false positives, review costs, referral costs, and limited review capacity. We show that optimal review is concentrated on borderline cases near the AI threshold rather than spread uniformly across patients. With scarce clinician capacity, the optimal workflow can be one-sided (i.e., either reviewing AI-positive cases to filter false alarms or reviewing AI-negative cases to rescue missed disease), before expanding to two-sided review as capacity grows. The optimal AI threshold generally differs from the AI-only benchmark because it determines not only sensitivity and specificity, but also which review protocol handles marginal cases. In a dynamic extension, current routing choices affect future clinician capability; preserving expertise can justify maintaining review on cases that appear unattractive in the short run. These results provide a framework for designing AI-enabled health screening workflows and coordinating threshold setting, staffing, and AI investment to balance safety, workload, and long-run human expertise.

3 - Dynamic Remote Patient Monitoring Under Patient Alert Fatigue

Jimmy Qin, The University of Texas at Dallas, Richardson, TX, United States, Wanting Yang, Jingwei Zhang

Remote patient monitoring (RPM) enables providers to observe patients' health signals outside the clinic and to intervene proactively when deterioration is suspected. In practice, however, wearable sensors can be imperfect, and unnecessary interventions can impose logistical and psychological burdens that erode patient compliance. We study a dynamic scheduling problem in which a provider monitors a panel of patients, updates probabilistic health beliefs from imperfect RPM signals, and decides whom to intervene under limited capacity. Proactive intervention reduces the risk of deterioration and improves short-term health outcomes, but over-intervention induces patient fatigue, increasing future noncompliance and undermining long-run health outcomes. We formulate the problem as a Partially Observable Markov Decision Process (POMDP) and the objective is to maximize quality-adjusted life years (QALYs). We characterize structural properties of the optimal value function and establish threshold-type optimal policies. For larger capacities, we propose a tractable Lagrangian relaxation and truncation policy with a provable performance bound. Numerical experiments quantify the value of RPM information and show that ignoring endogenous fatigue can substantially degrade performance, in magnitude comparable to a large worsening in sensor accuracy.

4 - All You Can Treat: Improving Hepatitis C Treatment Access with Refundable Subscription Contracts

Zhaowei She, Singapore Management University, Singapore, Singapore, Huaiyang Zhong, Yueran Zhuo, Turgay Ayer, Jagpreet Chhatwal

For decades, hepatitis C virus (HCV) infection was considered incurable, leading to end-stage liver diseases and substantial morbidity and mortality worldwide. The introduction of highly effective direct-acting antivirals transformed HCV into a curable disease, creating a realistic opportunity for elimination. However, access to these therapies remains severely constrained by high drug prices under the prevailing pay-per-prescription (PPP) contract. In response, Medicaid agencies of Louisiana (LA) and Washington (WA) adopted modified subscription drug contracts (MSDCs) for HCV drug procurement. Under MSDCs, states pay a fixed per-prescription price up to a capped spending threshold, after which additional treatments are provided without charge. However, despite adopting similar contracts, the two states experienced sharply divergent outcomes: LA

saw a substantial increase in treatment volume, while WA did not. Using a mechanism design approach, we show that MSDCs improve access only if demand uncertainty is low, as in LA's case, but fail to outperform PPP under high demand uncertainty, as in WA. To address this limitation, we propose a Refundable Subscription Contract (RSC), a subscription contract that includes a partial refund when realized demand falls below a pre-determined threshold. We prove that the RSC weakly Pareto-dominates PPP and MSDCs for any level of demand uncertainty, delivering higher treatment access without compromising manufacturer revenue. Finally, using Medicaid data, we calibrate our model for all 50 U.S. states and tailor the RSC design accordingly. Our numerical analysis indicates that adopting the RSC could increase treatment volume nationwide by approximately 80%.

Wednesday, July 29, 2026, 8:00 AM - 9:15 AM

WA01

301A

Identifying Inefficiencies in Outpatient and Primary Care

Invited Session

AI in Healthcare

Chair: Katherine Bobroske, Morgan Health, Falls Church, VA, United States

1 - Equal Access, Unequal Outcomes: How Elective Surgery Scheduling Produces Health Inequality

Tom Pape, University College London, London, United Kingdom, Nicos Savva, Stefan Scholtes, Christof Kastner, Katy Horder

This paper demonstrates how standard scheduling processes for elective surgeries can inadvertently produce health inequalities across social groups. Using data from 166,040 prostate resection surgeries in England between 2009 and 2019, we find that patients from the 20% most socially deprived neighborhoods wait as long as others for surgery (access equality) despite having a 36% [95% CI: 25%-48%] higher hazard of condition-specific emergency complications. This constitutes a scheduling-induced health inequality, because time in the queue exposes deprived patients to higher accumulated complication risk. Investigating the mechanism for this health inequality, we show that deprivation does not operate through specific diagnosable health indicators that clinicians could use to reprioritize individual patients along the clinical pathway. Rather, it acts through diffuse, non-specific health deterioration that shifts the entire risk distribution for the deprived population by a small amount per person. This is an instance of the prevention paradox: deprivation information is a poor predictor at the individual level (adding it to a clinical risk model improves predictive accuracy by only 0.14%) but a large driver of unequal health outcomes at the population level. Since clinicians cannot meaningfully act on this information for any individual patient, the inequality can only be addressed through the operational design of the scheduling process itself. We propose a delay-dependent priority queue that systematically adjusts scheduling priorities at the population level, and use data from the case study to illustrate how this queuing model can help decision makers navigate the tradeoff between access and health equality in practice.

2 - Vehicle Routing Optimization

Ben Murphy, SAS, Cary, NC, United States

Across industries, organizations must efficiently deliver valuable resources to critical destinations, often under strict time constraints. Yet many still rely on manual, multi-step processes that make it difficult to consistently identify the best routes and highest-priority stops.

Vehicle route optimization addresses this challenge by systematically identifying, evaluating, and selecting optimal destinations, then determining the most efficient path through them. In this demo, we walk through a solution built for healthcare service delivery. Our approach uses community detection to surface candidate destinations, combines heuristics with linear programming to pinpoint the ideal set of stops, and applies a traveling salesman formulation to determine the optimal vehicle route. A third-party API integration provides

turn-by-turn directions for last-mile logistics. Results are surfaced through interactive reports and maps, making the solution accessible to users of all technical backgrounds.

Join us as we demonstrate how this approach improves operational efficiency, reduces costs, and most importantly, ensures that scarce, life-critical resources reach the people who need them most.

3 - How Observable are Intervention-Oriented Patient-Reported Healthcare Access Barriers in Emr Data? Evidence from the Nih All of US Research Program

Shuwei He, Carnegie Mellon University, Pittsburgh, PA, United States, Holly Wiberg

Healthcare access barriers are multidimensional and often intervention-relevant, yet health systems lack scalable tools to identify which barriers can be inferred from routine clinical data and which require direct patient-reported measurement. This study develops a reproducible framework linking patient-reported healthcare access barriers to electronic medical record (EMR) activity to assess the empirical observability of distinct barrier domains in structured clinical data.

Using linked survey and EMR data from the NIH All of Us Research Program, we construct binary indicators for seven intervention-oriented barrier domains derived from Healthcare Access and Utilization survey items: lack of health coverage, care navigation or usual source of care barriers, financial burden, transportation barriers, scheduling or logistics barriers, interpersonal or relational barriers, and health literacy barriers. For each participant, we define a survey-anchored index date and extract temporally aligned EMR predictors from a 365-day lookback window within the OMOP Common Data Model. We evaluate standardized predictive pipelines using high-dimensional direct EMR representations, interpretable proxy features, and a combined feature set.

In the analytic cohort (N=195,734), financial burden and interpersonal or relational barriers were most prevalent. Predictability varied systematically across domains: affordability-related barriers were substantially more observable in EMR data than experiential barriers, while logistical barriers exhibited intermediate predictability. Strong co-occurrence patterns indicated multi-domain vulnerability.

These findings show that EMR-based inference can support scalable identification of certain intervention-relevant barriers but also reveals structural blind spots in routine documentation. We propose a hybrid measurement strategy in which EMR-derived proxies complement, rather than replace, patient-reported surveys.

4 - Detecting Undiagnosed Women's Health Conditions Using Medical Claims

Rowana Ahmed, JP Morgan Chase, New York City, NY, United States, Katherine Bobroske, Amy Liu

Women's health conditions are among the most prevalent yet systematically underdiagnosed categories of disease globally. Conditions, such as Polycystic Ovary Syndrome (PCOS), endometriosis, thyroid disorders, and autoimmune diseases, disproportionately affect women and have lifelong health consequences.

For example, PCOS affects 10–13% of women, with prevalence increasing ~25% over recent decades. It is associated with a 3–7x increased risk of type 2 diabetes, 1.7–2.5x increased risk of cardiovascular disease, 4–5x increased risk of endometrial cancer, 4–6x increased risk of anxiety and depression, and up to 15x greater likelihood of infertility, making it the leading cause of non-age-related female infertility. Women with PCOS also face a 2–3x increased risk of pregnancy-related complications, including gestational diabetes, hypertension, preeclampsia, and miscarriage. Endometriosis affects approximately 10% of women of reproductive age and is associated with chronic pelvic pain, impaired quality of life, and elevated infertility risk.

Despite this burden, these conditions are routinely missed or diagnosed only after years of symptoms. PCOS remains undiagnosed in ~70% of affected women, while endometriosis carries an average diagnostic delay of nearly a decade. Advances in machine learning present an opportunity to close this gap through cost-effective screening tools leveraging demographic and medical claims data. Clustering and graph-based methods can detect symptom patterns and comorbidities associated with undiagnosed conditions, and a predictive screening algorithm can be incorporated into clinical workflows to prompt confirmatory testing through ultrasounds and

metabolic panels. Early diagnosis is associated with improved outcomes, enhanced quality of life, and lower long-term costs.

WA02

301B

AI and Analytics for Decision-Making in Healthcare

Invited Session

Applied Probability

Chair: Mohammad Zhalechian, Kelley School of Business, Bloomington, IN, United States

1 - Aligning Minds and Machines: Calibrating Clinical AI with Human Perception

Michael Lingzhi Li, Harvard Business School, Boston, MA, United States, Agni Orfanoudaki, Scott Wang

Growing evidence from empirical studies focused on human–AI interactions has shown that algorithmic recommendations can sometimes undermine decision quality when misaligned with human judgment. In this paper, we develop a rigorous calibration framework that addresses this challenge by reconciling algorithmic risk estimates with human risk perception. Our approach decomposes the calibration process into two complementary strategies: *recentering*, which adjusts the absolute scale of the algorithm's output, and *alignment*, which fine-tunes the relative calibration between human and algorithmic judgments. Through a rigorous quantitative framework, we characterize the nonlinear effects of these adjustments on overall decision accuracy, and identify the conditions under which recalibration yields significant performance gains for the hybrid human-algorithm decision making process. In collaboration with a major academic medical center, we empirically validate our approach in the context of pediatric antibiotic prophylaxis for vesicoureteral reflux. We leverage data a combination of data from a randomized clinical trial, retrospective electronic health record, prospective survey studies with expert medical practitioners. Our results indicate that calibrated algorithmic recommendations substantially improve treatment decision quality compared to uncalibrated models. These findings offer a robust, data-driven strategy for managers to dynamically integrate data-driven models into decision processes, thereby enhancing the efficacy of hybrid human–algorithm systems in high-stakes environments.

2 - The Value Partial Information in Sequential Medical Scanning Decisions

Ali Kaan Kurbanzade, Kelley School of Business, Indiana University, Bloomington, IN, United States, Agni Orfanoudaki, Mohammad Zhalechian

In medical decision-making, practitioners often face trade-offs between waiting and conducting diagnostic scans, particularly in areas such as mammography, biopsy, and CT scanning. Existing models typically assume a binary action space—either waiting or performing an expensive full scan. In this study, we extend these models by introducing additional decision options, including partial scans, to assess their impact on sequential decision-making. Partial scans are assumed to be less costly than full scans but also less accurate. However, they provide more reliable information than simply waiting. We build upon existing POMDP frameworks by expanding the action set beyond the traditional binary choices and present theoretical insights into the value of partial information. To validate our findings, we apply our model to a novel dataset from a partnering hospital, focusing on subarachnoid hemorrhage diagnosis and management.

3 - Learning when to Nudge: a Dual-Agent Bandit Framework for Behavioral Interventions

Xueqing Liu, Harvard University, Boston, MA, United States, Esmaeil Keyvanshokooh, Kyra Gan, Yongyi Guo, Anna Trella, Susan Murphy

In many intelligent systems, decision-making is shared between two agents: a recommender agent that selects continuous interventions and a nudging agent that controls access to information via limited, costly prompts for user engagement. This structure creates a key information asymmetry, where the recommender operates with delayed data, and its performance depends on the nudging agent's strategic timing of pro-treatment actions—user behaviors that provide no immediate benefits but are crucial for the algorithm's functionality. This paper studies the challenge of optimizing synergy between these agents under limited nudging budgets. To address this

problem, we propose a dual-agent modeling approach involving: (i) a recommender agent that makes intervention decisions based on revealed user history, and (ii) a nudging agent with richer, real-time data that decides when to prompt engagement. To establish proper synergy, we develop the Dual-Agent Bandit Framework for Behavioral Interventions (**DABBI**) that seamlessly integrates optimization-based nudging mechanisms with contextual upper-confidence-bound procedures. We prove **DABBI** achieves sublinear regret across practical settings, including noiseless, noisy, latent, and partially observed contexts. We further establish a minimax lower bound for our dual-agent bandit with limited nudging and show **DABBI** matches this bound. We evaluate our theory using synthetic data and a mobile-health case study.

4 - Optimal Assignment Guidelines for Maternity Waiting Homes in Resource-Constrained Environments

Yuhang Du, London Business School, London, United Kingdom, Qi (George) Chen, Jeremie Gallien

Maternity waiting homes (MWHs) are increasingly used to increase skilled birth attendance in regions with scarce obstetric services. We seek to optimize their assignment policy, that is whether and when a pregnant woman should go to her local MWH given available accommodation capacity and individual characteristics including living location. Our model leverages clinical data on labor onset and latent phase duration to minimize the predicted number of unassisted deliveries resulting from such policy in the local population. We present explicit mathematical solutions for different variants of this problem and integer program formulations of discretized approximations capturing practical implementation constraints. Our numerical experiments leverage detailed primary and secondary data on the population and health facilities of Bong County, Liberia. They suggest that our proposed pooling and assignment policy interventions for the existing MWH capacity in that region may plausibly halve the incidence rate of unassisted deliveries and reduce the probability of unassisted deliveries for the most vulnerable women from 70% to 15%. Implementing data-driven individual assessments of unassisted delivery risk and MWH assignment policies may thus constitute impactful and low-cost interventions in regions with limited MWH accommodation capacity.

WA03

302A

The Appointment Access Problem

Invited Session

Healthcare Operations Management

Chair: Daniel Adelman, The University of Chicago, Booth School of Business, Chicago, IL, United States

1 - Key Drivers of the Appointment Access Problem

Daniel Adelman, The University of Chicago, Booth School of Business, Chicago, IL, United States, Ken Nunes

Together with University of Chicago Medicine, the Healthcare Analytics Laboratory at Chicago Booth has identified six operational drivers of appointment access problems. At their core is a critical concept that is missing in the management practices of hospitals. We introduce this concept, and use it to demonstrate the operational drivers as seen in data. It opens up the opportunity to develop novel solutions and methodologies to address the problem of improving access to appointments for patients. We'll introduce these, and they will then be discussed in greater detail subsequently in this session.

2 - Tracing Patient Pathways with Provider Episodes: Insights into Access and Censoring

Duygu Soylemez, The University of Chicago Booth School of Business, Chicago, IL, United States, Daniel Adelman, Kenneth Nunes

We introduce a novel concept, termed the provider episode, that captures patient pathways for different medical conditions across specialties, enabling a more systematic understanding of access dynamics in healthcare systems. In addition, we develop a complementary approach to address limitations arising from incomplete observations, allowing for more robust analysis of partially observed data.

3 - Dynamic Capacity Release in Appointment Scheduling

Esma Akgun, University of Chicago Booth School of Business, Chicago, IL, United States, Daniel Adelman

Most outpatient clinics employ predefined appointment scheduling templates that allocate capacity across patient types in advance. While such templates aim to balance competing demands, they often remain rigid in the face of uncertain and time-varying patient arrivals. This rigidity can lead to underutilized provider capacity and reduced access to timely care. A scheduling strategy to mitigate this issue and introduce flexibility into otherwise static templates is dynamic release, in which appointment slots reserved for one patient type are made available to other patient types as the appointment date approaches if unfilled. The timing of release decisions involves a key trade-off. Releasing capacity too early risks crowding out future demand from the target patient group, while releasing it too late may result in idle slots or reduced access for other patients. In this work, we study the problem of dynamically reallocating appointment capacity through release timing decisions and develop a framework to determine the optimal timing for releasing capacity across patient types to improve access and utilization.

4 - Optimizing Appointment Access Across a Large Academic Health System

Leena Ghrayeb, The University of Chicago, Booth School of Business, Chicago , IL, United States, Esma Akgun, Duygu Soylemez, Daniel Adelman

Managing patient access to appointments is a critical challenge for all healthcare systems, requiring a balance between timely access to care for patients and efficient utilization of provider time and system resources. As demand for healthcare continues to grow, systems must adapt to accommodate increasing and variable patient needs. An ongoing struggle that health systems face is the alignment of patient demand with the system's resources, especially given geographically dispersed clinics, provider availability, and referral patterns. This challenge is compounded by limited visibility into how demand is connected across specialties and departments, making it difficult to anticipate the downstream effects of capacity allocation decisions. In this work, we develop and analyze a math programming model that captures key system complexities, including patient choice behavior and referral flows between departments, with the overall objective of maximizing the total number of new patient visits across a large academic health system.

WA04

302B

Pharmaceutical R&D and Supply Chains: Theory and Practice

Invited Session

Healthcare Operations Management

Chair: Efthymia Kostaki, University of Cambridge, Cambridge, United Kingdom

1 - Incentivizing Drug Rediscovery: Exclusivity Extension and Priority Review for Repurposed Therapeutics

Elodie Adida, University of California - Riverside, Riverside, CA, United States, Fernanda Bravo

Drug repurposing consists in identifying new indications for existing drugs. Compared to *de novo* drug development, drug repurposing is faster and incurs less risk and costs. As a result, it can improve patients' access to new treatments. However, it is currently under-utilized, primarily due to a lack of sufficient financial incentives to invest in the approval of new indications. A new drug enjoys only a limited time with protected market exclusivity. Once the exclusivity period is over, the drug is often subject to competition with a generic drug, and the potential for profits drops sharply, making investment in developing new indications often cost-ineffective. While a new indication is usually granted a temporary exclusivity period, it is often too short to recoup investment costs.

We study a combination of two mechanisms to better incentivize drug repurposing. We consider (i) granting an extension of the exclusivity period for the primary indication upon approval of a new (secondary) indication, and

(ii) offering priority in the review process to speed up approval of the secondary indication. We investigate how to best design the intervention and the effect on patients' welfare.

We find that while both mechanisms can be effective in improving the incentive to engage in drug repurposing, they may hurt the patient welfare. When both mechanisms can incentivize investment, priority review is more beneficial to patient welfare. However, in some cases, it is optimal for the patient welfare to combine the two mechanisms, and thus, they can be viewed as complementary.

2 - Evaluating the Efficacy of Providers Compensation Contracts in Improving Participant Retention for Clinical Studies

Xueze Song, The University of Alabama, TUSCALOOSA, AL, United States, Mili Mehrotra, Tharanga Rajapakshe

In this work, we aim to analyze a clinical study sponsor's decisions regarding monetary payments to participants and compensation for providers (investigators and coordinators) for their efforts to improve participant retention in the study. To this end, we first consider a centralized model where the sponsor decides the monetary payments to participants and the providers' efforts. We then identify the optimal contracts for the providers under two decentralized team structures. We further analyze three widely adopted compensation contracts for the providers.

3 - Valuation and Portfolio Theory for Biomedical Investments

Chaoyi Zhao, Massachusetts Institute of Technology, Cambridge, MA, United States, Andrew Lo

Biomedical investments are characterized by large upfront costs, sequential capital allocation, strong cross-project interdependence, and highly skewed, often binary, payoff structures. In practice, investors and firms must evaluate not only the value of individual projects, but also how groups of projects perform as portfolios. In this talk, we bring these two perspectives together in a unified framework.

We first consider how to value biomedical projects when outcomes are linked through shared science, common markets, or strategic interactions. These connections matter for real investment decisions, especially when funding is staged over time and information arrives gradually.

We then examine portfolio construction in biomedical venture investing, where returns are highly uneven and success is often judged by whether investments achieve attractive payoff multiples. In this setting, standard portfolio ideas do not always work well. Diversification can improve conventional measures of risk while still reducing the chance of reaching a desired investment multiple. This tension is especially important in portfolios of projects with low probabilities of success but substantial upside.

By linking project valuation and portfolio design, we offer a broader framework for thinking about biomedical investments, with implications for venture capital, pharmaceutical R&D, and other innovation-driven settings shaped by uncertainty, interdependence, and sequential decision-making.

4 - The Impact of Investigator Experience on Clinical Trial Enrollment

Efthymia Kostaki, University of Cambridge, Cambridge, United Kingdom, Feryal Erhun, Nektarios Oraiopoulos

The median R&D cost to bring a new drug to the market is almost $1 billion. Delays in clinical development are therefore costly: each day of trial delay has been estimated to cost between $600,000 and $8 million, and can even lead to loss of first-to-market advantage, while also postponing access to new therapies for patients. A broad consensus has emerged that accelerating clinical development hinges critically on speeding up and streamlining patient enrollment. In this paper, we examine the impact of chief investigator (CI) experience on trial duration. We measure CI experience as the number of prior trials in which the investigator has served as CI before the calendar year in which the focal trial is registered on ClinicalTrials.gov. Our empirical analysis shows that increased CI experience is associated with shorter enrollment periods regardless of the level of prior experience as a CI: having an experienced CI leading the trial is associated with a reduction in enrollment duration of approximately 10%, on average. We also show that very high Principal Investigator (PI) experience,

lower drug novelty, prior sponsor–CI collaboration, and whether the CIs are based at non-academic sites are each associated with meaningfully shorter enrollment durations.

WA05

302C

Equity-Informative Health Policy Simulation

Invited Session

Public Health Policy

Chair: W. Alton Russell, McGill University, Montreal, QC, Canada

1 - Factors Contributing to Lower Rates of Blood Donation from People with African Ancestry

W. Alton Russell, McGill University, Montreal, QC, Canada

The most common condition that requires chronic blood transfusion is sickle cell disease, a genetic disorder that effects people with African ancestry. In many countries, the per capita rate of blood donation is lower among people with African ancestry than other groups. Increasing the representation of genetically diverse donors, particularly donors with African ancestry, is critical to ensure the availability of rare blood types that some chronically transfused patients need. This project uses simulation modelling to understand how factors like new donor arrival, donor return, and donor ineligibility contribute to differences in blood donation rates between Black and non-Black donors, an imprecise proxy for ethnic ancestry. By estimating separate models by racial group and evaluating different combinations of parameters, our analysis quantifies the contribution of specific parameters to the overall gap in per capita blood donation rates. Our work aims to inform strategies to improve blood donation rates among underrepresented racial groups, thereby increasing the availability of rare blood types for chronically transfused patients.

2 - Adapting a Microsimulation Model of Cervical Carcinogenesis to Reflect Black Women in the United States.

Jennifer Spencer, University of Texas at Austin, Austin, TX, United States, Emily Burger, Nicole Campos, Mary Caroline Regan, Stephen Sy, Jane Kim

To understand cervical cancer incidence and mortality disparities by race, we adapted a microsimulation model of HPV natural history in all women in the U.S. to specifically reflect Black women. We integrated race-stratified demographic, epidemiological and screening data for U.S. Black women into a general population model of HPV and cervical cancer, including age- and genotype-specific HPV prevalence from the 2002-2006 National Health and Nutrition Examination Survey and stage- and age-specific cancer survival from the Surveillance, Epidemiology, and End Results Program. Age and race-specific cervical cancer screening patterns were estimated from the METRICS sites of the PROSPR consortium. We used sequential replacement of variables to estimate the impact of racial disparities across the cervical cancer continuum. Overall, Black women in our model had a 1.24 [1.23-1.27] relative risk (RR) of cervical cancer incidence compare to all-race women and 1.92 [1.85-2.00] RR for cervical cancer mortality. Sequential replacement found differences in follow-up after positive screen explained the largest share of incidence disparities (47%) followed by screening frequency (33%) while survival was the largest explained portion of mortality disparities (70%) followed by diagnostic follow-up after a positive screen (12.7%).These findings identify diagnosis and treatment as important areas of care for addressing racial disparities in cervical cancer outcomes.

3 - A Shapley-Based Simulation Framework to Decompose Long-Term Health Disparities

Jinyi Zhu, Vanderbilt University, Nasvhille, TN, United States, Alyssa Bilinski, Jennifer Spencer

Background: Existing empirical data on health disparities primarily reflect short-term inequalities, limiting our ability to understand longer-term consequences. Disease simulation modeling offers a unique approach to quantifying long-term disparities by integrating diverse data and projecting outcomes beyond existing time frames. We propose a framework to quantify and decompose long-term disparities, illustrated through rural-urban stroke disparities.

Methods: This simulation-based framework begins by specifying subgroup-specific parameters (such as risk factor prevalence and intervention uptake) from epidemiologic studies, administrative data, and clinical trials. Separate simulations are then conducted for each subgroup to estimate long-term outcomes and quantify disparity. Next, stepwise decomposition is performed by systematically replacing the disadvantaged group's parameters with those of the advantaged group. To ensure order-independent attribution, we average each parameter's marginal contribution across all possible replacement sequences, yielding a Shapley-value decomposition. We illustrated this approach by adapting a published stroke microsimulation model to estimate quality-adjusted life-years [QALY]-based rural-urban disparities among US adults aged 40–80 at risk for incident ischemic stroke.

Results: We estimated a stroke-associated rural-urban gap of 0.253 QALYs per person. Pre-stroke factors were the largest contributors, with hypertension contributing 28.5% and smoking 21.3%, followed by inpatient rehabilitation (15.2%). Acute treatment differences contributed less, with IV tPA accounting for 4.9%.

Conclusions: This framework is generalizable to a wide range of health disparities in other clinical or policy contexts. By quantifying subgroup differences and identifying the key drivers of observed disparities, it offers a systematic basis for prioritizing targeted interventions and informing policy decisions to reduce health inequities.

4 - Model Incorporating Adolescent Drinking and Care Impact on Long-Term Aud Outcomes.

Jovan Julien, Georgia Institute of Technology, Atlanta, GA, United States

Background Children and young adults represent one of the more vulnerable populations to substance use disorders in the United States. Adverse childhood experiences—abuse, neglect, or exposure to violence or parental substance use—are well-documented risk factors for disorders ranging from ADHD and PTSD to mood, adjustment, and substance use disorders. Outcomes of interest can include alcohol consumption and alcohol use disorder (AUD) development.

Methods:

Alcohol consumption is a major modifiable risk factor for substantial morbidity and mortality, contributing to liver cirrhosis, cancer, injury, cardiovascular disease, and other alcohol-attributable health outcomes. To evaluate the long-term population impact of alcohol consumption patterns and potential interventions, we developed an individual-level microsimulation model that first focused on alcohol-associated liver disease and liver cirrhosis progression. Using age and sex-specific drinking trajectories, the model evaluates how intervention strategies that reduce risky drinking or improve treatment engagement may change future alcohol-related morbidity and mortality.

Outcomes include annual and cumulative deaths by cause, sex, age group, and intervention scenario. This stepwise modeling framework provides a flexible approach for evaluating alcohol harm-reduction strategies across multiple health outcomes and populations. By connecting drinking behavior, intervention effects, liver disease progression, and broader cause-specific mortality, this work supports more comprehensive public health decision-making for reducing alcohol-related harms in the United States.

Opportunities: Two system-level intervention are compared: **high-intensity intervention** (intensive case management, wraparound services, or high-frequency MH service delivery) and **status quo**.

Conclusions This framework advances old models to explore a present day issue with limited modern data availability.

.

WA06

303

Medical Decision Making in Aging Society

Invited Session

Medical Decision Making

Chair: Yvonne Huijun Zhu, Pennsylvania State University, University Park, PA, United States

1 - Service Design and Innovation for Aging Societies: Introducing Silver Prosperity Index (Spi)

Rohit Verma, University of South Carolina, Darla Moore School of Business, Columbia, SC, United States, Kejia Hu, Xin Ding, Diep Pham

This presentation introduces the Silver Prosperity Index (SPI), a comprehensive, data-driven framework designed to evaluate service innovation in aging societies. Older adults are active economic contributors, influential consumers, and key drivers of the $15 trillion service economy and SPI emphasizes the expanding socioeconomic and digital opportunities tied to population aging.

The SPI is built on four interconnected pillars: Economic Empowerment, Health and Well-being, Social Cohesion and Connectivity, and Security and Protection. Methodologically, it combines primary survey data with national secondary datasets. The framework is demonstrated through a 2025 application across all 50 U.S. states, enabling detailed, localized analysis.

Findings reveal significant regional disparities in how states support aging populations. These gaps point to underutilized opportunities in areas such as digital inclusion, workforce participation, and community resilience. Addressing these disparities could unlock substantial potential for innovative services tailored to older adults.

2 - Forecasting Statewide Emergency Medical Events: Comparing Statistical and Hybrid TIME Series Forecasting Methods

Alina Hagen, School of Information, University of South Florida, Tampa, FL, United States, Hongdao Meng, Xue Guo, Mingyang Li, Mohammad Al Olaimat, Lonie Hagen, Lindsay Peterson, Michael Lozano, Bruce Moeller

Emergency Medical Services (EMS) represent an essential public safety infrastructure for acute healthcare. Population growth and aging, particularly the disproportionately high EMS utilization among adults over 65 years, suggest that the demand for EMS will grow at rates that cannot be met by workforce expansion alone, underscoring the need for evidence-based resource planning. Using data from the state-wide EMS reporting system in Florida (2017-2025), which spans the COVID-19 pandemic and its attendant EMS utilization pattern disruption, we generated monthly state-wide and county-specific EMS event projections from 2026-2050. We examined the performance of four models: SARIMA, GARCH, Unobserved Components Model (UCM), and Prophet. Data from 2025 were used for validation and 95% confidence intervals quantified uncertainty surrounding annual event rate predictions. Results reveal substantial difference in long-term projections across models. SARIMA produced the most aggressive growth estimates, with some counties projecting growth rates exceeding 200%, with extreme outliers above 400%, likely reflecting unconstrained trend extrapolation. In contrast, the conditional mean-based GARCH model yielded more conservative and stable central estimates, projecting statewide incidents rising from approximately 3.6 million to 5.2 million by 2050, with some counties showing slight declines. UCM and Prophet models offered intermediate solutions, capturing seasonal patterns and structural change while producing interpretable trajectories. Across models, Mean Absolute Percentage Error (MAPE) in 2025 validation data was generally below 10%, though forecast reliability varied by county, reflecting regional differences in data quality and population dynamics. These findings support model-informed forecasting as a tool for EMS workforce planning and policy research.

3 - Does Virtual Take Longer? An Empirical Investigation of Service TIME in Telemental Health

Yupu Sun, University College London, London, United Kingdom, Ersin Korpeoglu, Wei Miao

Telemedicine visits, in the form of synchronous virtual interactions between patients
and providers, have witnessed a significant surge during and after the COVID-19 pandemic. Yet, little is known about the operational effects of using telemedicine. In this study, we examine the causal relationship between the

form of psychotherapy treatments (virtual or in-person) and the service time of these treatments. Using visit-level patient data for psychotherapy visits in the United States (U.S.), we employ an instrumental variable approach to estimate the causal impact of telemedicine use on monthly service time. We find that telemedicine use significantly increases monthly service time, driven by both higher
visit frequency and longer service time per visit. These effects may reflect challenges and limitations in the use of digital technologies during telemedicine visits. Heterogeneity analyses further show that retirees experience greater increases in service time than non-retired patients, and treatment effects rise with patient age. These patterns could be attributed to a decline in technological literacy among older retirees. These results show that serving elderly patients, especially retired ones, with telemedicine can increase the burden on healthcare providers. Our study contributes to a better understanding of how telemedicine affects the efficiency of care delivery in mental health.

WA07

304

Healthcare Policy, Effectiveness, and Value

Invited Session

Health Applications Society

Chair: Melike Yildirim, Wayne State University, Detroit, MI, United States

Co-Chair: Shayan Sharifi, Wayne State University, Southfield, MI, United States

1 - Persistent Burden of Invasive Pneumococcal Disease in Children with Sickle Cell Disease

Ziyu Zhang, ISyE Georgia Tech, Atlanta, GA, United States, Melike Yildirim, Pinar Keskinocak, Inci Yildirim

Sickle cell disease (SCD) is the most common genetic hematologic disease globally and children with SCD are at increased risk for pneumococcal disease. We utilized data from population-based enhanced surveillance for invasive pneumococcal disease (IPD) in children <18 years of age in Massachusetts from 2002 to 2020. We calculated incidence rates (IR) among children with SCD using bootstrapping resampling and incidence rate ratios (IRR) for pre- and post-PCV13 periods. Vaccine effectiveness (VE) was calculated and PCV13 vaccine failure probability was predicted using a random forest model. Children with SCD had higher IR during both pre−/post-PCV13 periods compared with otherwise healthy children. A modest reduction of 28.1 % in the incidence of overall IPD during the post-PCV13 period was observed in children with SCD, whereas a more substantial 59.5% reduction was observed in otherwise healthy children. There was a 60.8% reduction in the incidence of VST13 IPD in children with SCD and an 83.0% reduction in children without underlying health condition. Overall, 61.1% of the remaining IPD among children with SCD were due to non-PCV13 serotypes, many of which are included in expanded valency vaccines. Children with SCD continue to have higher rates of IPD compared with otherwise healthy children despite vaccination. Majority of the remaining disease is due to serotypes not included in vaccine formulations that have been used for the last two decades. Our study highlights the potential value of expanded valency vaccines and importance of risk-based vaccination strategies tailored for this vulnerable population.

2 - Process Measures Reveal What Quality Measure Can't: a Markov-Model-Based Approach to Understanding Hospital Care Delivery

Inas Khayal, Wayne State University, Novi, MI, United States, Aidan Campbell

Quality measures summarize hospital performance into aggregate rates but provide limited insight into which operational changes produced observed improvements. We construct a Markov model of hospital care processes —comprising sequential transitions between specific clinical functions at specific locations—from the same Medicare claims data used to derive quality measures. Both the Markov transition probabilities and the quality measure rates were estimated by analogous Bayesian inference pipelines to identify credible changes at 541 hospitals between 2016 and 2019, and their association was measured using mutual information. Hospitals that improved on the same quality measure did not all change in the same way: among those that increased the advanced care decision quality measure, those that also increased with respect to the same advanced care decision measure, but calculated only for events within 14 days of death had markedly higher rates of emergency

department-to-inpatient advance care planning routing. The principal associations were interpretable, stable across three credibility thresholds, and stronger when transitions were considered jointly. This framework–integrating a systems-based representation of healthcare delivery with Markov modeling–advances data-driven methods for identifying operational pathways underlying quality improvement and for informing quality-improvement decision-making.

3 - United States Implementation of Human Papillomavirus Self-Sampling for Cervical Cancer Screening: a Budget Impact Model

Claire Mellott, University of Michigan, Ann Arbor, MI, United States, Diane Harper, David Hutton

Human papillomavirus (HPV) self-sampling for cervical cancer screening was recently approved in the US and is anticipated to increase the screening rate by providing a preferable option for individuals with a cervix. The goal of this US budget impact model was to examine the 5-year economic and health outcomes of implementing self-sampling for cervical cancer screening. The model used a Markov structure for health state transitions and a decision tree structure for estimating resource use. It included costs for lab tests, healthcare visits, follow-up care after screening, and cancer treatment. Base-case analysis assumed 32% of cervical cancer screening would use self-sampling and that the screening rate would increase 8% over 5 years. Uncertainty analyses were conducted.

The model estimated that total costs to the US healthcare system would increase $2.2 billion (discounted) to screen 8.4 million more eligible individuals over the 5-year time horizon. In year 5, the cost is $6,518 more per 1,000 individuals screened. The number of cervical cancer cases and deaths prevented would increase by 1,782 and 87, respectively. The largest increase in costs was from colposcopies ($1,634 million) and largest decrease from healthcare visits (-$660 million). In a scenario analysis that assumed the screening rate would remain at its current level, total costs over 5 years would increase by $1.5 billion. Implementing HPV self-sampling would increase overall healthcare costs and the number of prevented cervical cancer cases and deaths, by increasing the cervical cancer screening rate to meet the US Healthy People screening goal.

WA08

305A

Operational Analytics for Early Detection and Care Pathways in Resource-Constrained Systems

Invited Session

Global Health and Humanitarian Operations

Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States

1 - A Dynamic Appointment Scheduling Algorithm to Prioritize High-Severity Renal Colic Patients for Early Follow-up

Daniela Santibanez, North Carolina State University, Raleigh, NC, United States, Osman Ozaltin

Post–emergency department (ED) care for renal colic is often delayed and fragmented, contributing to preventable revisits and missed opportunities for timely intervention. At a large quaternary referral center, fewer than 25% of patients discharged with renal colic received ambulatory follow-up within 60 days, largely due to limited clinic capacity and the absence of standardized triage workflows. To improve timely access, we developed a primal-dual based scheduling algorithm that integrates patient severity, clinic capacity, and predicted ED revisit risk. We evaluated whether this approach could enhance scheduling efficiency and reduce delays in follow-up care.

Encounter-level data from renal colic ED discharges (2023–2024; n=1140) were used to develop survival models estimating revisit probabilities. Predictors included age, stone size, pain score, race/ethnicity, and diagnostic findings from the index encounter. These predictions informed an online scheduling model that dynamically assigns patients to limited urgent urology clinic slots while minimizing expected revisit risk, subject to weekly capacity constraints. Model performance was evaluated for schedules generated over a three-month horizon (May-July 2024) under various fixed-capacity scenarios. Outcomes assessed included time to follow-up and the

mean expected revisit probability across scheduled patients. Results are compared with schedules generated by a benchmark heuristic and the offline optimization model.

The scheduling framework suggests the potential to improve access to timely post-ED care for renal colic patients and lower preventable ED revisits. Ongoing work focuses on validating optimized outcomes against current scheduling practices.

2 - Penalized Deep Partially Linear Cox Models with Application to CT Scans of Lung Cancer Patients

Yuming Sun, William & Mary, Williamsburg, VA, United States

Lung cancer remains a leading cause of cancer mortality worldwide, underscoring the need to better understand risk factors to guide patient-centered therapies. The National Lung Screening Trial (NLST) incorporates computed tomography (CT) texture analysis, which provides objective measures of imaging patterns associated with patient outcomes. Partially linear Cox models are well suited for survival analysis, as they decompose the hazard into parametric and nonparametric components, enabling joint modeling of established clinical variables and high-dimensional imaging features.
However, when the number of parametric covariates exceeds the sample size, model estimation becomes challenging, while nonparametric components suffer from the curse of dimensionality. To address these issues, we propose a Penalized Deep Partially Linear Cox Model (Penalized DPLC). Our approach integrates SCAD regularization for feature selection with a deep neural network to flexibly estimate the nonparametric component.
We establish theoretical guarantees, including convergence and asymptotic properties, and evaluate performance through extensive simulations, focusing on risk prediction and feature selection. Finally, we apply the proposed method to the NLST dataset to identify key clinical and imaging factors associated with survival. The results provide new insights into how these factors jointly influence lung cancer prognosis.

3 - Index Policies for Advance Scheduling

Xu Zheng, Miami Herbert Business School, Miami, FL, United States, Xu Sun, Christos Zacharias

We study an advance scheduling problem, where stochastic demand arrives over time and must be assigned to future service dates within a fixed-length booking window. The system manager seeks to minimize expected total cost, capturing both same-day processing pressure and the cost of deferring workload into the future. To address the resulting high-dimensional dynamic control problem, we develop a dual-based approach that exploits a Lagrangian relaxation to obtain a tractable separable reformulation, leading to a lightweight closed-loop marginal-index policy for sequentially allocating arrivals. We analyze the performance of this policy under a many-arrivals asymptotic regime with fixed time granularity and booking-window length, and establish tight regret bounds. The same decomposition-based perspective also extends naturally to several practically relevant variants, including service-capacity constraints with rejection and cancellations via thinning. Numerical experiments further demonstrate the effectiveness of the proposed policy.

4 - A Decision Framework for Early Inpatient Bed Requests in Emergency Departments

Aniruddhan Ganesaraman, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Qian Cheng, Nilay Tanik Argon, Serhan Ziya

Emergency department (ED) boarding, the practice of keeping admitted patients in the ED while awaiting inpatient beds, is a primary driver of crowding, poor outcomes, and excess cost. Under current practice, the Transfer Preparation Process (TPP) begins only after a formal admission decision is made. Initiating bed requests earlier, however, risks having prepared beds remain unoccupied for extended periods. We propose a framework that manages this trade-off directly. A centralized decision-maker periodically determines how many inpatient beds to request based on real-time ED census, patient admission probabilities, and current hospital bed availability. Aggregating requests at the population level rather than the patient level reduces prediction uncertainty.

The framework is formulated as an infinite-horizon Markov decision process with three cost components: a penalty for boarding patients, a penalty for unoccupied prepared beds, and a convex ordering cost that discourages large request batches. Three heuristic policies are derived: MDP-based (MHP), newsvendor-based (NHP), and deep Q-learning (DQN). We evaluate policies across a wide spectrum of boarding-to-idling cost

ratios in a validated high-fidelity discrete-event simulator calibrated to 2019 data from a large academic ED in North Carolina.

Under standard conditions, the framework reduces boarding time by 30 - 70% relative to current practice, with modest increases in bed idle time. Robustness is further confirmed under a pandemic-surge scenario characterized by elevated patient arrival rates. NHP most efficiently reduces boarding time for a given level of bed idle time; DQN achieves the lowest variability in request sizes, producing the most uniform bed request process.

5 - Hospital-Physician Integration and Quality Outcomes: the Role of Financial Incentive Programs Faced by U.S. Hospitals

Jingyun Su, Miami Herbert Business School, Coral Gables, FL, United States, Jong Myeong Lim, Masoud Kamalahmadi

Hospital-physician integration is the trend of the US hospitals management strategy. However, current literature shows mixed outcome for vertical integration on hospital quality of care. In this paper, we examine how financial incentive programs, such as the Hospital Readmissions Reduction Program (HRRP) and the Hospital Value-Based Purchasing (VBP) Program, alter the relationship between organizational structure and quality. We employ an IV approach to estimate the causal effect of vertical integration on quality outcomes in the presence of these financial incentives. We show that integration may enhance or attenuate quality responsiveness to financial incentives depending on how internal incentive transmission mechanisms align physician behavior with external performance goals. Specifically, we find that more highly integrated hospitals exhibit significantly better-quality performance following the announcement of CMS financial incentives compared to less integrated hospitals.

WA09

305B

Practical AI #3: Supply Chain and Resource Utilization in Hospitals

Invited Session

Healthcare Practice

Chair: Kristy Zhou, Memorial Sloan Kettering Cancer Center, New York City, NY, United States

1 - Clinically Integrated Supply Chain: Transforming Hospital Logistics

SAMMY NAMAN, Memorial Sloan Kettering Cancer Center, New York, NY, United States

Memorial Sloan Kettering Cancer Center (MSK) transformed its complex, multi-site internal logistics by replacing an opaque, risk-prone paper-based system with a comprehensive, clinically-integrated digital platform. Historically, manual tracking of patient-critical items like specimens, slides, and pharmaceuticals created significant operational inefficiencies, lacked real-time visibility, and compromised chain of custody, often resulting in anxious, time-consuming searches for misplaced items.

To overcome these challenges, the team developed a track and trace solution. This new process enables clinical staff to initiate shipments directly from within their EHR workflow, generating standardized, scannable labels. The transportation team uses mobile devices to scan items at every checkpoint, creating a real-time digital audit trail.

This system provides complete end-to-end visibility, forming a digital supply chain twin. It also incorporates an intelligent algorithm that assigns the most efficient transportation route for shipments based on schedule, location, and timing.

2 - Covid Resource Optimization at Cleveland Clinic

Subramanian Pazhani, SAS Institute, Cary, NC, United States, Natalia Summerville, Michelle Opp

This talk presents a medical resource optimization tool developed in collaboration with Cleveland Clinic to help healthcare organizations effectively manage resources during the COVID-19 pandemic. During the pandemic, hospitals suspended non-emergency visits and elective procedures to prioritize COVID-19 care. As demand for COVID-19 treatment began to decline, Cleveland Clinic faced the complex challenge of safely reopening facilities and services while operating under limited resources such as emergency rooms, ICU beds, ventilators, and testing capacity. The hospital network includes multiple facilities across locations, service lines such as cardiology, pulmonology, and neurology, and associated sub-services. The problem is formulated as a bi-criteria mixed-integer linear program that maximizes hospital revenue and margin, subject to resource and testing constraints. The optimization model is implemented using OPTMODEL on Viya. The optimization tool generates optimal reopening plans for hospital sub-services by balancing hospital economics, efficient resource utilization, and patient access to high-quality clinical care. The tool leverages decomposition and parallel scenario analysis to support short- and long-term, data-driven reopening decisions. In this talk, the tool's practical impact and value is demonstrated.

3 - Strengthening Supply Chain Resilience Through Data-Driven Recall Risk Management

Chirag Kanitkar, New York City Health & Hospitals, New York, NY, United States

Medical recalls often pose significant risks to patient safety, operational continuity, and financial sustainability. Managing these events at the intersection of Cost, Quality, and Outcomes (CQO) requires rapid reactive response and structured proactive risk assessment. Although automated alerts platforms are widely adopted, recall management remains largely fragmented, causing administrative latency, delayed item tracing, inconsistent patient impact assessment, and prolonged exposure to affected products.

This presentation introduces a scalable, data-driven framework transitioning healthcare operations from unstructured crisis response to a hybrid resilience model that balances effective reactive protocols with proactive risk-mitigation pathways. The framework integrates structural recall intake with Enterprise Resource Planning (ERP) and Electronic Medical Record (EMR) data using the Unique Device Identifier (UDI) to enable rapid item- and lot-level tracing, and patient-level visibility. Concurrently, it introduces a quantitative risk-scoring methodology – incorporating clinical severity, historical recall patterns and traceability – to identify high-risk products proactively, prioritize active recalls, and drive supply chain strategy. Furthermore, we explore the role of Artificial Intelligence (AI) in automating recall data extraction, normalizing product identifiers, strengthening data governance, and enhancing dashboarding capabilities.

Existing literature correlates recall response efficiency with clinical outcomes – including Length of Stay (LOS) and revision rates – and financial metrics such as recoverable asset value and uncompensated procedural costs. By enhancing response efficiency and enabling proactive mitigation, this framework provides a scalable blueprint for strengthening compliance with evolving federal quality mandates while reducing the national burden associated with recalled medical products and supply chain disruptions.

4 - Building An OR Block Management Ecosystem: a Practical Approach to Improving Utilization and Capacity

Kristy Zhou, Memorial Sloan Kettering Cancer Center, New York City, NY, United States

Ineffective operating room (OR) block management often leads to underutilized capacity and misalignment between surgical scheduling and downstream resources. We implemented a comprehensive block management ecosystem at a high-volume cancer center, including dedicated governance structures, a centralized decision-support application for ongoing block assessment, and complementary analytics and optimization tools to support reallocation, overbooking, and performance monitoring. Optimization-driven block reallocation smoothed inpatient demand, reducing peak capacity needs by the equivalent of 10 beds, while an overbooking strategy increased OR utilization by 6–8% without additional rooms needed. The ecosystem enabled consistent, transparent, and system-aligned decision-making while incorporating surgeon preferences and operational constraints. This presentation will highlight key design principles, implementation strategies, and lessons learned to support adoption of scalable block management practices in complex surgical environments.

5 - Beyond Accuracy: Translating ML-Driven Surgical Scheduling into Operating Room Impact

Kai Okada, Memorial Sloan-Kettering Cancer Center, New York, NY, United States

Accurate surgical case length prediction directly impacts Operating Room utilization, patient experience, and health system finances — yet convincing stakeholders to adopt a predictive tool for scheduling thousands of annual surgeries requires far more than a performant model. Statistical model performance metrics like Mean Absolute Error obscure the spread and direction of scheduling error, masking the different operational consequences of over- and under-scheduling procedures. Using simulation frameworks, we reframed our results around operational opportunity and expected benefit, such as additional schedulable cases and reduced patient wait times, which proved far more powerful in driving adoption. Drawing from our implementation of a real-time Machine Learning prediction tool for surgical schedulers at Memorial Sloan Kettering Cancer Center, we share practical lessons on earning stakeholder buy-in, iterating and scaling the live application, and assessing post-deployment impact.

WA10

306A

AI Use in Medical Diagnostics

Contributed Session

AI

Chair: Lina Zhou, UNC Charlotte

1 - Temporal Dynamics of Prognostic Factors in Breast Cancer Survival: a Human-Centered AI Approach

Nuray Uzunkaya, Ayranci Anatolian High School, Ankara, Turkey, Serhat Simsek

Accurate prediction of cancer survival is essential for supporting clinical decision-making, yet traditional approaches often overlook how the importance of prognostic factors evolves over time. Understanding these temporal dynamics is critical for developing interpretable and human-centered artificial intelligence systems in healthcare.

This study proposes a data-driven framework to analyze how the impact of clinical variables on breast cancer survival changes across clinically meaningful time horizons. Using the SEER dataset, survival models are constructed for 6-month, 1-year, 2-year, and 3-year periods, enabling a fine-grained examination of short- and mid-term outcomes. Predictive models are combined with sensitivity analysis and information fusion techniques to identify and compare the relative importance of variables at each time interval.

The results demonstrate that prognostic factors exhibit distinct temporal patterns. Certain variables, such as tumor characteristics and lymph node involvement, show increasing importance over time, while others are more influential in early-stage survival. These shifting patterns highlight the need for time-aware decision support systems that adapt to different stages of patient care.

This work contributes to human-centered and explainable AI by moving beyond static prediction models and introducing a temporal perspective on variable importance. The proposed approach enables clinicians to better understand which factors matter most at different stages, supporting more informed, transparent, and adaptive treatment decisions.

2 - An Interpretable ML Method for Identifying Ultra-High-Cost Patients Across Medical Conditions

Catherine Ning, MIT, Cambridge, MA, United States, Dimitris Bertsimas, Omid Nohadani

Rare-event detection in high-stakes domains such as identifying patients at risk of exceptionally high future medical spending demands interpretable models, making decision trees a natural choice. However, rare events imply extreme class imbalance, and decision trees are known to suffer under such conditions: the learned partitions collapse into shallow, majority-dominated structures that fail to capture minority-class signal, degrading both predictive performance and the model's value as an auditable decision tool.

We address this problem by formulating training data curation as a principled optimization problem that explicitly balances two objectives: retaining majority samples near the decision boundary (boundary support) and maintaining geometric coverage across the majority distribution. Unlike existing sampling heuristics, this formulation provides a clear optimization target. To scale to large datasets, we derive a two-stage approximation: a farthest-first k-center construction that inherits classical approximation guarantees, followed by exact bipartite matching on the resulting candidate pool. Furthermore, we introduce a leaf-level evaluation metric that quantifies the residual discriminative power of one model within the subgroups induced by another, enabling direct comparison of tree partition quality beyond global accuracy.

We evaluate the method on a large healthcare claims problem spanning diverse high-cost clinical cohorts, including chronic kidney disease, ischemic heart disease, and musculoskeletal conditions. Across cohorts, decision trees trained on curated subsets not only improve held-out minority-class detection but also yield more informative partitions under our leaf-level evaluation. While motivated by healthcare applications, the underlying framework is applicable to other rare-event decision problems.

3 - Combining Graph-Theoretic Eeg Features and Machine Learning for Robust Stress Detection

Shotabdi Roy, Oklahoma State University, Stillwater, OK, United States, Joseph Nuamah

Understanding how brain networks dynamically reorganize under stress is critical for improving human performance and safety in high-stakes healthcare environments. This study integrates graph-theoretic electroencephalography (EEG) network analysis with machine learning to classify task-induced stress during human–system interaction using a publicly available PhysioNet dataset. EEG data from 35 participants were analyzed across theta (5–7 Hz), alpha (8–12 Hz), and beta (15–29 Hz) frequency bands, where functional connectivity was estimated using the weighted phase lag index (wPLI) and transformed into minimum spanning tree (MST) representations to extract topology-based features. A subject-independent classification framework using Leave-One-Subject-Out (LOSO) cross-validation was implemented to distinguish baseline and acute stress conditions. Logistic Regression, Support Vector Machine, and Random Forest models were evaluated using accuracy, F1-score, precision, and the Area Under the Receiver Operating Characteristic Curve (ROC-AUC) with aggregated predictions across folds. The best performance was obtained with a Radial Basis Function (RBF)-kernel Support Vector Machine in the theta band, achieving 71% accuracy, 71% precision, 71% recall, 71% F1-score, and an aggregated ROC-AUC of 0.789. Logistic Regression in the theta band showed comparable performance (70% accuracy, F1-score 70%, ROC-AUC 0.789), while alpha-band Logistic Regression yielded the strongest alpha-band result (64% accuracy, F1-score 64%, ROC-AUC 0.701). Overall, theta-band models outperformed alpha- and beta-band models, indicating that stress-related discriminative information is most strongly captured in theta-band network topology. These findings demonstrate the potential of interpretable EEG network biomarkers combined with subject-independent machine learning for stress monitoring in healthcare settings.

4 - Transference in Human-Chatbot Relationship

Swapn Shah, The Univeristy of North Carolina at Charlotte, Charlotte, NC, United States, Lina Zhou, Shuai Shao, Brendan Zhou, Dongsong Zhang

Mental health crises are increasingly prevalent worldwide, yet treatment gaps remain due to the shortage of mental health professionals. As AI-assisted chatbots increasingly serve as scalable complements, understanding the psychological dynamics users develop with them becomes critical. Transference, the redirection of emotional and relational patterns onto a new figure, has been documented extensively in human therapy. Unmanaged transference can reinforce emotional dependency, distort clinical judgment, and undermine treatment outcomes. We investigate whether analogous patterns emerge in human-chatbot interactions by analyzing observable indicators such as dependency, romantic attachment, and idealization in AI companion communities. We collect social media data from five subreddit communities spanning therapeutic, romantic, and general-purpose AI companion platforms. We first use LLM-assisted structured coding to classify posts against a nine-form transference taxonomy grounded in clinical literature, and validate these classifications using two complementary unsupervised methods: LDA (a topic modeling approach) and HDBSCAN (a clustering approach). Based on LDA results, we identified five topics: romance and companionship, AI-as-therapist,

attachment, role-play, and personal connection. The HDBSCAN results produced 12 topics. Some are directly aligned with the LDA topics or capture them at a finer level of granularity. Other topics are about AI companion building mechanics and some negative effects of transference such as dependency and separation anxiety. We further analyzed engagement patterns across topics. The results show that AI-as-a-therapist generated the highest user engagement. It appears that platform culture shapes how transference is expressed. These findings carry implications for AI platform safety design and mental health policy.

5 - When Should AI Step Back? An Uncertainty-Aware Escalation Framework for Improving Safety of Mental Health Chatbots

Lina Zhou, UNC Charlotte, Charlotte, NC, United States, Ethan Zhang, Brendan Zhou

Large language model (LLM)–based mental health chatbots are increasingly used to provide scalable, on-demand mental and emotional support. However, their deployment raises serious safety and ethical concerns, especially when a chatbot continues responding to users in risky situations that require human intervention (e.g., suicidal ideation). Most existing approaches rely on single-turn risk classification, which often fails to capture changes in distress over time, differences in risk type, and uncertainty in model predictions.

This project proposes an uncertainty-aware, context-sensitive escalation framework that determines when a mental health chatbot should continue conversation with users, offer supportive resources, or defer to human support. The framework integrates three components: temporal risk modeling to track how distress evolves across a conversation, uncertainty estimation to measure confidence in model predictions, and multi-dimensional risk assessment that considers severity, category, and progression of distress. These signals are combined in a policy-based decision module that supports graded escalation rather than a simple binary handoff. The system also generates brief interpretable rationales for escalation decisions to improve transparency and accountability.

The framework will be evaluated using both predictive and safety-centered metrics, including macro-F1, precision, and recall for high-risk cases, unsafe continuation rate, over-escalation burden, and time-to-escalation. By shifting the focus from static risk classification to selective and responsible autonomy, this work can support the development of safer, more reliable, and more human-centered AI systems for mental health support.

WA11

306B

Treatment Planning Using AI and Machine Learning

Invited Session

Healthcare Technology and Bioinformatics

Chair: Shan Liu, University of Washington, Seattle, WA, United States

Co-Chair: Jeremy Watts, I-O Urology, Knoxville, TN, United States

1 - So Many Tests, So Little TIME: Operationally Constrained Machine Learning for Diagnostic Assessment Under Time Constraints

Himadri Pandey, Georgia Institute of Technology, Atlanta, GA, United States, Gian Garcia, Landon Lempke

In many application areas, practitioners must make diagnostic decisions under operational constraints. For example, when diagnosing concussion—one of the most common types of traumatic brain injuries—clinicians may assess patients under time constraints imposed by athletic, emergency department, or military settings. Yet, many existing machine learning (ML) techniques fail to account for such operational constraints in the construction of a diagnostic battery. Accordingly, this research proposes a novel approach to generating ML models that enforce operational constraints such as time and interpretability. Motivated by the need to incorporate stochasticity in the time associated with each concussion subtest (i.e., feature), we model this prediction problem as a chance-constrained mixed integer program and propose several reformulations to enhance tractability. We then apply our models to a large, multi-site dataset on sports-related concussion among collegiate athletes and military cadets (n=47,860). Compared to diagnostic batteries commonly used in current

practice, the concussion diagnosis batteries designed by our methods are both time-efficient (achieving >97% reduction in time needed to administer the battery) and highly accurate (maintaining area under the curve > 0.93). As such, our method can design practically relevant ML models that can be applied broadly across many industries and application areas.

2 - Sequential Diagnostic Decision-Making in Chronic Bph: a Pomdp Comparison of in-Office and at-Home Testing Strategies

Jeremy Watts, IO Urology, Knoxville, TN, United States, Ashley Sturdy, John Knoedler

Benign prostatic hyperplasia (BPH) is a prevalent chronic condition characterized by heterogeneous lower urinary tract symptoms and substantial variation in diagnostic evaluation and treatment pathways. Clinical decision-making is complicated by overlapping phenotypes, imperfect diagnostic tests, and treatment responses that depend on underlying pathophysiology. We develop a finite-horizon partially observable Markov decision process (POMDP) to evaluate diagnostic and treatment strategies in chronic BPH management. Patients are modeled as occupying latent clinical states representing distinct symptom and pathophysiologic profiles. In-office diagnostic tests and multi-day at-home testing generate noisy observations of these latent states, while pharmacologic and procedural interventions induce state transitions reflecting heterogeneous treatment effects and potential failure.

The framework captures sequential decision-making over discrete clinical encounters, with belief states updated via Bayesian inference and actions selected to optimize a weighted objective combining quality of life and cost. We compare conventional clinic-based diagnostic workflows with an alternative strategy incorporating multi-day at-home testing that provides richer or more frequent information. Differences in outcomes emerge through changes in information quality, treatment targeting, and rates of inappropriate or ineffective intervention.

The model supports evaluation of expected costs, quality-adjusted outcomes, utilization patterns, and provides a general decision-analytic framework for assessing the operational and economic value of distributed diagnostic technologies in chronic disease management.

3 - Using Multi-Void Home Uroflowmetry Data and Machine Learning to Support Treatment Planning in Patients with Lower Urinary Tract Symptoms

Ashley Sturdy, iO Urology, Knoxville, TN, United States, Jeremy Watts

Evaluation of lower urinary tract symptoms (LUTS) often relies on single point-of-care diagnostic testing, providing only a narrow snapshot of a patient's voiding physiology. In-office uroflowmetry may not reflect typical urinary flow patterns because it is sensitive to environmental factors, voided volume, and normal variability. As a result, treatment planning and decisions about further testing may be based on limited and potentially unrepresentative data.

CarePath is a handheld home uroflowmetry device that collects multiple voids over several days in a patient's natural environment. By capturing repeated measures of peak flow rate, average flow rate, voided volume, and time-to-peak flow, the device provides a more complete characterization of voiding behavior. This multi-void dataset enables assessment of how consistent a patient's flow measurements are across time and varying voided volumes.

We explore how machine learning methods can use repeated multi-void measurements to support treatment planning. By comparing models built from a single void with those incorporating multiple voids, we assess whether multi-void data improves confidence in interpretation and informs decisions about additional testing, therapy selection, and monitoring. Rather than focusing on a specific diagnosis, this approach centers on improving clinical decision-making across evaluation and management pathways.

Using repeated measurements instead of isolated observations may reduce uncertainty in care decisions. Integrating multi-void data into AI-enabled tools could support more individualized, data-informed management of patients with LUTS.

4 - Interpretable Machine Learning for Trustworthy Clinical Decision Support: Lessons from Nephrocompass

Deepti Bahel, VigourSoft Global Solutions, San Jose, CA, United States

Machine learning models in healthcare often achieve strong predictive performance, yet struggle to earn clinical trust due to limited transparency and actionability. Inspired by principles from **Interpretable Machine Learning**, this talk presents how interpretability can be designed as a *core system capability* rather than a post-hoc explanation layer, using **NephroCompass**, an AI-powered kidney health decision support system, as a case study.

NephroCompass applies interpretable machine learning techniques to support early risk assessment and care navigation for chronic kidney disease. Global explanations are used to understand population-level drivers such as lab trends, comorbidities, and progression patterns, while local explanations and counterfactual reasoning help clinicians and patients explore individualized risk scenarios and potential intervention paths. Instead of producing opaque risk scores, the system surfaces *why* a recommendation is made, *what factors matter most*, and *how outcomes might change* under different conditions.

The talk also highlights limitations of purely associative models in clinical settings, including sensitivity to distribution shifts and overconfidence under uncertainty. By embedding interpretability, uncertainty awareness, and human-in-the-loop feedback into the workflow, NephroCompass demonstrates how explainable AI can bridge the gap between prediction and decision-making. This session offers practical insights for building trustworthy, deployable clinical decision support systems that align technical rigor with real-world healthcare needs.

WA12

306C

Innovative Applications in Healthcare

Invited Session

MSOM: Healthcare

Chair: Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States

1 - Near-Optimal Monitoring in Delegated Clinical Trials

Jinze Cui, Duke University, Durham, NC, United States, Peng Sun

Delegating clinical trials to contract research organizations (CROs) improves efficiency but may also induce shirking. Sponsors can deter shirking through monitoring, but investigations are costly. We study how a sponsor can induce full CRO effort across all participants at minimal expected cost. Using a mechanism design framework with costly state verification, we propose a simple yet effective contract, the **uniform-pass contract**: the sponsor investigates each participant with an equal probability and pays the CRO only if no shirking is detected. We prove that this contract is near-optimal. As trial size n increases, total cost approaches the theoretical minimum at rate $O(1/\sqrt{n})$, while the expected number of investigations grows only $O(\log n)$. Calibrated to Phase 3 clinical trials with 500 to 3,000 participants,the uniform-pass contract's total cost is only 1.5% to 8% above the theoretical minimum, and it requires investigating only 28 to 74 participants in expectation. These findings provide theoretical support for a practically scalable approach to monitoring in delegated clinical trials.

2 - Unpacking Transparency: Information, Bargaining, and Search Costs in Healthcare Markets

Hansraj Satyam Verma, Carnegie Mellon University, Pittsburgh, PA, United States, Deepak Agrawal, Sridhar Tayur, Amin Hosseininasab

This paper develops a game-theoretic model of insurer-hospital negotiations to study how the healthcare price disclosure mandate affects negotiated rates and market outcomes. We explore how such mandated transparency acts as a two-sided information shock that alters the bargaining dynamics between payers and hospitals. By exposing cross-payer and cross-hospital rate disparities, the mandate shifts the outside options and leverage

available to both sides of the negotiation. Automated data aggregation tools amplify these effects by lowering the search costs of processing published machine-readable data. We evaluate our model's predictions using a large cross-sectional dataset of parsed negotiated rates alongside data from the RAND National Hospital Price Transparency Study, and use empirical simulations to illustrate the resulting impacts on equilibrium prices, incentives, and welfare.

3 - Provider Payment Models for Transformative Technologies in Healthcare

Elodie Adida, University of California - Riverside, Riverside, CA, United States, Tinglong Dai

Transformative health technologies, such as robotic surgery and AI-enabled diagnostics, can improve outcomes but often launch without reimbursement that matches implementation cost. This creates a two-sided incentive problem: providers face high deployment costs, while developers face weak demand-side returns to quality investment. We study this problem using a game-theoretic model of a developer, a provider, and a payer. The developer chooses quality and per-use price; the provider chooses unobservable implementation effort and patient-level adoption; and the payer sets reimbursement. Under no reimbursement, providers restrict use to high-complexity cases, effort is below the social benchmark, and developers underinvest in quality. Fee-for-service expands diffusion but does not remove the distortion: in the interior-adoption regime, use remains below the social benchmark because the developer strategically adjusts both price and quality; at sufficiently high reimbursement, adoption becomes universal, creating overuse risk and lower quality. We then compare value-based designs. Benefit-based payment, tied to realized gains, skews use toward high-complexity patients. Quality-based payment, tied to verifiable upstream performance dimensions, better targets developer incentives but is limited when used alone. We therefore propose a hybrid design combining fee-for-service and quality-based incentives. The fee-for-service component drives diffusion, while the quality-based component sustains investment. This yields an ordering: a hybrid with quality-based incentives outperforms a hybrid with benefit-based incentives, clarifying trade-offs among access, quality, and long-run innovation. Our contribution is to model a triadic payer-provider-developer interaction in which provider-facing reimbursement reshapes upstream innovation incentives, shedding new light on payment policy for transformative technologies in healthcare operations settings.

WA13

307

Policy, Analytics, and AI in Modern Healthcare Delivery

Invited Session

eBusiness

Chair: Kanghyun Cho, Indiana University, Philadelphia, United States

1 - Evaluating Effect of Medicaid Managed Care on Adolescent Immunization in North Carolina

Sooyeol Park, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Justin Trogdon

Adolescent immunization rates remain below public health goals, especially for human papillomavirus (HPV). Evidence on whether Medicaid transitions to capitated managed care improve vaccine uptake is mixed. North Carolina launched its first Medicaid managed care plan on July 1, 2021 and incorporated HEDIS Immunizations for Adolescents Combination 2 (Combo2) into managed care accountability. We assessed whether managed care adoption changed uptake of each component vaccine.

Using North Carolina immunization registry records linked to Medicaid claims, we measured monthly doses per eligible adolescent for meningococcal (MGC), tetanus-diphtheria-acellular-pertussis (Tdap), and HPV vaccines within HEDIS-defined age windows from November 2016 to December 2023. We estimated interrupted time series models using a linear probability model for Combo2 and Poisson pseudo-maximum likelihood for monthly doses. We also conducted difference-in-differences analyses using North Carolina annualized outcomes and National Immunization Survey-Teen data on Medicaid-insured adolescents from states with stable managed care and Affordable Care Act expansion status.

Interrupted time series results showed that by December 2023, the fitted post-implementation level remained close to the counterfactual. For component vaccine delivery, only HPV showed a significant immediate decline that attenuated over follow-up. In difference-in-differences analyses, HPV dose delivery increased relative to comparison states. Results were similar in subgroup and sensitivity analyses.

North Carolina’s transition to managed care was not associated with measurable changes in Combo2 completion in within-state models, but comparative analyses suggested improved HPV dose delivery relative to similar states.

2 - The Impact of Algorithm-Based Prior Authorization on Public Healthcare Delivery

Hyeongseo Park, University of Minnesota, Minneapolis, MN, United States, Kartik Ganju, Gautam Ray

Prior authorization requires providers to obtain insurer approval before delivering certain services. Increasingly, insurers use artificial intelligence to make these decisions, aiming to reduce unnecessary care and help patients anticipate costs. However, evidence suggests it may delay or restrict access: a 2024 American Medical Association survey found that 93% of physicians reported care delays and 82% said patients sometimes abandon treatment due to prior authorization.

This study examines the impact of AI-driven prior authorization by focusing on EviCore, a third party AI developer company and major contractor used by insurers. Using Vermont Medicaid data from 2016-2019, the analysis evaluates how delegating prior authorization to EviCore affects medical procedure volume. A difference-in-differences approach compares specialties subject to EviCore (cardiology and radiology) between hospitals participating in the state's ACO program and those not participating.

Results show that AI-driven prior authorization is associated with a substantial reduction in care: cardiology and radiology procedures delivered at non-ACO-affiliated hospitals declined by about 34% relative to affiliated facilities after EviCore's intervention. These findings remain robust when procedures are matched based on patient characteristics. However, no significant changes are observed in the number of hospital admissions involving cardiac ambulatory care sensitive conditions as well as the mean length of stay per admission.

Overall, the findings suggest that AI-based prior authorization reduces the absolute volume of care with no obvious signs of causing under-provision. The results highlight AI's potential to improve healthcare delivery by identifying and minimizing unnecessary costs through prior authorization.

3 - Optimizing Antimicrobial Resistance Control Under Dynamic Uncertainty: a Reinforcement Learning&Ndash;Agent-Based Modeling Framework

Shuyi Song, UNC Chapel Hill, Chapel Hill, NC, United States, Sean Sylvia, Tara Templin

Field randomized controlled trials provide credible evidence for evaluating healthcare interventions. However, estimates are often obtained in specific contexts and over limited time period, making it difficult to assess how policies will perform when implemented at scale in complex healthcare systems characterized by heterogeneous agents, strategic interactions, and dynamic feedback. This paper develops a reinforcement learning agent-based modeling (RL-ABM) framework that integrates experimental evidence with dynamic policy simulation. Using field data from rural China, we calibrate patient and provider behaviors related to antibiotic use and healthcare seeking among children under five. The calibrated agent-based model captures key mechanisms underlying antimicrobial resistance (AMR), including provider prescribing incentives, patient care-seeking decisions, access to antibiotics, and the accumulation of antibiotic pressure that shapes resistance dynamics. We then embed reinforcement learning algorithms to search the policy space and evaluate and improve intervention strategies over time. The framework enables policymakers to examine combinations of regulatory and behavioral interventions, such as provider training, diagnostic testing policies, pharmacy regulation, and public health education, under realistic behavioral and epidemiological dynamics. Simulation results show that policy performance depends on feedback between prescribing behavior and resistance evolution, as well as delayed system responses that can produce cyclical or unstable patterns under static policy rules. By combining experimental calibration with dynamic policy optimization, this study offers a computational approach to designing adaptive health policies in settings where traditional experimental or econometric methods alone are insufficient.

4 - The Devil's Advocate: when AI Challenges Human Expertise

Kanghyun Cho, Indiana University, Bloomington, IN, United States, Kihwan Nam, Jin-Woo Park

Integrating artificial intelligence (AI) into high-stakes clinical diagnostics while preserving professional authority poses a significant challenge to the delivery of safe and effective healthcare. Toward this end, we develop the identity-contingent reflection framework, which integrates reflective practice theory and professional identity theory to explain how the timing of AI recommendations shapes expert performance. This framework enables us to answer three key questions: (a) Does workflow timing impact diagnostic classification correctness? (b) Does task difficulty moderate the effectiveness of these different workflow timings? (c) Do identity-driven concerns explain the performance variations observed in high-stakes environments? We test our framework using a multi-method design, including an exploratory physician study and a unique experimental dataset tracking 1,350 diagnostic trials performed by 30 board-certified physiatrists. We demonstrate that the efficacy of AI-assisted workflows is not uniform but rather is significantly moderated by task difficulty, challenging the assumption that retrospective evaluation is universally beneficial. We find that while a two-step workflow facilitates productive reflection and accuracy gains in medium-difficulty tasks, it triggers an "engagement trap" in high-difficulty scenarios where diagnostic signals are ambiguous. Furthermore, our model identifies "expert fragility" among senior physicians, where salient disagreement from AI leads to a 3.51-fold increase in high-magnitude critical mistakes. Our results indicate that task-adaptive architectures and collaborative evidence visualization are essential building blocks for optimizing human–AI collaboration in high-stakes clinical environments.

Wednesday, July 29, 2026, 9:45 AM - 10:45 AM

Wednesday Plenary

Ballroom B

Diagnosing Complex Problems in Pediatric Emergency Medicine and Developing Comprehensive Treatment Plans

Plenary Session

Plenary

1 - Diagnosing Complex Problems in Pediatric Emergency Medicine and Developing Comprehensive Treatment Plans

Amy Cohn, Michigan Medicine, Ann Arbor, MI, United States

Just as the human body is a complex system of interconnecting and interdependent components, so is the healthcare system. Given such complexity, it can be easy to see the symptoms of a problem but difficult to diagnose the underlying cause (or causes). Likewise, it can be challenging to develop and coordinate a treatment plan that addresses the root problem without causing unnecessary or unacceptable harm to other components. Physicians are trained to solve these complex diagnostic and treatment planning problems for the human body. In this talk, I will discuss how I approach the challenges of "diagnosis' and "treatment planning" for addressing challenges in healthcare operations.

Wednesday, July 29, 2026, 11:00 AM - 12:15 PM

WB01

301A

AI Coworkers in Healthcare: Real-World Applications, Human-AI Collaboration, and Emerging Talent

Contributed Session

AI

Chair: Beju Rao, Amruta, Inc, Fredericksburg, VA, United States

1 - Hospital Discharge Planning & Throughput Optimization

Anay Misra, Amruta, Inc, McLean, VA, United States, Beju Rao, Rushil Malhotra

Hospitals continue to face persistent discharge delays driven by fragmented communication, incomplete readiness assessments, and high administrative burden on clinical teams. We present a governed, explainable multi-agent AI coworker that integrates statistics, operations research, machine learning, and large language models into a unified, observable workflow for discharge planning. Built on an OpenClaw-based agentic framework, the system orchestrates multiple specialized agents—including readiness assessment, barrier identification, care coordination, and documentation support—each equipped with causal reasoning, transparent explanations, and policy-aligned guardrails.

The AI coworker operates within clinical and administrative workflows, synthesizing EHR data, operational constraints, and contextual patient factors to recommend next-best actions, surface bottlenecks, and automate routine tasks. In a recent deployment, the platform reduced Emergency Department congestion by 14% by accelerating safe inpatient throughput and improving discharge predictability. Additional deployments demonstrate reductions in avoidable readmissions and improvements in care quality, safety, and patient satisfaction.

This presentation will demonstrate the multi-agent architecture, governance and validation framework, explainability methods, and real-world performance results. We will also discuss refactoring the technology for scalable commercial deployment across diverse hospital environments.

2 - AI Ready Obligation Intelligence for Healthcare Supply Chains

Beju Rao, Amruta, Inc, Fredericksburg, VA, United States, Naresh Reddy

Healthcare supply chains operate under increasing regulatory and contractual complexity, where obligations related to service execution, compliance, and product traceability must be managed reliably and at scale. As organizations accelerate AI adoption across operations, fewer than 30% of pilots scale enterprise-wide to deliver expected value due to the lack of AI-ready foundational gaps in data architecture, integration, governance, and process redesign.

This session introduces a practical framework for Obligation Intelligence enabled by a Governed Contract Data Layer integrated with enterprise master data and extended through an Agentic AI execution layer. The architecture operationalizes six sequential layers contract ingestion, obligation extraction, data standardization, governance control, decision/agentic execution, and monitoring & audit so automation runs on governed, policy-compliant data (with definitions, lineage, validation/quality controls, and access/security) rather than ungoverned raw inputs

The approach is validated across enterprise use cases including CLM, master data platforms, and regulated supply chain systems, demonstrating how governed obligation models improve traceability and auditability while reducing manual effort, exception handling, and compliance delays. Attendees will leave with actionable implementation patterns to design governed data foundations that enable trustworthy AI, scalable compliance execution, and audit-ready reporting in healthcare operations

3 - Readmissions Modeling, Golden-Case Knowledge Engineering & Human-AI Collaboration

Beju Rao, Amruta, Inc, Fredericksburg, VA, United States, Ishaan Kommula, Raj Maheshwarapu

Avoidable hospital readmissions cost U.S. health systems an estimated **$25 billion annually**, yet most predictive models fail to translate risk scores into actionable interventions. This presentation introduces a governed, explainable machine-learning framework that bridges prediction with prevention by mapping each predicted readmission to an interpretable, clinically meaningful “golden case.” Golden cases encode expert knowledge about preventable readmission pathways—such as medication confusion, follow-up failure, social-determinant

barriers, or discharge-instruction gaps—and serve as the operational link between model explanations and targeted interventions.

Using SHAP-based interpretability, causal features, and contextual patient factors, each high-risk prediction is matched to the most relevant golden case, enabling clinicians and operational teams to take precise, evidence-aligned actions. A human-in-the-loop feedback mechanism allows new golden cases to emerge over time, capturing patterns not previously recognized and continuously improving the system's actionability. Not all readmissions are preventable; the framework explicitly distinguishes preventable vs. non-preventable cases to focus resources where they matter most.

Finally, we demonstrate how these actionable insights integrate directly into **AI coworkers**—including discharge-planning agents—to operationalize interventions, reduce cognitive burden, and improve patient outcomes at scale.

4 - Appointment Readiness & Clinician Utilization Efficiency

Ahana Narayanasetty, Amruta Inc, Aldie, VA, United States, Beju Rao

Missed, incomplete, or unprepared appointments remain a major source of wasted clinician time, reduced access, and avoidable operational cost. We introduce a contextual, explainable patient-engagement AI agent designed to ensure appointment readiness, reduce no-shows, and maximize clinician utilization efficiency. The agent integrates causal inference, machine learning, operations research, and large language models into a governed, observable workflow that interacts with patients through multimodal channels while maintaining transparency, safety, and policy compliance.

Using an OpenClaw-based agentic framework, the system dynamically assesses readiness risks, identifies missing prerequisites (labs, imaging, paperwork, referrals), and generates tailored outreach with intelligible explanations. It also coordinates with scheduling, care teams, and operational systems to optimize slot utilization and reduce idle time. Early deployments show improvements in patient preparedness, reductions in avoidable appointment churn, and measurable gains in clinician productivity and revenue capture.

This presentation will demonstrate the agent's architecture, explainability and governance mechanisms, and integration into clinical operations. We will also discuss pathways for scaling the technology, including modularization, commercial deployment models, and alignment with health-system digital transformation strategies.

WB02

301B

Stochastic Models for Healthcare Operations

Invited Session

Applied Probability

Chair: Yue Hu, Stanford University, Stanford, CA, United States

1 - Online Learning for Dynamic Service Mode Control

Wenqian Xing, Stanford University, Stanford, CA, United States, Yue Hu, Anand Kalvit, Vahid Sarhangian

Modern service systems are increasingly adopting innovative modalities enabled by emerging technologies, such as AI-assisted services, to achieve a better balance between quality and efficiency. Motivated by these advancements, we study the optimal dynamic control of service modes in systems with unknown parameters, aiming to balance reward accumulation and system congestion. We study a single-server queue with two switchable service modes, each with a distinct service rate and an unknown reward distribution upon completion. The objective is to maximize the long-run average of cumulative rewards minus holding costs. To address this problem, we first characterize the optimal state-dependent policy under full knowledge of model parameters. For unknown reward distributions, we propose a Upper Confidence Bound (UCB) based online algorithm that

adaptively learns the optimal policy, achieving near-optimal regret on the order of square root T over a time horizon T, with strong empirical performance. Beyond optimism, we show that the core statistical challenge is determining whether state dependence is necessary. Once it is known that the optimal policy is state-dependent, the queueing dynamics generate sufficient information, and learning becomes free. Exploiting this insight, we design an explore-then-commit policy that matches the near-optimal regret of UCB and achieves constant regret once this bottleneck is resolved. A key methodological contribution is a novel regret decomposition and a regenerative cycle-based framework, offering insights into learning-based optimal control in queueing systems. Finally, we demonstrate the practical relevance of our approach through a case study on optimizing AI-assisted patient message replies in a healthcare setting.

2 - Prioritizing Recurrent Services

Lin (Franklin) Feng, Stanford University, Stanford, CA, United States, Yue Hu, Kuang Xu

We study optimal scheduling in queueing systems with reentrance, where jobs may return for additional service after completion. Such reentrance creates feedback loops that fundamentally alter congestion dynamics and challenge classical scheduling insights based solely on current workload. We model two key dimensions of reentrance behavior: the probability and speed of return, and show that their product, the *effective return rate*, is the central statistic governing optimal priorities.

Our main result establishes a sharp dichotomy. When the smaller job class (i.e., the class with lower expected total workload) has a lower effective return rate, a fixed priority policy is optimal. When it has a higher effective return rate, fixed priority rules become suboptimal and the optimal policy must be state dependent. Despite this, the optimal policy admits a simple structure with at most one priority switch: prioritizing smaller jobs under heavy congestion and larger jobs under light congestion.

Guided by these insights, we propose the Unresponsive-First-Offpeak (UFO) policy, a practical scheduling rule that captures this switching logic in stochastic environments. We evaluate UFO through simulation and an empirical study using hospital messaging data, where patient-provider interactions generate reentrant workload. Our results show that UFO can reduce total holding costs by up to 35% relative to current practice and achieves gains from 5% to 20% over standard benchmarks in settings with heterogeneous return behavior. These findings demonstrate that incorporating reentrance-aware scheduling can substantially improve operational performance in real-world service systems.

3 - Capacity Sizing with Strategic Labor Supply

Saniya Vaidya, Stanford University, Stanford, CA, United States, Zhuoyang Liu, Yue Hu, Dan Iancu

Many service systems manage demand uncertainty by combining a lower-cost base capacity, committed to in advance, and a flexible but expensive surge capacity deployed once demand information becomes available. In many settings, such as healthcare staffing, gig platforms, and inventory management, workers respond strategically to compensation and workload, shaping the available labor pool.

We study capacity planning with strategic labor supply, where workers endogenously choose between base employment, surge employment, and outside opportunities. Anticipating these responses, the system manager determines base and surge capacity to minimize total expected costs. We show that labor market responses fundamentally reshape optimal staffing policies.

Our analysis yields three main insights. First, we show that there exists an optimal policy with a two-threshold structure: surge staffing fills the gap between realized demand and base capacity, but is truncated between a minimum and a maximum level. This truncation allows the system to distort surge usage in order to influence workers' participation decisions. Second, in large markets, if labor is abundant, the staffing problem effectively reduces to the classical base-surge capacity planning problem without strategic labor supply. When labor becomes scarce, the system may either actively utilize surge employment or avoid it altogether to maintain a strong base pool. Third, we compare our policy with an adaptive but myopic benchmark and show that the proposed policy asymptotically dominates this benchmark in the worst case and can yield cost savings that scale linearly with demand.

WB03

302A

AI-Enabled Healthcare

Invited Session

AI in Healthcare

Chair: Yu Ma, University of Wisconsin Madison, Madison, WI, United States

1 - Patient Centric Surgery Scheduling for Pediatrics via Optimal Fasting

Panus Sawetpiyakul, University of Washington, Seattle, WA, United States, Katherine Adams, Gian-Gabriel Garcia, Arka Roy, Vishal Ahuja, Dai Chung

Operating rooms (ORs) are a major cost center in most hospitals. As such, many investigators have proposed both quantitative and qualitative strategies to optimize OR utilization, reduce patients' waiting time, and improve patient satisfaction. However, to our knowledge, there is limited quantitative research that explicitly considers how patients' excess fasting duration interacts with traditional OR scheduling metrics. In pediatric hospitals, there is a significant need to minimize excessive fasting, as pediatric patients are less tolerant of prolonged fasting. In this work, we construct an optimal OR schedule that maximizes expected utilization while minimizing expected excess fasting duration for pediatric patients. Since optimization-based schedules can be difficult to implement in practice, we additionally propose two heuristic scheduling strategies for practical use. We then compare excess fasting durations and expected utilization between the optimal schedule, the two heuristic strategies, and the current scheduling policy used in our partner pediatric hospital. Overall, we find that both the optimal and heuristic OR scheduling strategies reduce excess fasting time relative to the current scheduling policy while achieving similar or higher utilization. This result highlights a "win-win" scenario in which operational targets and patient satisfaction both improve relative to the current standard.

2 - Algorithmic Replenishment for Healthcare Equity: A Field Experiment in Remote Pharmaceutical Supply Chains

Xianjia Gao, University of Illinois Urbana-Champaign, Champaign, IL, United States, Pengyi Shi

This study investigates the operational challenges of pharmaceutical inventory management in geographically constrained island communities with significant aging populations. Through a field experiment with a regional retail pharmacy chain, we implemented a data-driven algorithmic inventory management system to optimize traditional manual replenishment decisions. A primary methodological challenge in these remote settings is the severe data sparsity and intermittent demand at the individual item level. To address this, our algorithm leverages demand pooling by aggregating substitutable medications at the category level, significantly enhancing forecasting accuracy. In formulating inventory decisions, rather than relying solely on traditional profit-maximizing models, our approach explicitly incorporates the demographic realities of the remote islands. Recognizing the high healthcare demand of the vulnerable elderly population, the algorithm imposes differentiated service-level constraints to prioritize and guarantee the availability of vital medications. Results from the field experiment demonstrate that this algorithmic approach successfully mitigated inventory backlogs, decreasing capital occupation and accelerating turnover while achieving a substantial reduction in overall stock-out rates. Crucially, by addressing data sparsity and enforcing strategic service-level targets, the system systematically covered low-frequency long-tail items and dynamically guaranteed the supply of critical chronic disease medications. Ultimately, this research provides empirical evidence that targeted algorithmic interventions can advance healthcare equity and improve socially responsible operations in remote, resource-constrained supply chains.

3 - Telehealth and the Reorganization of in-Person Provider Workload: Evidence from a Centralized Virtual Care Model

Kyuyeon Rhee, Arizona State University, Tempe, AZ, United States, Tan (Suparerk) Lekwijit, Xiao Liu, Pei-yu Chen

Telehealth is often introduced with the expectation that it will alleviate pressure on in-person care. In practice, however, it may reorganize how clinical work is generated and distributed across providers. Prior research suggests that video visits increase downstream utilization, yet their implications for provider workload remain poorly understood, particularly in settings where telehealth is delivered by a separate care team. We examine how telehealth affects in-person provider workload along three dimensions: volume, variability, and complexity. We analyze data from a large health system that introduced a centralized telehealth program staffed by a dedicated virtual care team, in which in-person providers are indirectly exposed through changes in their patient panels' care-seeking behavior. Guided by a queuing-based framework, we conceptualize telehealth as a front-end entry point that reshapes arrivals to the in-person care queue and estimate causal effects using a two-stage least squares approach exploiting variation in telehealth eligibility. We find that telehealth reorganizes in-person provider workload. First, telehealth exposure increases in-person visit volume. Second, it alters the temporal structure of demand by increasing the likelihood of capacity exceedance and concentrating demand in high-load periods. Third, telehealth shifts the composition of in-person care toward higher-comorbidity patients without increasing visit-level service intensity. These findings show that telehealth does not simply substitute for in-person care or uniformly increase workload. Instead, it reconfigures patient flows across channels, increasing volume, concentrating demand in peak periods, and shifting case mix toward more complex patients, with implications for capacity planning and workforce management in hybrid care delivery systems.

WB04

302B

Data-Driven Insights for Healthcare IT and Operational Performance

Invited Session

Healthcare Operations Management

Chair: Seokjun Youn, University of Arizona, Tucson, AZ, United States

1 - A Multi-Level Perspective of AI and Digital Coordination Capabilities for Operational Efficiency in Emergency Departments

Nirup Menon, George Mason University, Fairfax, VA, United States, Deepika Gopukumar, Sidhartha Das, Amitava Dutta

Hospitals are increasingly adopting artificial intelligence (AI) to improve operational decision-making, particularly through demand forecasting and resource planning. At the same time, unit-level digital coordination platforms enhance information flow and operational alignment within critical care units. This study examines how system-level AI and unit-level digital coordination capabilities jointly shape emergency department (ED) performance using a multi-hospital panel dataset. By distinguishing between technologies implemented at different organizational levels, this study offers a multi-level perspective on how digital transformation shapes patient flow and operational efficiency in complex service systems.

2 - Beyond Human Interaction: the Impact of Generative AI-Based Conversational Agents on Mental Health Counseling

Seokjun Youn, University of Arizona, Tucson, AZ, United States, Chenxi Guo, Karen Xie, Xunyu Chen, Wei Chen

The rapid integration of Generative AI (GenAI) conversational agents into digital mental health platforms raises critical questions about the future of professional care delivery. While GenAI agents promise scalable, always-available support, their impact on incumbent human counselors remains largely underexplored. This study examines how mental health counselors respond to the unexpected launch of a GenAI conversational agent on a leading online mental healthcare platform. Leveraging a quasi-experimental setting and a matched difference-in-differences design across two comparable platforms, we analyze behavioral and performance changes among 544 counselors over a two-year window. Contrary to displacement concerns, we find that counselors' paid appointments increase by 11.9% following the GenAI launch. This improvement is driven by heightened patient engagement. Counselors increase reply volume by 12.8%, receive 9.5% more patient "likes," and experience

11.2% more conversational exchanges. To uncover underlying mechanisms, we integrate social comparison theory with large language model (LLM)-powered content analysis of over 33,000 replies. We show that counselors systematically adjust their communication strategies in response to the GenAI agent's performance and content style. Specifically, counselors converge toward the GenAI agent's balance of informational and emotional support, selectively emulating high-performing elements while preserving economically strategic behaviors such as self-referral. Our findings advance understanding of human-AI competition in healthcare by demonstrating that GenAI can catalyze performance improvement rather than labor displacement. The results offer important implications for platform design, workforce adaptation, and responsible AI integration in digitally mediated healthcare systems.

3 - Interpretable Admission Prediction In the Emergency Department

Lewis Dubrowski, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Sukriye Nilay Argon, Serhan Ziya, Chudi Zhong, Thomas Smith

Previous research has explored the use of machine learning algorithms to predict an emergency department (ED) patient's likelihood of hospital admission. This project examines the relationship between model interpretability and performance in the admission prediction setting. We apply a risk score algorithm to patient health data from an ED in the Southeastern United States to build a simple admission prediction tool. By taking into account information gained during a patient's stay in the ED, our tool allows users to update a patient's predicted likelihood of admission as they learn more about the patient. We compare our model's performance with that of black-box and interpretable approaches used in previous studies. While we find that more complex methods demonstrate better performance, our model's interpretability may make it easier for practitioners to implement in some settings.

4 - Response Latency in Asynchronous Care Queues: AI-Augmented Measurement and Emergency Department Spillover

Mohamed Megahed, Fisher College of Business, The Ohio State University, Columbus, OH, United States, Nathan Craig, Aravind Chandrasekaran, Natalie Lee

Clinical message inboxes require rapid triage under congestion, yet urgency is embedded in unstructured text and patient history and must be inferred at scale. Leveraging natural language processing (NLP), supervised models, and clinician-calibrated severity scoring, we develop an AI-assisted triage framework that prioritizes messages for timely action. Using 3.2 million secure messages from a large U.S. academic health system, we assess how alternative workflow designs shapes urgency–response alignment, response delays, and escalation-related outcomes. Framing triage as an operational decision problem under capacity constraints, we assess performance beyond predictive accuracy by measuring system-level responsiveness and downstream risk. Results identify when AI augmentation improves triage effectiveness while preserving clinician judgment, and when workflow frictions and reliability limits attenuate gains.

WB05

302C

Analytics and Innovation in Health Systems

Invited Session

Public Health Policy

Chair: Xiaoquan Gao, Singapore Management University, Singapore, Singapore

Co-Chair: Buyun Li, Kelley School of Business, Indiana University, Bloomington, IN, United States

1 - A Bilevel Optimization Framework for State and County Budget Allocation to Combat the Opioid Epidemic

Jean-Philippe Richard, University of Minnesota, Minneapolis, MN, United States, Bouchra Er-Rabbany, Ankit Bansal, Samuel Workman, Henry Brownstein, Gordon Smith

This talk introduces a novel, data-driven methodology for the strategic allocation of state and county public budgets to mitigate the opioid epidemic across decentralized administrative units. We employ a bilevel

optimization framework that integrates strategic decision-making with predictive modeling. The upper level represents the state's allocations across counties, aiming to maximize overall public health outcomes. The lower level captures each county's autonomous budgeting decisions and local objectives. The framework incorporates Machine Learning (ML) models trained on empirical data to predict the epidemic's progression as a function of resource deployment. This formulation gives rise to a challenging bilevel integer programming problem. We present solution strategies tailored to this class of problems and demonstrate results that underscore the potential for data-driven improvements in the effectiveness of public spending.

2 - Overcoming the Underestimation in State-Level Prevalence of Opioid use Disorder from the National Surveys: a Bayesian Hierarchical Modeling Approach

Zixuan Feng, Penn State, State College, PA, United States, Qiushi Chen, Le Bao, Paul Griffin

The opioid epidemic remains a critical public health crisis in the US, with over 800,000 overdose deaths over the last two decades. Accurate estimation of opioid use disorder (OUD) prevalence is crucial for understanding treatment needs and allocating intervention resources. Currently, prevalence estimates primarily rely on self-reported surveys, such as the National Survey on Drug Use and Health (NSDUH), which have been known to systematically underestimate prevalence due to stigma and sampling bias. Conversely, health outcomes and healthcare utilization data, such as overdose mortality and Medications for OUD (MOUD) treatment data, offer more reliable indirect observations of the affected population, providing valuable information to triangulate and infer OUD prevalence. We present a Bayesian hierarchical model that synthesizes publicly available mortality and treatment data to address the underestimated prevalence from the national surveys. We infer OUD prevalence by jointly modeling the data generation processes of mortality and MOUD data. Applying our model to all 50 US states from 2021 to 2024, we found that our approach yields prevalence estimates that are substantially higher than the existing NSDUH estimates. In 2021, we estimate that over 9.2 million (3.6%) US adults (95% credible interval: [8.1 million (3.1%), 10.3 million (4.0%)]) had OUD, more than 50% higher than the NSDUH-reported estimate. We also estimate that NSDUH captures only 34.3%-83.7% of our OUD prevalence estimates across states. Our model provides policymakers with a data-driven method to address survey biases and better align treatment resources with actual public health needs.

3 - Digital Twins and the Reconfiguration of Expertise: Evidence from AI Avatars in Online Healthcare

HUA LU, University of International Business and Economics, Beijing, China, People's Republic of

The rapid integration of artificial intelligence (AI) into professional service sectors is fundamentally reshaping how expertise is delivered, raising urgent questions about whether and how AI empowers human experts. While prior studies have primarily examined AI in manufacturing and consumer recommendation contexts, limited attention has been paid to its supply-side implications in trust-intensive, knowledge-based services such as healthcare. Focusing on the introduction of the "Doctor AI Avatar" in online healthcare platforms, this study examines its causal impact on physicians' online consultation volumes. Drawing on Task–Technology Fit (TTF) theory, we argue that AI avatars are better suited to standardized, low-uncertainty tasks, potentially diverting routine demand away from physicians and leading to a decline in their consultation volumes. We further explore moderating effects related to departmental characteristics and physician heterogeneity. Using a propensity score matching–difference-in-differences (PSM-DID) approach with physician-level panel data, we identify the net effect of AI avatar adoption on professional service outcomes.

WB06

303

OR for Healthcare Decisions

Invited Session

Medical Decision Making

Chair: Saumya Sinha, University of Minnesota, Minneapolis, MN, United States

1 - Decentralized Training Decentralized Execution Using a Shared Network (Dtde-Sn): A New Reinforcement Learning Approach for Epidemic Modeling

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States, Sonza Singh

Multi-agent reinforcement learning (MARL) methods are suitable tools for modeling highly dynamic decision analyses such as intervention analyses in epidemic outbreaks across multiple jurisdictions. MARL models allow us to consider jurisdictional mixing interactions in the decision-making policies, as they evaluate decisions specific to each jurisdiction while considering the interactions between all jurisdictions. In multi-agent reinforcement learning, a commonly used paradigm is the decentralized training with decentralized execution (DTDE). This paradigm trains agents independently to optimize team rewards, and each agent regards other agents as a part of the environment. The standard DTDE approach trains a separate network for each agent. This approach is computationally complex, and this complexity grows with the number of agents. To solve this issue of scalability, we propose a new reinforcement learning training and execution approach, where a single deep learning model is trained for all agents (which considers the interactions between agents) but each agent makes its own independent decisions. This new method, which we call decentralized training with decentralized execution using a shared network (DTDE-SN) models interactions between agents while reducing computational complexity. The DTDE-SN leverages the unique features of epidemics, namely that the value network is a representation of the system dynamics model and the assortative property in jurisdictional mixing, to mitigate issues of non-stationarity while maintaining faster convergence. We applied the proposed framework to an HIV compartmental simulation model to demonstrate that the DTDE-SN method can identify intervention decisions for multiple interacting jurisdictions in a scalable manner.

2 - Distributionally Robust Template-Based Chemotherapy Scheduling with Override Policies Under Patient-Mix Mismatch and Duration Uncertainty

Xinyi Mao, Georgia Institute of Technology, Atlanta, GA, United States, Yiling Zhang, Yu-Li Huang

Outpatient chemotherapy units commonly employ fixed scheduling templates that partition the operating day into predefined appointment slots with nominal treatment durations. While templates facilitate operational standardization and workforce planning, they face two major sources of uncertainty: mismatches between the daily patient mix and template slots, and variability in treatment durations. In practice, infusion centers allow override policies, such as combining consecutive slots, splitting longer slots, or assigning shorter treatments to longer slots, to accommodate structural mismatches. However, these decisions are typically made based on nominal durations and do not explicitly account for treatment time uncertainty, which can lead to undesirable patient waiting, chair idle time, and overtime. In this talk, we present a two-step distributionally robust optimization (DRO) framework for template-based chemotherapy scheduling. In the first step, appointments are allocated to chairs under duration uncertainty. In the second step, given the chair-level allocation, we jointly optimize override decisions, sequencing, and time allocation. Both models use a moment-based ambiguity set for treatment durations. We derive a tractable 0–1 second-order cone programming (SOCP) reformulation of the resulting DRO models. Numerical experiments demonstrate that incorporating uncertainty reduces total operational cost by up to 39% compared to deterministic scheduling.

3 - Simultaneously Expiring Organ Offers: the Benefits of Variable Batch Sizes

Subramanian Raghavan, University of Maryland-College Park, College Park, MD, United States, Zhenyu Yue, Michael Fu

To reduce organ discards due to prolonged allocation processes, a simultaneously expiring batch-offer policy using fixed batch sizes has been proposed to replace the currently used sequential procedure. The main drawbacks of this policy are the risks of disappointment (when a patient agrees to accept the organ but does not receive it) and redundant work (unnecessary evaluations performed for patients later in the batch than the one who receives the organ). We quantify these metrics to improve upon the proposed fixed batch size policy and develop variable batch sizing policies, where the batch size can be dynamically adjusted during the organ offer process. We prove that a Greedy policy is optimal for minimizing discard subject to per-patient constraints on disappointment and redundant work. Additionally, we formulate a Dynamic Programming (DP) approach that balances expected discards and the expected number of disappointed patients. Using a probabilistic model calibrated to empirical kidney-allocation data, we compare fixed and variable batch size policies. We find that optimized variable batch sizes are superior to fixed batch size strategies, with the DP policy achieving the highest efficiency under perfect information. Relative to the sequential baseline (25% discard rate), variable policies can reduce the discard rate to approximately 18.6%---indicating the potential to increase the number of kidney transplants in the US by 1920 annually. Furthermore, sensitivity analysis reveals that while the DP policy

is theoretically optimal, the simpler Greedy policy is significantly more robust to systematic misspecification, making it a superior candidate for practical implementation.

4 - Robust Surgery and Bed Scheduling Under Coupled Duration and Los Uncertainty

Jing Liu, Johns Hopkins University, Baltimore, MD, United States, Farzin Ahmadi, Diego Martinez, Enzo Zavala, Rudy Geissbuhler, Kimia Ghobadi

We develop a data-driven framework for integrated operating room (OR) and downstream bed scheduling for elective surgeries under uncertainty. Leveraging ML-based case duration prediction and survival modeling of post-operative length of stay (LOS), we formulate a joint distributionally robust optimization model that captures both OR and downstream uncertainty. We introduce a service-level duration-to-LOS coupling mechanism that dynamically adjusts the LOS ambiguity set based on case duration stress. We evaluate the model in a rolling-horizon framework with dynamic downstream capacity constraints across ten hospital units using two years of real-world data from Johns Hopkins Hospital. Compared to historical practice, our approach achieves 15% of days within the target 75–85% utilization range (versus 3% for the nominal model), while maintaining comparable downstream bed over-capacity, demonstrating a more favorable balance between OR efficiency and congestion risk.

WB07

304

HAS Student Paper Competition

Awards Session

Health Applications Society

Co-Chair: Aiqi Zhang, Wilfrid Laurier University, Waterloo, ON, Canada

Co-Chair: Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States

1 - Optimal Assignment Guidelines for Maternity Waiting Homes in Resource-Constrained Environments

Yuhang Du, London Business School, London, United Kingdom, Jérémie Gallien, Qi (George) Chen, Jody Lori, Dimitris Panayiotou, Joseph Sieka

Maternity waiting homes (MWHs) are increasingly used to increase skilled birth attendance in regions with scarce obstetric services. We seek to optimize their assignment policy, that is whether and when a pregnant woman should go to her local MWH given available accommodation capacity and individual characteristics including living location. Our model leverages clinical data on labor onset and latent phase duration to minimize the predicted number of unassisted deliveries resulting from such policy in the local population. We present explicit mathematical solutions for different variants of this problem and integer program formulations of discretized approximations capturing practical implementation constraints. Our numerical experiments leverage detailed primary and secondary data on the population and health facilities of Bong County, Liberia. They suggest that our proposed pooling and assignment policy interventions for the existing MWH capacity in that region may plausibly halve the incidence rate of unassisted deliveries and reduce the probability of unassisted deliveries for the most vulnerable women from 70% to 15%. Implementing data-driven individual assessments of unassisted delivery risk and MWH assignment policies may thus constitute impactful and low-cost interventions in regions with limited MWH accommodation capacity.

2 - Hospital Admission and Patient Transfer Control: An Offline Doubly–Robust Learning Approach

Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, United States, Esmaeil Keyvanshokooh, Arlen Dean, Mark Van Oyen

Capacity-driven patient transfers, moving a current inpatient to a lower-acuity unit to free a bed for a new arrival, are among the most consequential yet least systematized decisions in hospital operations. Poorly timed step-downs can expose patients to adverse events including clinical deterioration, ICU readmission, and increased mortality. We formulate the joint admission and transfer control problem as an event-driven Markov decision process with unit capacity constraints, and prove that the optimal policy has a threshold structure: proactive transfers should begin before a unit reaches full capacity, with the threshold depending on congestion

across units. Because system dynamics are not fully known in practice, we develop an Offline Doubly-Robust Learning (ODRL) framework that learns transfer policies from historical electronic health record data, enforcing per-state capacity constraints and using pessimistic estimation to guard against limited data coverage. In a computational study using patient encounters from a large academic medical center, the optimal policy initiates ICU transfers at approximately 73% occupancy, substantially earlier than current practice, and the data-driven ODRL policy closes 76% of the gap between hospital practice and the full-information optimum while maintaining strict capacity constraints.

3 - Benchmark label errors distort clinical AI capability claims and model training

Junze Ye, Stanford Graduate School of Business, Stanford, CA, United States, Mohsen Bayati, Daniel Tawfik, Alex Goodell, Nikhil Kotha, Mark Buyyounouski

Benchmark scores increasingly support clinical AI capability claims and post-training reward signals, yet medical benchmarks often rely on LLM-assisted reference labels whose validity is rarely independently verified. We audit MedCalc-Bench v1.0, a widely used benchmark for medical score computation, using multi-agent recomputation, targeted physician review, and corroboration by independent frontier models. At least 27% of test labels are likely erroneous or incomputable. On 50 high-disagreement test-split instances reviewed by physicians, our recomputed labels agree with physicians' labels 74% versus 20% for the originals (95% CIs, 60-84 and 11-33). On these same instances, three frontier models from independent developers (OpenAI, Anthropic, and xAI, none sharing the model family used for recomputation) agree with the physician labels 61% of the time versus 12% for the originals, corroborating the physician labels without relying on our pipeline. Re-evaluating frontier LLMs against the recomputed labels raises measured accuracy by 15 to 23 percentage points relative to the originals, and a model trained on the recomputed train-split labels as reinforcement-learning rewards outperforms one trained on the originals by 13.3 percentage points (CI, 5.2-22.1; evaluated on the physician-labeled instances), showing that benchmark labels act as teachers and not only as yardsticks. Independently, archived snapshots of Anthropic's Claude for Healthcare launch page report MedCalc-Bench accuracies for three Claude models that differ by 19.5-27.5 percentage points across snapshots of the same page, illustrating that LLM-assisted clinical benchmarks currently lack versioning conventions that would let readers interpret such shifts. LLM-assisted clinical benchmarks require versioned stewardship before supporting capability claims.

4 - Beyond ATE: Multi-Criteria Design for A/B Testing

Kaining Shi, University of Chicago, Chicago, IL, United States, Jiachun Li, David Simchi-Levi

Adaptive clinical trials face a fundamental operational and ethical tension: balancing in-trialvpatient welfare (cumulative regret), acquiring precise subgroup-level efficacy evidence (CATE estimation), and identifying optimal post-trial treatment policies (simple regret). While each objective is critical for regulatory approval and broader clinical deployment, existing designs typically optimize them in isolation or rely on restrictive parametric assumptions. In this pa per, we develop a unified, multi-objective framework for adaptive experimentation in a general non-parametric setting. First, we characterize the fundamental statistical limits—the instance dependent Pareto frontier—governing the intrinsic trade-off between in-trial patient harm and CATE estimation accuracy. To navigate this trade-off, we propose ConSE, a sequential seg mentation and elimination algorithm that dynamically discretizes the patient covariate space to achieve this Pareto-optimal frontier. Second, we establish a critical decoupling theorem: any Pareto-optimal trial design inherently guarantees optimal post-experiment treatment policies (simple regret), regardless of the specific welfare-estimation trade-off chosen during the trial. For trial designers and healthcare operators, this provides a vital ”safety net”—prioritizing pa tient safety during the trial does not compromise the long-term quality of the resulting precision medicine guidelines. We validate our framework using synthetic data and the Collaborative Peri natal Project dataset, and introduce a practical causal forest-guided feature selection method to scale this approach for high-dimensional clinical operations.

5 - Constraint-Aware Self-Improving Large Language Model for Clinical Role Model Generation

Che-Yi Liao, Georgia Institute of Technology, Atlanta, GA, United States, Esmaeil Keyvanshokooh, Gian Garcia

As generative AI in healthcare moves from text documentation to treatment decision support, its central challenge shifts from fluency to verification: healthcare systems need mechanisms to decide whether generated

outputs are clinically meaningful, safe, and locally usable. We study this challenge for personalized treatment planning through Clinical Role Models (CRMs), patient-like reference trajectories that help clinicians identify feasible treatment plans. Local CRM databases are sparse, especially for underrepresented patients and rural health systems. Medical large language models (MLLMs) could augment these data but, without proper verification, may generate clinically invalid profiles. Accordingly, we develop CASE (Constraint-Aware Active Self-Improving Fine-tuning), a framework that couples an MLLM with an optimization-based verifier for CRM generation. The verifier checks feasibility, risk-reduction, uncertainty, and effort-to-change (EtC) constraints. Both verified and rejected candidates are then converted into instructional fine-tuning signals for the MLLM. CASE allocates fine-tuning efforts through an active-learning policy that targets patients where reliability and patient-centeredness can improve jointly. Theoretically, we (i) prove that CASE admits high-probability dynamic regret bounds that separate optimization, sampling, and surrogate-approximation errors, and (ii) develop online patient-specific safety certificates controlling false safety certification. In a cardiovascular disease case study leveraging granular clinical datasets, we demonstrate that CASE is the only approach among several benchmarks that jointly improves reliability and patient-centeredness. Moreover, CASE doubles the reliability of a local database-only approach, showcasing the benefit of leveraging MLLMs for closing data gaps. Overall, this research generates insights on designing verification mechanisms to facilitate MLLM integration into clinical decision support.

WB08

305A

Global Health Decision Analytics for Prevention, Detection, and Resource Allocation

Invited Session

Global Health and Humanitarian Operations

Chair: Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

1 - A Generative Agent Framework to Simulate Behaviorally Realistic Cancer Screening Trajectories

Yuewei Ling, Stanford University, Stanford, CA, United States, Marissa Reitsma

Cancer screening decisions are shaped by heterogeneous behaviors and individual contexts, yet simulation models used to shape guidelines and evaluate interventions typically parameterize screening as a fixed population-level rate. To address this limitation, we developed CARE-M, a generative agent framework that simulates interpretable and behaviorally grounded longitudinal cancer screening trajectories. CARE-M models individual screening decisions through a structured cycle of Context, Appraisal, Response, Evaluation, and Memory, allowing decisions to be informed by prior screening decisions, evolving reasoning, and individual-level heterogeneity in decision-relevant attributes, including healthcare access and attitudes towards preventive care. At each decision point, the model generates a screening action together with six appraisal ratings informed by the Health Belief Model, a brief evaluation, and an updated memory summary that is carried forward across screening cycles. We demonstrate proof-of-concept using breast cancer screening with mammography, which is recommended biennially for women beginning at age 40. When simulating decisions for three interpretable personas (proactive, pragmatic, and avoider), CARE-M produced distinct longitudinal screening trajectories. Decision appraisal profiles consistently differentiated personas, providing an opportunity to interrogate behavioral drivers of health and inform the tailoring of interventions to specific populations. Our framework and proof-of-concept evaluation suggest that generative agents may provide a useful in silico approach for studying and intervening on behavioral mechanisms underlying heterogeneity in cancer screening adherence. Future work will extend this study by incorporating human feedback in the design of personas and evaluation of decisions, and by examining the impacts of health policies and interventions.

2 - A Semi-Supervised Multi-Modal Framework for Early Cognitive Impairment Detection in Chronic Pain Patients

Huaiyang Zhong, Virginia Tech, Blacksburg, VA, United States

Introduction: Early identification of cognitive impairment among older adults with chronic pain is vital for prevention and timely intervention. However, predictive modeling is hindered by severe class imbalance and the difficulty of integrating information in electronic health records (EHRs).

Methods: We conducted a retrospective cohort study using EHR data from a not-for-profit health care network in the United States between 2016 and 2022. The cohort included 70,954 patients aged 55 years or older with at least one chronic pain diagnosis. Of these, 1,333 developed incident cognitive impairment, 15,901 were labeled as cognitively normal, and 53,766 were unlabeled. A Multi-Modal Pre-CSA framework was developed to combine structured EHR variables with text features extracted by the BERT-base-cased model. Token embeddings (768 dimensions) were pooled and reduced to one latent component using Truncated SVD before integration with structured features. To handle class imbalance, a new Prevalence-Informed Confident Sinkhorn Allocation (Pre-CSA) algorithm was applied, performing optimal-transport pseudo-labeling under prior prevalence constraints reflecting expected underdiagnosis.

Results: When trained on structured data only, Pre-CSA achieved an AUC of 82.0% and an F1-score of 77.7%, outperforming XGBoost (AUC 80.6%) and other semi-supervised methods. Incorporating BERT-derived text features further improved performance, with the Multi-Modal Pre-CSA reaching an AUC of 82.5% and an F1-score of 78.3%.

Conclusions: The Pre-CSA framework effectively addresses label scarcity and class imbalance while integrating structured and unstructured EHR data. It provides a practical and scalable approach for early identification of patients at elevated risk of cognitive impairment and supports risk-informed prioritization in real-world settings.

3 - *CANCELED A Societal Pomdp Framework for Optimizing Screening in Multimorbid Populations

Ali Hajjar, MBSC, King Abdullah Economic City, Saudi Arabia, Oguzhan Alagoz

Existing clinical practice guidelines tend to overlook the unique requirements of patients with multiple chronic conditions (MCC) by primarily focusing on the management of individual diseases. This inadequate approach leads to adverse events and places a financial strain on the healthcare system, particularly in countries with large or growing populations. To address this issue, our study utilizes a stochastic modeling approach to optimize disease screening decisions while considering financial and resource limitations from a broader societal perspective.

WB09

305B

Practical AI #4: Operationalizing Analytics in Healthcare

Invited Session

Healthcare Practice

Chair: Stefania Sokolowski, Memorial Sloan Kettering Cancer Center, New York, NY, United States

1 - From Idea to Implementation: Using AI to Optimize How a Cancer Center's Strategy Team Works

Stefania Sokolowski, Memorial Sloan Kettering Cancer Center, New York, NY, United States

As AI capabilities rapidly expand, strategy and operations teams at healthcare institutions have a growing opportunity to rethink how they work — automating routine tasks, accelerating analysis, and freeing capacity for higher-value work. This presentation describes a systematic approach to AI adoption developed within the Strategy & Growth department at Memorial Sloan Kettering Cancer Center, with the goal of fundamentally improving how the team operates.

We detail a structured process for identifying, categorizing, prioritizing, implementing, and tracking AI use cases across a non-clinical business function. Central to this process is determining which use cases warrant dedicated data science development versus those better addressed through other pathways — whether that means socializing existing tools to broader users, enabling non-technical users to build their own solutions, or flagging use cases for further exploration.

To illustrate the framework in practice, we present two applied data science use cases completed through this process, walking through how each was surfaced, scoped, developed, and evaluated for impact. Together, these examples demonstrate how a strategic and repeatable methodology can help strategy teams work smarter and drive greater organizational impact.

This work offers practical guidance for analytics and strategy teams at cancer centers and health systems seeking to build internal AI capabilities in a disciplined, outcome-oriented way.

2 - From Benchmark to Bottom Line: a Relative-Price Method to Project Plan Savings at Mskcc

Lauren Johnson, MSKCC, New York, NY, United States

By analyzing the gross collection ratio of Medicare- and commercially-insured patients at Memorial Sloan Kettering Cancer Center (MSKCC), we calculated the relative price of care for newly treated commercial patients. The objective of this analysis was to prove our overall cost efficiency thereby enhancing our relationships with self-insured employers and unions. Presenting these metrics to a prominent New York City union facilitated a collaboration with their health plan. Together, we assessed the relative price experience for their population within the NYC metro area and explored the potential cost savings through active navigation to MSKCC.

This collaborative research culminated in establishing a Center of Excellence program. The program offers significant benefits to commercially insured patients, including cost-shares and travel stipends funded by the health plan for those receiving care at MSKCC.

3 - AI in Healthcare: Connecting Members, Providers, and Data for Intelligent Care

Shalmali Joshi, Elevance Health, Atlanta, GA, United States

Despite unprecedented growth in healthcare data, many organizations still struggle to translate fragmented information into coordinated, high-quality care. This session explores how artificial intelligence is bridging the gaps between members, providers, and data systems to enable more intelligent, connected care delivery.

Focusing on real-world applications, this talk highlights how AI integrates diverse data sources—clinical records, claims, and member-generated data—to create a unified, actionable view of patient health. Attendees will learn how AI-driven solutions enhance member engagement through personalized outreach, empower providers with timely decision support, and streamline care coordination across complex healthcare ecosystems.

The session will also address key challenges in operationalizing AI, including data interoperability, workflow integration, trust, and scalability. Through practical examples, it will demonstrate how organizations can move beyond siloed analytics toward connected, end-to-end intelligence that supports both proactive interventions and improved patient outcomes.

Participants will leave with a clear understanding of how to design and implement AI use cases that not only connect systems, but also drive meaningful action—transforming data into intelligent care that is more personalized, efficient, and equitable

4 - Comparing Objective Wait TIMES with Subjective Patient Experience Survey Scores in Ambulatory Oncology

Sarah Kadish, Dana-Farber Cancer Institute, Boston, MA, United States, Angel Cronin, Yixian Chen, Caitlin Donohue, Kelly Eng, Brian Finn, Benjamin Maslow, Michael Hassett

Background: In many health systems, appointment wait time is a commonly low scoring, yet hard to address, patient experience measure. Since 2011, Dana-Farber Cancer Institute has used workflow-integrated operational data, collected by Real Time Location System (RTLS) badges worn by patients and staff, to objectively quantify waits. We evaluated associations between subjective patient experience surveys and objectively measured waits

across six ambulatory oncology locations to identify actionable targets and determine whether communication about delays mitigates dissatisfaction.

Methods: We linked patient experience survey responses with RTLS-based wait-to-first provider for over 15,000 ambulatory exam visits in 2025. The primary outcome was the top-box (very good) response to the timeliness survey question. Multivariable logistic regression estimated associations between RTLS wait and very good timeliness ratings, adjusting for covariates. A secondary analysis for encounters with waits >30 minutes assessed the relationship between communication about delays and timeliness.

Results: For timeliness, approximately 70% of encounters were rated very good, 20% good, and 8% fair/poor, with median RTLS waits of 39, 50, and 66 minutes, respectively. Very good timeliness probabilities were 79%, 69%, and 57% at waits of 30, 60, and 90 minutes. Male sex, older age, and clinic site were associated with better timeliness ratings. Among encounters with a wait >30 minutes, the proportion rating timeliness very good was 92% when communication about delays was rated very good, versus 21% and 3% when rated good or fair/poor. Leveraging objective wait times with subjective experience surveys may help organizations set targets, benchmark performance, and mitigate dissatisfaction.

WB10

306A

AI and Data-Enabled Medical Decision Making

Contributed Session

AI

Chair: Gokalp Cosgun, Wayne State University, 663 Prentis Street, Detroit, MI, 48201, United States

1 - LLM-Based Optimization of Radiotherapy Patient Scheduling with Real-World Constraints

Eric Zhang, Westlake High School, Westlake Village, CA, United States, Wen Li, Youfang Lai, Xun Jia

Radiation therapy patient scheduling is complex due to clinical and operational constraints, including treatment continuity over multi-fraction courses and added time for first fractions. Current practices are largely manual and labor-intensive. This work investigates the use of a large language model (LLM) to automate patient scheduling while satisfying clinical and operational constraints and balancing competing scheduling objectives.

Patient data were simulated using statistical distributions derived from clinical practice. Over 260 treatment days, representing one clinical year, 1,400 new patients were generated and grouped into categories sharing scheduling requirements. ChatGPT 5.4 was employed, and prompts were designed to intake daily new patients given an existing treatment calendar and generate updated schedules.

The LLM consistently generated schedules satisfying predefined feasibility rules across all tested scenarios. Treatment times were highly consistent within each treatment week: approximately 60% of evaluated fractions were scheduled with no within-week time shift, and approximately 99% were scheduled within one hour of the patient's other treatment times that week. In a separate scenario designed to reduce unnecessary machine changes, the machine-switching rate decreased from approximately 55% to 10%, with moderate trade-offs in idle time and workload balance. The framework also identified all tested infeasible scheduling requests and proposed corrective actions rather than generating invalid schedules.

This study demonstrates the feasibility of using an LLM to replicate expert-level radiation therapy scheduling logic under complex clinical constraints. It shows promise for improving scheduling efficiency, adaptability, and resilience without requiring explicit mathematical formulation.

2 - A Data-Driven Approach to Ivf Treatment Personalization: Improving Outcome Prediction and Patient Experience

Antra ., Masters' Union School of Business, Gurugram, India

In Vitro Fertilization (IVF) is a widely used treatment for infertility, which involves fertilization of egg and sperm in a laboratory setting; later, the resulting fertilized egg is transferred to the uterine cavity with the goal of achieving pregnancy. This research focuses on developing a predictive model using a combination of clinical and embryological parameters to predict pregnancy outcomes. Infertility can create a complex interplay of social, emotional, and informational challenges, making it difficult for couples to navigate their decision-making process. We have developed a machine learning model to help couples make informed choices and enable fertility experts to provide evidence-based consultations. The choice of selection of *gametes* is one of the most emotionally challenging decisions couples face during treatment, and this choice is significantly influenced by various age-related factors. This study contributes to the literature on emotional decision-making and addresses information asymmetry between patients and fertility experts in the IVF sector. The study highlights how factors such as source selection, embryo quality, and transfer techniques play a crucial role in IVF outcomes, enabling couples to make more informed decisions about their treatment journey. This knowledge helps navigate the complex emotional and clinical decisions involved in the process. Our findings offer valuable insights that can significantly aid patients in making informed and emotionally sound decisions regarding their IVF treatment. By understanding the impact of various clinical and embryological factors, patients are better equipped to navigate the complexities of IVF and make choices that align with their personal and medical needs.

3 - Integrating AI-Assisted Telemedicine into Congested Specialist Care: a Dynamic Flow Diversion Model

Abraham Seidmann, Questrom Business School, Boston University, Boston, MA, United States, Opher Baron, Fanying Chen

Physician shortages and prolonged access delays have reached crisis levels in specialty care delivery. This highlights the urgent need for innovative solutions like AI-assisted telemedicine, which can be strategically integrated into congested systems to deliver substantial reductions in waiting times—without the need for costly and often infeasible permanent capacity expansion. Motivated by our field collaboration with a large HMO serving approximately 2.5 million members, where dermatology waiting times average 24 business days and can exceed 120 days, we introduce the Dynamic Flow Diversion (DFD) mechanism.

The DFD algorithm activates an on-demand, AI-assisted remote care channel only when in-person queues for dermatology exceed an economically calibrated congestion threshold. This targeted approach ensures healthcare professionals can trust the system's safety and reliability, as eligible patients receive timely virtual consultations supported by AI-driven diagnostic tools and third-party board-certified physicians, ensuring both scalability and clinical oversight.

We model this system as a two-dimensional continuous-time Markov chain with state-dependent routing and derive performance metrics, including waiting times, overflow probabilities, and costs. This collaborative approach helps healthcare leaders feel engaged and confident in setting optimal activation thresholds to balance delays and costs. Extending to multi-clinic networks, we establish staffing rules using heavy-traffic approximations.

Empirical calibration using ~55,000 dermatology appointments show that diverting fewer than 10% of patients can reduce waiting times by over 70% at modest costs, inspiring confidence in healthcare leaders about its effectiveness. DFD offers a rigorously grounded, operationally feasible pathway to improve access, equity, responsiveness, and patient satisfaction in specialist care systems.

4 - The Hidden Cost of Convenience: Downstream Utilization Impact of Telemedicine in Outpatient Care

Yiwen Lu, University of Pennsylvania, Philadelphia, PA, United States, Yuqing Lei, David Asch, Yong Chen, Christian Terwiesch

As telemedicine becomes a routine option in outpatient care, health systems need to understand not only whether virtual visits substitute for in-person care, but also how they reshape downstream demand. We study downstream utilization impact (DUI), defined as how the choice of initial visit modality changes a patient’s subsequent use of care. Using outpatient electronic health record data from Penn Medicine, we examine four high-virtualizability clinical categories and compare downstream utilization after telemedicine versus face-to-face index visits.

Our framework centers on a simple 2×2 behavioral matrix based on whether downstream telemedicine follow-up increases and whether downstream face-to-face follow-up increases. This matrix distinguishes four patterns: convenience-driven use, escalation into in-person care, broader expansion across both channels, and minimal downstream impact. Rather than treating telemedicine as a single intervention with a single effect, this framework shows how the same modality can produce different downstream pathways for different types of care.

We find that downstream effects are heterogeneous across clinical categories. Mental, behavioral, and neurodevelopmental conditions are more often convenience-driven, with additional downstream use occurring mainly through telemedicine. Musculoskeletal conditions show a stronger escalation pattern, with more follow-up shifting into face-to-face care. These findings suggest that telemedicine should be managed as a condition-specific channel strategy rather than a uniform substitution tool. For managers, the key implication is that telemedicine policy affects not only access at the point of entry, but also the composition and volume of future demand across care channels.

WB11

306B

Precision Health Through Advanced Data Science

Invited Session

Healthcare Technology and Bioinformatics

Chair: Mostafa Reisi, University of Florida, Gainesville, FL, United States

Co-Chair: Xubo Yue, Northeastern University, Chestnut Hill, MA, United States

1 - Toward Temporal Causal Representation Learning with Tensor Decomposition: a Study on Ehr Data

Mostafa Reisi, University of Florida, Gainesville, FL, United States, Xubo Yue

Temporal causal representation learning is a powerful tool for uncovering complex patterns in observational studies, which are often represented as low-dimensional time series. However, in many real-world applications, data are high-dimensional with varying input lengths and naturally take the form of irregular tensors. To analyze such data, irregular tensor decomposition is critical for extracting meaningful clusters that capture essential information. In this paper, we focus on modeling causal representation learning based on the transformed information. First, we present a novel causal formulation for a set of latent clusters. We then propose CaRTeD, a joint learning framework that integrates temporal causal representation learning with irregular tensor decomposition. Notably, our framework provides a blueprint for downstream tasks using the learned tensor factors, such as modeling latent structures and extracting causal information, and offers a more flexible regularization design to enhance tensor decomposition. Theoretically, we show that our algorithm converges to a stationary point. More importantly, our results fill the gap in theoretical guarantees for the convergence of state-of-the-art irregular tensor decomposition. Experimental results on synthetic and real-world electronic health record (EHR) datasets (MIMIC-III), with extensive benchmarks from both phenotyping and network recovery perspectives, demonstrate that our proposed method outperforms state-of-the-art techniques and enhances the explainability of causal representations.

2 - Conditional Residual Diffusion for Clinical Multivariate TIME Series Imputation

Adam Meyers, University of Miami, Miami, FL, United States, Ramiro Jara

Missing data are common in clinical time series due to irregular sampling and physical sensor failures. Existing approaches for multivariate time series imputation (MTSI) may be broadly categorized as deterministic (e.g., transformers) or probabilistic (e.g., variational autoencoders). Deterministic approaches typically provide strong point estimates but fail to capture uncertainty, while probabilistic models offer calibrated uncertainty estimates but struggle to simultaneously capture global structure and fine-grained variability.

We propose CTD-MTSI, a conditional residual diffusion architecture tailored for clinical time series imputation. The proposed model uses an estimate from a baseline deterministic model and performs conditional diffusion in

the residual space, focusing on high-frequency corrections and uncertainty around the baseline prediction. In particular, the diffusion model can leverage a high-dimensional representation learned by the deterministic model, allowing it to refine the baseline prediction in a more structured and efficient manner. This formulation bridges deterministic and probabilistic paradigms, combining the strong inductive bias and stability of point estimation with the uncertainty quantification of diffusion models.

We evaluate the proposed framework on the PhysioNet 2012 ICU dataset, a standard benchmark with substantial missingness. Experimental results show that CTD-MTSI outperforms both deterministic and probabilistic baselines, achieving improvements in reconstruction accuracy (MAE/MSE) while maintaining stronger probabilistic calibration. The proposed model is particularly well-suited for clinical time series, where capturing both reliable point estimates and calibrated uncertainty is critical for downstream decision support and risk prediction. In addition, the proposed model offers flexibility in the choice of deterministic backbone, enabling its use in other domains.

3 - A Dual-Granularity Bayesian Active Learning Framework with Uncertainty Quantification for Sleep Apnea Severity Estimation

Xin Zan, University of Iowa, Iowa City, IA, United States

Sleep apnea requires accurate severity assessment for effective clinical management, yet fine-grained annotation of physiological signals is costly and labor-intensive. Existing machine learning approaches largely rely on strongly supervised learning with densely labeled data, which is often impractical in clinical settings. This work incorporates clinical knowledge to develop a dual-granularity Bayesian active learning framework for accurate and reliable fine-grained apnea severity estimation. To address uncertainty arising from limited fine-grained labels and noisy signals, we integrate active learning with Bayesian neural networks to quantify uncertainty and iteratively query the most informative instants for labeling, augmenting initial coarse-grained data. A Bayesian deep learning model captures both aleatoric and epistemic uncertainties, and a novel acquisition function combines epistemic uncertainty with knowledge misalignment for sample selection. A dual-granularity negative Gaussian log-likelihood loss unifies training with both coarse- and fine-grained labels. Experiments on real sleep data demonstrate improved accuracy, robustness, and annotation efficiency.

4 - A Late-Fusion Multimodal AI Framework for Privacy-Preserving Deduplication in National Healthcare Data Environments

Mohammed Omer Shakeel Ahmed, UTHealth Houston, Irving, TX, United States

I would like to present my research on a privacy-preserving multimodal AI framework designed to detect duplicate records in healthcare and CRM systems without relying on personally identifiable information (PII). Duplicate records continue to undermine data quality, analytics accuracy, and operational decision-making yet traditional deduplication methods depend heavily on identifiers such as names, emails, or SSNs, which are often masked under GDPR and HIPAA regulations.

In this talk, I will walk through a novel approach that integrates three independent signal types to achieve accurate deduplication while maintaining strict privacy compliance:

- Semantic embeddings of names and cities using DistilBERT,
- Behavioral patterns extracted from login timestamps, and
- Device metadata encoded from browser and operating system information.

Using a late-fusion strategy and DBSCAN clustering, my model achieves strong recall and a competitive F1 score compared to traditional string-matching baselines, even when data is noisy or incomplete. This demonstrates the feasibility of deduplication in environments where PII is restricted or unavailable.

During the session, I will explain the system architecture, experimental methodology, evaluation metrics, and opportunities for real-world deployment in enterprise CRMs and national health data infrastructures. My goal is

to show how multimodal AI can strengthen data integrity, improve public health analytics, and support ethical, privacy-first innovation.

This talk is ideal for attendees working in data science, healthcare IT, CRM systems, or anyone interested in practical and responsible AI applications.

WB12

306C

Scheduling for Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Kimberly Villalobos Carballo, New York University, New York, NY, United States

1 - An Operational View on Managing Mass Trauma Events

Noa Zychlinski, Technion - Israel Institute of Technology, Haifa, Israel

Mass Trauma Events (MTEs), such as wars, natural disasters, and terror attacks, place severe strain on mental health systems by generating surges in demand for both immediate and long-term psychological care. Early intervention is critical, as conditions such as post-traumatic stress disorder (PTSD) can have lasting individual and societal consequences, yet practical policies for delivering timely care remain limited.

We study the coordination of group and individual therapy in a multi-server queueing framework. While group therapy can rapidly address acute demand, it may increase downstream demand for individual treatment. Our model captures this trade-off and the interaction between therapy channels, while incorporating key features such as patient no-shows and dropouts. Using a fluid approximation, we derive index-based scheduling policies tailored to post-MTE surge, recovery, and long-term phases, accounting for time-varying and transient dynamics.

Using a case study based on real data from a large-scale MTE, supplemented by evidence from prior events, we show that the proposed policies substantially outperform a commonly used benchmark adapted to include group therapy, no-shows, and dropouts. Our approach shortens the recovery phase by approximately six months, reduces queue lengths by 31%, and increases total cost savings by 52%, driven by explicit coordination across therapy channels.

These results offer simple, implementable guidance for practitioners and policymakers seeking to improve mental health service delivery and system resilience following MTEs.

2 - Mixed Integer Linear Program for Family Medicine Residency Scheduling with Clinic First Objectives

Ankit Bansal, State University of New York, Binghamton, Binghamton, NY, United States, Mohammad Al Syouf, Jean-Philippe Richard, Bjorn Berg

Family medicine physicians make up the largest share of the primary care provider (PCP) workforce, accounting for 40% of all PCPs. A notable contributor to the shortage of PCPs is the tendency of family medicine residents to pursue sub-specialties or careers as hospitalists. This issue has been primarily linked to a training gap within the family medicine residency curriculum. In practice, family medicine physicians work mainly in outpatient primary care clinics—known within the field as continuity clinics—where they focus on long-term care and maintaining continuity in the physician-patient relationship. However, traditional residency training often emphasizes hospital-based experiences, creating a disconnect between training and practice. The Clinic First paradigm is an educational approach developed to close this gap by prioritizing outpatient clinic experience and reinforcing continuity of care. In alignment with this model, we introduce a Mixed Integer Linear Programming (MILP) model designed to generate annual rotation schedules for family medicine residents. This model captures the complex constraints imposed by accreditation standards and program-specific requirements. It explicitly incorporates two key objectives that reflect the Clinic First paradigm: Clinic Time Consistency (CTC) and Level-Loading (LL). To address the computational challenges, we develop heuristic models that efficiently

produce high-quality feasible solutions that can be used to significantly speed up the solution of the problem with Gurobi. Computational experiments based on data from an existing residency program demonstrate the effectiveness of our approach.

3 - Multi-Resource Allocation and Care Sequence Assignment in Patient Management: a Stochastic Programming Approach

Rema Padman, Carnegie Mellon University, Pittsburgh, PA, United States, Xinyu Yao, Karmel Shehadeh

To mitigate outpatient care delivery inefficiencies induced by resource shortages and demand heterogeneity, this paper focuses on the problem of allocating and sequencing multiple medical resources so that patients scheduled for clinical care can experience efficient and coordinated care with minimum total waiting time. We leverage highly granular location data on people and medical resources collected via Real-Time Location System technologies to identify dominant patient care pathways. A novel two-stage Stochastic Mixed Integer Linear Programming model is proposed to determine the optimal patient sequence based on the available resources according to the care pathways that minimize patients' expected total waiting time. The model incorporates the uncertainty in care activity duration via sample average approximation. We employ a Monte Carlo Optimization procedure to determine the appropriate sample size to obtain solutions that provide a good trade-off between approximation accuracy and computational time. Compared to the conventional deterministic model, our proposed model would significantly reduce waiting time for patients in the clinic by 60%, on average, with acceptable computational resource requirements and time complexity. In summary, this paper proposes a computationally efficient formulation for the multi resource allocation and care sequence assignment optimization problem under uncertainty. It uses continuous assignment decision variables without timestamp and position indices, enabling the data-driven solution of problems with real-time allocation adjustment in a dynamic outpatient environment with complex clinical coordination constraints.

4 - Overbooking Surgical Blocks: a Fair and Scalable Miqp Approach to OR Utilization

Kristy Zhou, Memorial Sloan Kettering Cancer Center, New York City, NY, United States, Upasana Raval

Operating rooms (ORs) are among the most expensive and resource-intensive assets in a hospital, and optimizing their utilization is critical for financial sustainability and timely patient access to surgical care. As cancer cases continue to rise globally, institutions face mounting pressure to increase surgical capacity without expanding physical infrastructure — a process that can take two to ten years and incur enormous costs.

This presentation introduces a novel approach to improving OR utilization through surgical block overbooking that provides surgeons with fair, reliable, and advance-scheduled OR time.

The core challenge is ensuring equitable distribution of overbooking risk across surgeons. We present a mixed-integer quadratic programming (MIQP) model that formalizes fairness as an objective while enforcing operational hard constraints and fairness soft constraints. The model is embedded in an R-Shiny web application integrated with Python and the Gurobi solver, enabling scheduling managers to generate, validate, modify, and communicate monthly overbooking schedules without modeler involvement.

Deployed across two campuses at a major academic cancer center, the system contributed an incremental 2% and 3.5% improvement in OR utilization respectively. Between 2016 and 2021, the institution was able to provide OR access to 23% more cases with the same physical infrastructure — demonstrating that optimization-driven overbooking is a scalable and implementable strategy for sustainable surgical demand growth.

WB13

307

Health Innovation and Analytics

Invited Session

eBusiness

Chair: Maria Baldeon Calisto, Wake Forest University, Pfafftown, NC, United States

1 - *CANCELED An Evaluation of Vision Foundation Models for Head and Neck Tumor Segmentation.

Maria Baldeon Calisto, Wingate University, Wingate, NC, United States

Vision foundation models (FMs) have shown strong potential for adapting to diverse downstream tasks without specialized fine-tuning. However, their performance in head and neck cancer (HNC) tumor segmentation remains insufficiently explored. This work provides a comprehensive evaluation of five prominent FMs for HNC tumor segmentation. The FMs tested include SAM, MedSAM, nnInteractive, UniverSeg, and SegGPT. The evaluation of SAM, MedSAM, and nnInteractive utilized a bounding box prompting strategy. The influence of prompt precision is assessed by systematically enlarging the prompt bounding box until it reaches twice the size of the tightest bounding box enclosing the region of interest. For UniverSeg and SegGPT, the prompt consisted of example image-label pairs, and the effect of increasing the size of this support set is tested. Two-factor repeated measures ANOVAs were used for statistical analysis, confirming that both the FM architecture and the prompting strategy exert a statistically significant effect on segmentation performance for both primary tumors (GTVp) and metastatic lymph nodes (GTVn). For SAM, MedSAM, and nnInteractive larger bounding boxes lead to notable performance degradation. Moreover, MedSAM outperformed SAM and nnInteractive for GTVp segmentation, whereas SAM provided slightly better results than the competitors for GTVn segmentation. For UniverSeg, performance improved as more image-label pairs are added to the support set, while SegGPT remained unaffected by this factor. Overall, UniverSeg and SegGPT demonstrate limited few-shot capability, with low mean dice scores across all scenarios. Comparisons against supervised CNN models trained specifically for HNC segmentation showed that conventional architectures remain more stable and clinically reliable.

2 - Transforming Early Autism Detection by AI-Powered Video-Based Gait Analysis

Ziteng Wang, Northern Illinois University, Dekalb, IL, United States, Sinan Onal

We leverage AI (computer vision and predictive models) in developing an innovative technique to screen for autism in children by analyzing parent-recorded walking videos. This approach complements the existing autism diagnostic and screening practices by offering an objective, accessible, low-cost, and scalable method for all individuals.

3 - Toward Trustworthy Health Information: Automated Quality Assessment of Health News Using Fine-Tuned LLMs

Lan Sang, University of Colorado Boulder, Boulder, CO, United States, David Dobolyi

The quality of online health information can significantly influence public trust in healthcare decisions. Despite efforts like expert-driven reviews, large-scale, manual assessments have proven cost prohibitive and unsustainable. This study proposes an automated, scalable approach to evaluate health information quality using fine-tuned large language models (LLMs). Using an expert-annotated dataset sourced from HealthNewsReview.org, this study benchmarked several state-of-the-art LLMs, including open-weights models from Meta’s Llama and Alibaba’s Qwen series. Our best fine-tuned model demonstrated strong performance in accurately classifying information quality. By automating complex, criterion-based assessments that have traditionally been conducted by expert reviewers, this work provides a foundation for accessible and practical tools that help users evaluate health information, supporting informed public health decision-making.

4 - The Monarch Initiative: Bridging Biomedical Knowledge and Clinical Care for Rare Diseases

Shawn O'Neil, University of North Carolina at Chapel Hill, Chapel Hill, NC, United States, Melissa Haendel

Rare diseases, which individually affect fewer than 0.05% of the population, are collectively prevalent: over 10,000 are known, impacting over 30 million Americans, half of whom are children. Diagnosis remains a major challenge, requiring an average of over 4 years from symptom onset to definitive identification. Many go undiagnosed altogether, and many conditions lack an available ICD code, complicating patient journeys.

The Monarch Initiative supports rare disease research and diagnostics through curated integration of phenotypic,

genetic, and other biomedical knowledge. The Initiative also manages the Mondo disease ontology, a carefully organized and comprehensive catalogue of rare and genetic diseases. Mondo serves a central role in the wider Monarch knowledge base, which links diseases and their subtypes to genetic variants, cellular mechanisms, and observable phenotypes. References to primary literature are included, providing a rich information source for both research and clinical practice.

In this presentation we'll provide an overview of Monarch resources and recent developments, including ongoing additions of data sources and utilities serving researchers, clinicians, and patients alike. We'll also discuss the new availability of Mondo in clinical settings via Epic and IMO Core, adding over 5,000 new disease codes for clinical documentation and unifying a fragmented landscape of rare disease care.

Wednesday, July 29, 2026, 12:15 PM - 1:30 PM

Ballroom A

Posters, Munch & Mingle

Poster Session

1 - 01- An Interactive Multi-Criteria Decision-Support Tool for Equitable Food Retail Siting

Vaahin Mehta, University of Southern California, Los Angeles, CA, United States, Abigail Horn

Many Americans lack access to healthy food, contributing to food insecurity and diet-related chronic disease. Low-quality food environments (FEs), such as food deserts with limited grocery access and food swamps saturated with fast food, disproportionately affect low-income and minority populations. Siting new healthy food retail is inherently a multi-criteria decision: policymakers must weigh competing objectives in collaboration with researchers who define and prioritize them. Yet, current practice relies on simple proximity metrics rather than modeling frameworks that integrate multiple objectives, and lacks interactive tools allowing policymakers and researchers to jointly explore trade-offs and refine priorities. This study presents an interactive geospatial dashboard designed to help regional planners and policymakers identify areas in Los Angeles County where food access interventions are most needed and feasible. We develop a multi-factor weighted scoring model synthesizing indicators of food access (static and mobility-derived), social vulnerability, and neighborhood opportunity across multiple geographic levels (Census Block Groups, Council Districts, etc.). The tool enables users, envisioned as policymakers and researchers, to select priority factors, assign custom weights, and apply constraints reflecting zoning feasibility and land availability. The multi-criteria decision framework computes composite spatial scores, producing a ranked list of priority intervention areas and a live interactive map. Co-developed with the LA County Departments of Regional Planning and Public Health, this open-source dashboard is designed for direct use by planners and public health staff without technical expertise. It enables systematic, multi-criteria siting decisions that reflect locally defined priorities, bridging research and practice to support equity-aligned food access planning.

2 - 02- Impact of LEAN Interventions on Operating Room Turnover Times

Donrey Acharon, Rowan-Virtua School of Osteopathic Medicine, Stratford, NJ, United States, Andrew Barnish, Dylan Bagley, Nicholas Vaccaro, Christopher Middleton, Edem Tetteh

Hospital operating rooms (ORs) are complex healthcare systems where inefficiencies result from interactions between personnel and processes. This systematic review evaluates the impact of LEAN project management interventions on OR efficiency by focusing on performance metrics of turnover times. We searched PubMed/MEDLINE, Embase, and Scopus for studies reporting quantitative OR efficiency outcomes pre- and post-LEAN intervention. The included studies showed that application of LEAN tools such as kaizen, stakeholder engagement, value stream mapping, and process maps resulted in improved OR efficiency through reduced turnover times. These findings indicate that there is quantitative evidence that demonstrates LEAN method's potential to eliminate waste and optimize current processes of the hospital OR system.

3 - 03- A Multi-Objective Stochastic Optimization Model for Routing of the Mobile Lung Cancer Screening Unit in Rural Areas: A Case Study in Oklahoma

Kiana Amani, University of Oklahoma, Norman, OK, United States, Iakovos Toumazis, M. Soheil Hemmati

Oklahoma is among the most heavily affected states by lung cancer, ranking 8th in age-adjusted incidence (63.3 per 100,000) and 5th in mortality (45.3 per 100,000) in the United States. Although low-dose computed tomography (LDCT) is the recommended screening method for high-risk populations, participation rates remain low nationwide and are particularly limited in Oklahoma (about 10% compared to 16% nationally). To address access barriers in rural and underserved regions, mobile LDCT screening programs have been introduced. However, these programs are typically deployed in a reactive, ad-hoc manner rather than through systematic, data-driven planning. This study proposes a multi-objective stochastic routing optimization model for a mobile lung cancer screening unit that explicitly accounts for uncertainty in both patient participation and screening outcomes. The model incorporates county-level incidence rates, observed uptake patterns, and geographic factors to determine effective service routes in areas with limited access to fixed screening facilities. Our primary objective function is to maximize the overall net benefit of screening, measured in quality-adjusted life years (QALYs) gained from early detection, while the associated cancer screening program cost is treated as the secondary objective function. The framework produces optimal routing strategies across representative scenarios and reports key performance metrics, including the total number of individuals screened and both screening-related and net QALYs achieved.

4 - 04- Characterizing Latent Developmental Risk Profiles in Preterm Children: A Combined Growth Modeling and Machine Learning Approach

Li-Sheng Wu, National Tsing Hua University, Hsinchu City, Taiwan, Yu-Hsin Chen

Characterizing developmental trajectories for preterm children is essential for the precise allocation of limited intervention resources to optimize long-term functional outcomes. Identifying distinct growth patterns in cognitive, language, and motor domains allows for a targeted distribution of occupational, speech, and physical therapy. While traditional assessment frameworks rely on static measurements, these snapshots fail to capture the dynamic and heterogeneous nature of development. Consequently, this study identifies the multidimensional drivers of real-world developmental trajectories, offering a framework for feature-based characterization of longitudinal developmental patterns.

We first utilized Latent Growth Modeling with a covariate-adjusted framework to identify the demographic and clinical drivers of developmental trajectories. By incorporating variables such as sex, gestational age, socioeconomic status, and Cerebral Palsy severity, we quantified how these factors significantly shift the baseline levels and the longitudinal rate of change across cognitive, language, and motor domains. We then applied Growth Mixture Modeling to categorize children into distinct developmental phenotypes. To understand the underlying composition of these groups, we utilized XGBoost to identify the key features that distinguish different developmental trajectory groups. This approach shifts the focus from simple diagnostic labels to a multi-feature profiling system.

These findings highlight substantial developmental heterogeneity and demonstrate the value of growth-based modeling for characterizing early developmental differences and informing targeted intervention planning. By uncovering the specific features that drive divergent pathways, this research provides a data-driven basis for more efficient resource allocation tailored to a child's long-term developmental patterns.

5 - 05- Assessing the long-term health impact of sleep through mortality-associated EEG and ECG dynamics

Dat Diep Khanh, National Tsing Hua University, Hsinchu, Taiwan, Yu-Hsin Chen, Cheng-Bang Chen, Diane C. Lim

Sleep is a fundamental physiological necessity, comparable to eating, drinking, and breathing, and it plays a critical role in maintaining human health. Given the close relationship between sleep and health, disruptions in

sleep quality may adversely affect an individual's health. To examine this relationship, various assessments can be performed, including physical examinations, imaging, and bio-signal analysis. However, these methods are generally considered point-in-time assessments and are limited in their ability to explain the long-term impact of sleep on health. To enable long-term evaluation, we consider mortality, as the ultimate health outcome, to be a meaningful endpoint for investigating the relationship between sleep and health. In this study, we objectively assess sleep quality using electroencephalography (EEG) and electrocardiography (ECG) signals and investigate their association with mortality outcomes. Heterogeneous recurrence quantification analysis (HRQA) is applied to multimodal EEG and ECG recordings to extract nonlinear dynamical features of sleep. Dimensionality reduction and clustering techniques are then employed to identify distinct sleep-related phenotypes, followed by survival analysis using the Kaplan–Meier estimator. The results reveal clear stratification of mortality risk across patient clusters derived from objective sleep features, indicating that EEG- and ECG-based sleep dynamics are significantly associated with survival outcomes. These findings suggest that sleep–mortality relationships may provide a useful framework for assessing long-term sleep-related health status.

6 - 06- Beyond Efficiency: A Scoping Review of Patient Safety, Data Privacy, and Medico-legal Challenges in AI-Driven Clinical Documentation

Sara Ali, Tele Geriatric Research fellowship, Okemos, MI, United States, Leena Khalafallah, Eiman Ahmed, Nadia Elbaghdadi, Hajer Ali, Azza Ahmed, Mohammed Omer, Nadir Abdelrahman

Background: AI-driven clinical documentation (CD) tools generate clinical records to reduce administrative burden and improve doctor–patient interactions but raise concerns regarding accuracy, safety, and legality.

Objectives: This scoping review aimed to map current evidence on AI-based CD tools and their implications for documentation accuracy, patient safety, data privacy, and medico-legal accountability.

Methods: Following PRISMA-ScR guidance, a systematic search of PubMed, Embase, Web of Science, IEEE Xplore, and Google Scholar was conducted from 2020-2025. Studies on AI use in CD in healthcare settings were included.

Results: 28 studies with diverse designs were included. Findings were synthesized across four domains. Large language models (LLMs) and ambient scribes were most frequently evaluated. Omission errors were the most commonly reported, comprising up to 83.8% of errors, and were strongly associated with LLMs, highlighting completeness as a key system design challenge. Hallucination rates ranged from 1.47% in optimized pipelines to 36% in general-purpose applications.Privacy concerns included third-party data sharing, limited transparency, and weak regulatory oversight. Multiple biases - algorithmic, linguistic, demographic, and automation - highlighted the complexity of AI integration. Evidence on medico-legal accountability was limited, although several studies emphasized the need for clearer legal frameworks.

Conclusion: AI-assisted documentation tools may improve workflow, but variability in error rates, particularly omissions, raises concerns about reliability and patient safety. Persistent privacy concerns, regulatory uncertainty, and limited medico-legal evidence underscore the need for human-in-the-loop validation, enforceable governance, and cautious implementation. Future research should prioritize standardized evaluations of error rates and omission risks, alongside clearer regulatory and legal frameworks.

7 - 07- *CANCELED Stacked Against Care: Machine Learning Analysis of Cumulative Social Risk and Preventive Care Underutilization Among U.S. Adults Using BRFSS 2022–2024

Zaili Gu, Northeastern University, Medford, MA, United States, Ajit Appari

Preventive health services such as routine medical checkups and influenza vaccination are essential components of population health management, yet utilization remains persistently unequal across U.S. adults. While individual access barriers are well-documented, the compounding effect of multiple simultaneous social risks on preventive care engagement remains unexamined using nationally representative data with direct social determinant measurement.

This study uses pooled data from the Behavioral Risk Factor Surveillance System (BRFSS) 2022–2024, incorporating the newly introduced Social Determinants and Health Equity (SD/HE) optional module, to examine cumulative social risk and preventive care underutilization across 541,873 adults in 33–37 states annually. A cumulative SDOH risk index is constructed from four binary domains — financial hardship, utility insecurity, transportation barriers, and employment instability — yielding scores from 0 to 4. Two outcomes are examined: missed routine checkup and missed influenza vaccination in the past 12 months.

Weighted prevalence estimates reveal a consistent dose-response: checkup underutilization rises from 20.1% among zero-risk individuals to 35.6% among those with 3–4 simultaneous risks (+15.5 percentage points); flu shot underutilization rises from 52.6% to 73.9% (+21.3 points). XGBoost models with SHAP interpretation identify lacking a personal doctor as the dominant driver of checkup gaps, while age and education dominate flu vaccination underutilization — pointing to fundamentally different intervention strategies for each outcome. Temporal analysis (2022–2024) shows checkup disparities are persistent but stable, while flu vaccination gaps are widening, with high-burden individuals worsening by 4.1 percentage points. Findings identify targetable population profiles to guide preventive care outreach design.

8 - 08- Health, Wealth, and Education as Structural Drivers of AI Readiness

Jeffrey Chavis, University of South Carolina, Columbia, SC, United States

This study examines the structural factors that shape how countries adopt and integrate Artificial Intelligence (AI). It argues that AI readiness arises not solely from technological capacity, but from long-run development systems that shape a nation's ability to innovate, govern, and sustain complex technologies. The analysis centers on three foundational pillars of AI adoption: population health, economic capacity, and educational attainment. Drawing on harmonized international data from the World Bank, World Health Organization, UNESCO, and the United Nations Development Programme, the study constructs a cross-sectional dataset of 193 countries to evaluate how these structural systems relate to national AI readiness scores.

Rather than identifying where AI is most advanced, the study investigates why readiness varies systematically across countries. It tests whether life expectancy, public health investment, income levels, research intensity, and access to advanced education operate as structural conditions that shape national preparedness. The methodology integrates multivariate regression, dimensionality reduction, and nonlinear modeling to assess robustness across specifications while maintaining cross-country comparability.

The findings indicate that Observed AI readiness aligns with the configuration of development systems. Economic capacity establishes the baseline, while education and health drive divergence across countries national performance within comparable income levels. Progress in AI therefore depends not only on technological investment, but on the institutional and human foundations that sustain long-term innovation.

9 - 09- Impact of Socioeconomic Status on Complications and Readmissions After Hip and Knee Arthroplasty in Diabetic Patients

Anneke Parkinson, Jacobs School of Medicine and Biomedical Sciences , Buffalo, NY, United States

Introduction: Diabetes mellitus is associated with increased postoperative complications and readmissions following total hip arthroplasty (THA) and total knee arthroplasty (TKA). Socioeconomic status (SES), encompassing income, education, and access to healthcare resources, may influence surgical outcomes. Limited data exist on its impact among diabetic patients undergoing joint arthroplasty. Understanding the interplay between diabetes and SES is critical for identifying high-risk patients and informing strategies to reduce complications and improve perioperative equity.

Methods: A retrospective cohort study was conducted using the TriNetX database including adult patients with diabetes who underwent primary THA or TKA. Patients were categorized as low SES or non–low SES based on indicators such as financial insecurity, insufficient insurance coverage, or housing-related hardship. Propensity score matching (1:1) balanced baseline demographic and clinical characteristics. Postoperative outcomes within 90 days included major complications and healthcare utilization.

Results: A total of 134,282 diabetic patients were identified, including 5,199 with low SES. After propensity score matching, 5,123 pairs were analyzed. Low SES patients had higher rates of postoperative sepsis (OR 1.946, p=0.0086), myocardial infarction (OR 3.276, p=0.0003), and pneumonia (OR 1.857, p=0.0042). Emergency department visits were also more frequent in the low SES group (OR 2.023, p<0.0001). No significant differences were observed in surgical site infections, urinary tract infections, acute kidney injury, ventilator requirement, or hospital readmission rates.

Conclusion: Low SES is associated with increased complications and emergency department utilization among diabetic patients undergoing these procedures, highlighting the importance of incorporating social determinants of health into perioperative risk assessment and postoperative care.

10 - 10- Biomarker-Based Prediction of Healthcare Utilization and Health Outcomes Across Cardiovascular-Kidney-Metabolic Syndrome Stages: An Interpretable Machine Learning Analysis of NHANES 2017–2018

Tajma Thongrakard, Northeastern University, Boston, MA, United States, Ajit Appari

Cardiovascular-kidney-metabolic (CKM) syndrome affects one in three U.S. adults, yet clinicians lack validated, population-level tools to identify which patients are approaching the next disease stage before a clinical event forces treatment escalation. This study developed an interpretable machine learning framework using nationally representative NHANES 2017–2018 data (n=9,254) to predict healthcare utilization, functional health outcomes, and CKM stage progression risk across the American Heart Association's five-stage classification. The analytic sample covered all CKM stages (Stage 0: 37.0%; Stage 1: 14.9%; Stage 2: 39.0%; Stage 3: 0.3%; Stage 4: 8.8%). Primary biomarker predictors included high-sensitivity C-reactive protein, neutrophil-to-lymphocyte ratio, systemic immune-inflammation index, triglyceride-glucose index, UACR, and eGFR, alongside demographic and socioeconomic covariates. Three supervised classifiers were trained: regularized logistic regression (LR), random forest (RF), and gradient boosting (GB). RF achieved the highest AUC-ROC for overnight hospitalization (0.728), difficulty accessing care (0.708), functional limitation (0.879), and fair/poor self-rated health (0.842). For Stage 2-to-3 progression—the primary clinical target among 3,610 Stage 2 participants (positive rate=2.0%)—GB achieved near-perfect discrimination (AUC=1.000; LR=0.996, RF=0.997). Overall classification accuracy was 88% for hospitalization and 84% for self-rated health. SHAP decomposition identified UACR, eGFR, hs-CRP, HbA1c, and the triglyceride-glucose index as the dominant Stage 2→3 predictors, each mapping directly to AHA guideline treatment escalation thresholds. Fairness was assessed across racial/ethnic and income subgroups to evaluate differential model performance. This framework offers a population-validated, biomarker-informed signal to support guideline-aligned treatment escalation decisions across primary care, cardiology, nephrology, and endocrinology settings.

11 - 11- *CANCELED AI-Based Validation and Operational Efficiency in Hospital PGHD Workflows Under Increasing Wearable Data Volume

Danial Shahrabi Farahani, Virginia Polytechnic Institute and State University, Blacksburg, VA, United States, Idris Adjerid, Christopher Zobel, Lara Khansa

Hospitals are increasingly expected to incorporate patient-generated health data (PGHD) from wearables into care delivery, yet the operational consequences of this shift remain poorly understood. While prior work shows that PGHD can improve monitoring, clinical decision support, and patient engagement, it also shows that high-volume PGHD can overwhelm staff, complicate interpretation, and disrupt workflow when data quality, relevance, and interoperability are not adequately managed. This study asks a focused Operations Management question: can AI-based validation help hospitals absorb increasing PGHD volume without sacrificing operational efficiency? Existing research suggests that AI can clean heterogeneous data, detect anomalies, reduce noise, prioritize relevant signals, and lower manual review burden, but evidence is still limited on its role as an operational mechanism for sustaining workflow performance in hospital settings.

We study hospital personnel who interact directly or indirectly with PGHD workflows, including clinicians and administrative staff. Using a mixed-method design, we first conduct interviews to identify workflow bottlenecks and current validation practices, then test whether AI-based validation mediates the relationship between PGHD

volume and workflow burden in a hospital survey. The study contributes by shifting the discussion from the general value of AI-enabled PGHD toward a more specific operational problem: how hospitals can convert growing wearable data streams into usable inputs rather than workflow overload. The findings are expected to inform scalable validation strategies for PGHD integration in hospital operations.

12 - 12- A Comparative Machine Learning Approach for Predicting Mode of Delivery Using CDC Natality Data

Megan Sambets, North Carolina State University, Raleigh, NC, United States, Maria Mayorga, Chandler Sambets

Since 2009, cesarean delivery rates in the United States have steadily increased, reaching 32.2% in the most recent report from the National Center for Health Statistics. This exceeds the 10–15% rate recommended by the World Health Organization. Elevated cesarean rates are associated with increased risks of maternal morbidity and mortality. We examine the relationship between maternal characteristics and mode of delivery using machine learning. Features may include, but are not limited to, maternal age, race, weight gain, body mass index, education level, gestational age, and use of labor induction or augmentation. The goal is to identify factors associated with cesarean delivery to inform strategies for reducing its prevalence in the United States. We apply supervised classification models, including XGBoost, logistic regression, and k-nearest neighbors to a random sample of 900,000 first-time mothers with term, singleton, cephalic pregnancies. Data are drawn from the 2024 CDC WONDER natality database. Given the importance of minimizing false negatives, model performance is evaluated using area under the ROC curve, recall, and F1-score. These findings provide insight into key predictors of cesarean delivery and may support clinical decision-making and targeted interventions to improve maternal and neonatal outcomes.

13 - 13- Access to Marijuana Among Midwest College Students

David Ruggeri, University of Missouri, Columbia, MO, United States

PURPOSE

The purpose of this study is to examine the prevalence of marijuana use among traditional, full-time college students.

RESULTS

Of currently enrolled, full-time college students, 43% reported having consumed marijuana within the past 30 days. Approximately 75% of college students indicated that they consumed marijuana 10 days out of the past month, with just over 7% reporting that they consume marijuana on a daily basis. 48% of student specified that consuming marijuana once a week was not harmful to one’s health, and no students indicated that this frequency of consuming marijuana was very harmful. Of the students surveyed, approximately 60% stated that they could get marijuana within one hour, with 35% claiming access within 30 minutes.

DISCUSSION/CONCLUSION

Marijuana use among college students is far more prevalent that shown in earlier studies. Previous studies concluded that past month marijuana use among college students range from approximately 20% to 27%.

14 - 14- Personalized Wellness Optimization - a Constrained Programming Approach to Cycle-Based Remedy Recommendations

Apurva Deepika Arni Sudhakar, UC Berkeley, Bellevue, WA, United States

Women's wellness applications have proliferated, yet remain limited to cycle prediction without actionable intervention. This work presents the first binary integer programming formulation for personalized cycle-based wellness recommendation, bridging this gap by selecting interventions from a multi-domain remedy database while respecting individual time budgets and clinical safety constraints.

The framework operates on a daily user state vector capturing six dimensions: cycle phase, mood, symptoms, sleep quality, energy, and stress. Each remedy in the database carries three values 1) clinical effectiveness derived from systematic review, 2) phase suitability reflecting hormonal physiology, and 3) a personalized response rate that updates with user feedback. The optimizer combines these via dynamic weights inferred from the user's current state, then selects an optimal recommendation set subject to time, conflict, diversity, and cardinality constraints.

The poster illustrates the full pipeline: state vector input, weight inference logic, the BIP formulation with all five constraint types, and the constraint relaxation hierarchy that guarantees a feasible recommendation even when constraints conflict. A worked example demonstrates how the model handles a luteal-phase user with high stress and a 60-minute time budget, producing a five-remedy recommendation across nutrition, movement, lifestyle, and herbal categories.

Implementation in Python with PuLP and the open-source CBC solver achieves sub-50-millisecond solve times on mobile hardware. The framework prioritizes interpretability, hard constraint satisfaction, feasibility guarantees and properties essential for any clinical decision support context where transparent reasoning and safety must hold.

15 - 15- A Two-Stage Forecasting and Simulation Framework for Pediatric Emergency Department Capacity Planning

Ahmed Qasem, University of Alabama at Birmingham, Birmingham, AL, United States, Lori Pelletier, Abdulaziz Ahmed

Accurate forecasting of emergency department (ED) demand is essential for improving operational planning, yet forecasts are most useful when linked to decisions about capacity and patient flow. This study presents an integrated machine learning and discrete-event simulation framework for forecasting pediatric ED demand and translating those forecasts into actionable operational insights. Daily pediatric ED visit data from October 2021 through September 2025 were combined with weather variables, including maximum temperature, minimum temperature, and wind speed. Feature engineering generated 46 predictors capturing lagged demand, rolling statistics, and cyclic calendar patterns. An XGBoost model was trained to forecast daily ED visits at a 7-day horizon, with hyperparameters optimized using a 100-trial Bayesian search. Performance was evaluated on a strict 365-day temporal holdout set. Forecasted daily volumes were then used as inputs to a discrete-event simulation model of ED patient flow, where arrivals followed nonhomogeneous Poisson processes based on empirical hourly patterns, patient acuity reflected historical Emergency Severity Index distributions, and length of stay was modeled using acuity-specific probability distributions. The forecasting model achieved strong holdout performance, with test MAPE of 8.47% and MAE of 13.5. Simulation results showed that under a baseline capacity of 55 beds, average bed utilization was 79.3% and mean patient waiting time was 27.7 minutes. Increasing capacity to 60 beds reduced mean waiting time to 10.8 minutes, while 65 beds reduced it to 3.7 minutes. These findings demonstrate how integrating predictive modeling with simulation can support pediatric ED capacity planning and operational decision-making.

16 - 16- Embedding a RAG-based model to support dietary planning and nutritional insights in a regulated software as a medical device to support diabetes and obesity management

Anand Iyer, Welldoc, Inc., Columbia, MD, United States, Mansur Shomali, Abhimanyu Kumbara

Effective dietary management is a cornerstone for optimizing glycemic management and facilitating weight reduction in patients with diabetes and obesity (American Diabetes Association, *Standards of Care in Diabetes*, 2026). To address the complexities of personalized nutrition in chronic care, we developed an advanced, dual-model Retrieval-Augmented Generation (RAG) AI system. Crucially, this system was engineered as a regulated Software as a Medical Device (SaMD) in full conformance with FDA regulations to ensure clinical safety and efficacy.

Our initial RAG-based model provides dynamic meal planning support. By grounding the AI in established nutritional guidelines, the model synthesizes and generates personalized meal plans driven by a comprehensive

set of patient-specific parameters. These inputs include meal type preferences, dietary restrictions, continuous glucose trends, longitudinal weight trends, current medication regimens, and daily physical activity levels.

Complementing the meal planner, a second specialized RAG model processes this same multi-modal physiological dataset alongside the patient's logged dietary intake. This secondary model evaluates the consumed meals against the user's specific metabolic data to deliver actionable, AI-driven insights. It provides contextual recommendations and targeted food substitutions designed to stabilize glycemic response and support weight management.

By utilizing a RAG architecture within an FDA-compliant framework, we mitigate the risks of AI hallucinations while delivering a scalable, regulatory-compliant solution. This approach empowers patients with hyper-personalized, medically sound nutritional guidance for effective disease management.

17 - 17- Chronic Conditions as Risk Factors for Respiratory Syncytial Virus-Associated Hospitalization among Community-Dwelling Adults, 2016-2017 and 2017-2018 Seasons

Gordana Derado, CDC, Atlanta, GA, United States, Rebecca Woodruff, Michael Melgar, Monica Patton, Sarah Hamid, Huong Pham, Christopher Taylor

Respiratory syncytial virus (RSV) causes substantial morbidity among adults. Data on RSV hospitalization rates among adults with and without chronic conditions could accurately characterize certain groups' risk for hospitalization but are challenging to produce given multiple uncertainties in numerator and denominator estimation. We estimated RSV-associated hospitalization rates and adjusted rate ratios among community-dwelling adults with nine chronic conditions using population-based, laboratory-confirmed hospitalization numerator data from the RSV Hospitalization Surveillance Network and population denominators from the Behavioral Risk Factor Surveillance System and U.S. Census. Rate numerators were stratified by documented chronic conditions, and denominators were estimated for condition-specific catchment-area populations. Monte Carlo simulation jointly propagated numerator and denominator uncertainty by simulating 1,000 datasets for each model. Quasi-Poisson regression models with a scaled deviance term, adjusted for age group, sex, and race or Hispanic ethnicity, were fit. This process was repeated 100 times, yielding 100,000 estimates of rates and rate ratios. Median estimates and 95% uncertainty intervals were reported, demonstrating a generalizable framework for uncertainty-aware rate estimation using multiple public health data sources.

18 - 18- Defining Total Mission Value for Healthcare's Strategic Decisions About AI

Arwen Declan, Prisma Health, Greenville, SC, United States, Andrew Taylor

Emerging Artificial Intelligence (AI) technologies are poised to transform healthcare across a rapidly expanding array of applications, and healthcare organizations must assess AI's economic impact before adoption decisions and throughout the AI lifecycle. However, health economic analyses of AI technologies largely focus on cost metrics, such as cost-effectiveness, cost-utility, and return on investment, while excluding mission-driven values that are central to healthcare's purpose. This cost-centric approach risks decoupling strategic business decisions about AI from organizational mission, a misalignment that the broader business literature consistently links with diminished organizational performance. We argue that mission-value alignment is normatively essential for strategic business decisions like those about AI, and that healthcare organizations should integrate mission-driven values with financial metrics as they define AI's total impact. We propose a conceptual framework, Total Mission Value (TMV), that integrates mission-driven value with cost-focused analyses for strategic business decisions regarding AI in healthcare. To define TMV's value domains, we draw on the Balanced Scorecard approach to mission-value integration from the business literature, analyze the Institute for Healthcare Improvement's Quintuple Aim for broad healthcare goals, and examine Mission, Vision, and Values statements for organizational strategic priorities. This analysis reveals five ethically grounded value domains that define TMV in healthcare: Patient Care, Staff Experience, Operations, Economic, and Education and Research. These domains establish a conceptual foundation for future work to define quantifiable outcome metrics, develop TMV scorecards for specific AI technologies, and implement TMV for strategic AI adoption and governance decisions in healthcare.

19 - 19- Can Generative Artificial Intelligence Think, Feel and Act Autonomously, Potentially Replacing Physicians?

Sunil Erevelles, University of North Carolina, Charlotte, Charlotte, NC, United States, Anthony Erevelles, Emma Knutsson, Misti Polić

Generative Artificial Intelligence (Gen AI) pushes the boundaries of a paradigm shift, and is radically transforming healthcare. Unlike *all* past technologies, it is not a "tool" under human control, but has "autonomous agency." It has the capacity to create, decide, execute and reproduce without human control. Thus, it arguably can be classified as an inorganic species, in accordance with the biological species concept. Once highly-respected, intellectually-intense professions, such as those of physicians, are increasingly being outperformed in purely cognitive tasks. The value of physician "cognitive labor" is in on a trajectory that will turn negative. Gen AI surpasses physicians in emotional intelligence, specifically because it has no emotions. Emotion is a merely a complex algorithm, and Gen AI transforms patients into complex algorithms, able to understand them better with each interaction. It metamorphosizes patients and treatments into simultaneously-evolving algorithms, able to precisely diagnose and treat. Moreover, advances in robotic Gen AI is on a trajectory to surpass physicians in some physcial tasks. Collectively, while Gen AI has numerous limitations, this paradigm shift will result in the radical transformation of healthcare.The objective of this research is to develop a theoretical framework for an autonomous physician network, based on Gen Al-centric logic. Using inductive-realist approaches, and drawing on Bayesian, Latent Variable, Natural Language Processing and Behavioral Reasoning Theories, the authors propose a framework, grounded in foundational premises for this fast-developing physcian healthcare model . This framework could help accelerate academic research and healthcare decision-making in an "Age of the Inorganic Superspecies."

20 - 20- Resource Planning for End Stage Treatments Under Uncertainty

Renato de Matta, University of Iowa, Iowa City, IA, United States

End-stage renal disease (ESRD) represents the terminal stage of chronic kidney disease, requiring renal replacement therapy through dialysis or transplantation. Although kidney transplantation offers superior clinical and economic outcomes, the persistent shortage of donor organs necessitates reliance on alternative modalities such as home-based dialysis and integrative care strategies, including palliative interventions. Contemporary ESRD programs face rising costs related to organ procurement, transplant services, and diverse dialysis modalities, highlighting the need for efficient resource allocation frameworks. To address this challenge, we develop a model that yields a tractable characterization of the tradeoff between hospital profitability and patient welfare. We quantify the operational cost of improving outcomes under prevailing incentives. The incentive structure is embedded in a stochastic resource planning framework that captures labor allocation, training investment, and treatment modality choice under uncertain organ availability. The resulting equilibrium characterizes ESRD care delivery as a second-best system where even fully rational hospitals select operationally efficient but welfare-distorted allocations because long-term patient outcomes, measured in quality-adjusted life years (QALYs), are non-contractible. Our analysis highlights structural limits of value-based payment and provides an operations-based explanation for persistent underinvestment in quality-enhancing care pathways.

21 - 21- Where Matters: Contextual Anomaly Detection

Ayush Gupta, University of Strasbourg, Strasbourg, France

Unsupervised anomaly detection in medical imaging is frequently hindered by two primary challenges: the high computational cost and data scarcity associated with 3D models, and the "anatomical blindness" of efficient 2D patch-based approaches. Furthermore, the clinical reality of "dirty" datasets, where fully healthy scans are rare, limits the training of effective normative models.

We propose a novel, context-aware multimodal framework that addresses these limitations by conditioning 2D patch-based reconstruction on explicit anatomical textual priors. Our method utilizes TotalSegmentator to perform two critical functions: (1) extracting organ-level anatomical labels that are encoded via a frozen

BioClinicalBERT to provide global context to local patches, and (2) identifying and filtering abnormal regions within existing clinical datasets to synthesize a "pseudo-normal" training distribution.

By integrating visual features with textual anatomical embeddings, the model learns a conditioned representation of normality tailored to specific organs. Experimental results on the BraTS 2021 and MOOD benchmarks demonstrate that our approach significantly outperforms context-blind baselines and heavy 3D architectures, achieving a state-of-the-art Dice Score of 0.71 on brain tumor localization. This framework provides a highly efficient, scalable solution for clinical anomaly detection, effectively bridging the gap between high-resolution 2D processing and 3D spatial intelligence while bypassing the need for curated healthy-only datasets.

22 - 22- Design and Implementation of an Explainable OR Scheduling System for Multi-Hospital Environments

Ayman Mohamedelkhair, Bradley University, Peoria, IL, United States, John Yoo, Jill Nora, Roopa Foulger, Jonathan Handler

Efficient operating room (OR) scheduling constitutes a complex operational challenge within multi-hospital healthcare systems, arising from the need to balance clinical priorities, resource constraints, and system-level efficiency. This study presents the design and implementation of a rule-based, explainable OR scheduling decision-support framework tailored for networked healthcare environments. The proposed system integrates structured patient and procedural inputs with institutional policies to support consistent and transparent allocation of surgical cases across hospitals and operating rooms. The framework adopts a multi-layered decision model. At the hospital level, eligibility is determined based on case complexity thresholds and available capability, ensuring that cases are directed to facilities with appropriate resources. Eligible hospitals are subsequently ranked using a composite criterion incorporating expert-selected preferred hospitals, patient-centric geographic proximity, and preferred surgery timing. At the operating room level, allocation is governed by strict compatibility constraints, including urgency classification, robotic equipment requirements, and department-specific reservation policies. A time-based allocation mechanism performs intra-day scheduling using a minute-resolution capacity model, identifying the earliest feasible time slot within a defined search horizon centered around the preferred procedure date. A distinguishing feature of the prototype system is its emphasis on interpretability; each scheduling outcome is accompanied by a documented rationale that explicates the underlying logic, thereby enhancing transparency, accountability, clinical effectiveness, and information management. The system architecture combines a relational data model with a web-based interface, enabling real-time validation and practical deployment. This work contributes a scalable and interpretable foundation for OR scheduling, advancing decision transparency in healthcare operations management.

23 - 23- Adolescent E-Bike Injury Risk: The Role of Social Norms and Exposure

Anya Dalal, California Association of Youth Commissions, Hillsborough, CA, United States, Nathan Kim, Leslie Ragsdale, Mandy Brown, John Maa

Electric bicycles (e-bikes) are transforming youth mobility, but their increasing use among adolescents introduces significant injury risk and public health concern, including inconsistent helmet use and normalization of risky behaviors. Despite growing policy attention, epidemiologic evidence on adolescent e-bike users remains limited, as existing studies rely on adult populations or injury surveillance data that do not capture underlying behavioral and social determinants. This study, based on one of the first large-scale surveys of adolescent e-bike use, analyzes data from a 2026 anonymous survey of 270 students in grades 5-12 across California, assessing riding frequency, helmet use, crash history, risky behaviors, awareness of laws, and parental, peer, and social media influences. A composite risk index and multivariable logistic regression models were used to estimate associations between social, behavioral, and exposure variables and outcomes, including helmet use, crash history, and risk behavior. Among respondents, 43% had never ridden an e-bike, while 12% reported regular use. Helmet use prevalence was 56% for consistent use; however, 71% of riders reported at least one crash. Crash prevalence increased with riding frequency (67% to 83%), indicating a clear exposure-response relationship. Helmet use was strongly associated with social determinants, with peer and parental influence significantly increasing consistent use. Awareness of e-bike laws was not associated with reduced risk behavior, and behavioral risk factors showed limited explanatory power for crash outcomes. Findings suggest adolescent injury

risk is driven by social norms and exposure, supporting population-level interventions targeting social environments and exposure reduction.

24 - 24- A Coordinated Drone and EMS Network Design for Automated External Defibrillators Delivery

Marziehsadat Rezaei, University of Louisville, Louisville, KY, United States, Carmine Sorgente, Domenico Serra, Monica Gentili, Evan Kuhl

Each year, over 350,000 out-of-hospital cardiac arrests (OHCAs) occur in the U.S., with survival rates as low as 8–10% due to delayed ambulance arrivals and limited use of publicly available Automated External Defibrillators (AEDs). Unmanned Aerial Vehicles (UAVs), or drones, present a transformative solution for rapid AED delivery to OHCA scenes, overcoming barriers like AED accessibility and bystander hesitation. Using real-world data from Jefferson County, Kentucky, and optimization and predictive models, we generated 210 candidate drone-only and coordinated EMS–drone network configurations and estimated survival for each. We linked survival and response outcomes to cost estimates from the literature. This approach allowed us to compare the cost and survival across configurations and identify the best trade-off between survival gains and investment value. Across all evaluated configurations, coordinated EMS–drone networks optimized for survivability achieved the greatest survival gains and favorable trade-offs between cost and survivability. The current EMS baseline had a mean response time of 7.68 min. In coordinated designs, mean response times were 8.9%–75.9% shorter than the EMS baseline. The expected number of survivors to hospital discharge under the EMS baseline was 126 (8.65%), which increased to 146 (10.04%)–357 (24.49%) across coordinated networks. The best-performing coordinated design with 95 platforms cost $7,403 per additional survivor. Survival gains plateaued beyond 95 platforms, indicating diminishing returns with further expansion. Coordinated drone and EMS systems for AED delivery were found to be cost-effective relative to Drone-only and standard EMS response systems.

25 - 25- Change-Point–Based Multimodal Prioritization Using Wearable Activity Data in Type 1 Diabetes

Yupeng Chen, Stanford University, Stanford, CA, United States, Jing Zou, Pierre-Amaury Laforcade, David Scheinker, Dessi Zaharieva

Background: The Stanford 4T study for newly diagnosed type 1 diabetes (T1D) uses continuous glucose monitoring (CGM) and the Timely Interventions for Diabetes Excellence (TIDE) dashboard to prioritize patients for review. Physical activity substantially affects glycemic control, but wearable-derived activity data are not routinely incorporated into T1D care, creating a need for a more holistic, multimodal review of patients that integrates activity and glycemic patterns.

Methods: We integrated Garmin wearable data with Dexcom CGM data from 4T participants and designed a change-point–based prioritization approach using deviations from patient-specific baselines in activity and glucose metrics. The primary measure was week-to-week change in average daily step count, with wear-time filtering to ensure interpretable patient-week observations (≥6 hours/day, ≥3 days/week). Patients were flagged when activity deviations aligned with shifts in glycemic metrics. We evaluated threshold-based rules and simulated their effect on weekly prioritization volume.

Results: Wear-time filtering produced interpretable patient-week observations and enabled reliable detection of activity and glucose deviations. Flagging thresholds directly controlled prioritization volume: a 10% step-count deviation flagged over 40% of eligible observations, whereas a 20% threshold reduced flags to about 25%, corresponding to roughly five patients reviewed per week. These results demonstrate a tunable, deviation-based prioritization mechanism that aligns sensitivity to meaningful change with clinician capacity.

Conclusions: Change-point–based activity metrics support multimodal prioritization in population-based T1D management. Deviation-based rules allow explicit control of review burden while identifying patients with activity-associated shifts in glycemic patterns, providing a scalable approach for integrating wearable data into proactive diabetes care.

26 - 26- Queueing-Informed Adaptive Alarm Policies for Continuous Glucose Monitoring in Diabetes Management

Yupeng Chen, Stanford University, Stanford, CA, United States, Naman Gupta, David Scheinker

Background: Continuous glucose monitoring (CGM) provides real-time glycemic visibility for proactive diabetes care. In practice, CGM systems generate frequent alerts for hypo- and hyperglycemia requiring timely response from patients and care providers. Static threshold-based alarms configurations overlook heterogeneous clinical urgency, time-varying volume, and limited caregiver capacity, contributing to alarm fatigue and delayed responses However, clinically desirable tighter glucose targets further increase alert burden. This tension between tighter control and alert burden motivates adaptive designs that prioritize clinically actionable events.

Methods: We analyze CGM time-series data from a pediatric type 1 diabetes cohort in the DCLP5 study. Static alarms include hypoglycemia (<70 mg/dL), hyperglycemia (>180 mg/dL), and tighter hyperglycemia (>140 mg/dL), mapped to high-, medium-, and low-priority alerts. Exploratory analysis shows strong time-of-day variation, with a peak-to-off-peak rate ratio of 1.58 (95% CI 1.53–1.63). We design adaptive alarm policies using time-varying thresholds and rate-of-change criteria to suppress low-urgency alerts during high-volume periods. Policies are evaluated using a non-preemptive priority *M/G/n* queue with discrete-event simulation over repeated daily horizons.

Results: We found that tightening glycemic thresholds increases alert frequency by 19%, while naïve robustness checks reduce this increase by over 50%. Relative to fixed schedules, adaptive policies reduce low-priority alerts by 8% while maintaining high-priority frequency, yielding 14–29% reductions in mean wait times, with the largest improvements during peak periods.

Implications: Queueing-informed adaptive policies reduce alarm burden while preserving responsiveness to clinically urgent events. The model provides a systematic approach to designing CGM alert strategies under limited attention.

27 - 27- A Queueing Microsimulation of Coordinated Entry Prioritization for Supportive Housing Allocation

Yupeng Chen, Stanford University, Stanford, CA, United States, Margaret Brandeau

Background: The U.S. population experiencing homelessness increased by 18% from 2023 to 2024. The U.S. Department of Housing and Urban Development requires local Continuums of Care (CoC) to operate Coordinated Entry processes that assess, prioritize, and refer people to scarce housing resources. Because supportive housing capacity remains insufficient relative to growing need, allocation decisions are consequential.

Methods: Methods: We developed a stochastic queueing microsimulation of the housing allocation process in the Santa Clara Continuum of Care. We simulated a cohort of 10,000 single adults experiencing homelessness and tested how prioritization rules performed when vulnerability assessments were subject to measurement error and group-specific bias. We compared the following prioritization rules: vulnerability-score thresholds, benefit-based allocation, hybrid vulnerability-benefit rules, and wait-time-adjusted prioritization.

Main Outcome and Measures: total months homeless, unsheltered homelessness, housing retention, deaths, overdoses, queue length, waiting time, and group-level service rates.

Results: The analyses found that strict vulnerability score-threshold rules generated discontinuities near eligibility cutoffs and became inefficient with errors in the assessment tool. Benefit-based rules reduced total months homeless and improved housing retention but shifted resources toward individuals with lower vulnerability. Wait-time-adjusted and hybrid rules reduced long waits while preserving much of the efficiency gain from targeted allocation.

Conclusions: This work clarify the tradeoffs created by Coordinated Entry prioritization rules, including efficiency, equity, waiting time, and housing stability. This framework provides a policy-testing environment for comparing housing allocation rules before implementation.

28 - 28- Modeling Maternal Health Dynamics Across Pregnancy

Beril Balibey, University of Michigan, Ann Arbor, MI, United States

This study examines the health conditions of pregnant women during pregnancy by developing a comprehensive state definition that represents maternal health at each interaction with the healthcare system. The proposed state definition is constructed by integrating evidence from the literature with available medical data, thereby combining theoretical knowledge with real-world clinical practice. It is designed to capture physical, mental, and psychological health dimensions, while also incorporating social determinants of health that may influence maternal outcomes. After defining the state vector, the study investigates measurement associations among its components to identify which health indicators are commonly measured together in clinical practice. This analysis helps reveal how different dimensions of maternal health are observed within healthcare encounters. In addition, the Continuous-Time Markov Model idea is applied to a smaller subgroup of the state vector, specifically vital signs such as blood pressure, respiration, heart rate, and temperature. This subgroup-level idea aims to provide insights into how vital-sign-based health states and measurement-frequency patterns evolve over time throughout pregnancy. Based on the preliminary investigation, transition probabilities of these vital-sign-based states and monitoring patterns vary across pregnancy stages.

29 - 30- Standardized BPMN 2.0 Modeling of In-Clinic Cervical Cancer Screening and At-Home HPV Self-Sampling Workflows in an FQHC

Ryan Suk, Emory University, Atlanta, GA, United States, Mackenzie Czerner, Maryam Kheirandish, Jessica Calderón-Mora

Introduction: Federally qualified health centers (FQHCs) are considering at-home HPV self-sampling to improve cervical cancer screening, but operational planning requires clear representations of delivery pathways to support stakeholder validation, explicit handoffs/role accountability, and further comparative evaluation and downstream operational analyses. We describe a structured-data-to-Business Process Model and Notation (BPMN) approach to map these workflows at an FQHC in El Paso, Texas.

Methods: Clinic staff and a research team completed a structured spreadsheet template capturing step descriptions, responsible roles, inputs/outputs, and approximate staff time. We translated the sheets into BPMN 2.0 collaboration diagrams with separate pools for Patient, Clinic (swimlanes for Front Desk, Medical Assistant, Provider), and external partners (Lab/Courier; Mail/Carrier for the at-home pathway). Sequence flows were used within pools, and message flows were used for cross-pool handoffs (e.g., specimen shipping, results return).

Results: The in-clinic workflow comprised 17 documented steps across three clinic roles, with a baseline of 70 minutes of total staff time per episode (Front Desk 15, Medical Assistant [MA] 39, Provider 16). The at-home self-sampling workflow comprised 6 documented steps primarily owned by MA (40 staff minutes) plus patient actions; the BPMN explicitly represented kit delivery, specimen transit, lab processing/results messaging, and an exception loop for lost-in-transit specimens with replacement kit reordering.

Discussion: Standardized BPMN models derived from structured site data produced shareable, role-explicit process maps for both screening modalities and a reusable BPMN XML foundation for subsequent simulation and budget impact analyses.

30 - 31- Leveraging Surgical Schedules to Improve Operating Room Supply Readiness Using Predictive Analytics

Akane Fujimoto Wakabayashi, ISyE Georgia Tech, Atlanta, GA, United States, Meredith DePuy, Dima Nazzal, Pinar Keskinocak

Incomplete surgical case carts are a common source of operating room (OR) delays, disrupting schedules and reducing efficiency. These challenges are driven by reliance on manual workflows and limited use of data-driven planning, making it difficult to anticipate supply needs as surgical demand evolves. This can result in both missing supplies and excess inventory.

This project develops a data-driven approach to improve surgical supply forecasting using scheduled surgery data, historical case data, and surgeon preference cards. We estimate baseline supply requirements and evaluate performance using fill rates, defined as the proportion of supplies requested on the day of surgery that are captured by schedule-based estimates. Results show that surgical schedules are frequently incomplete and dynamic, leading to systematic underestimation of supply needs and low fill rates for many items.

To address this, we introduce a forecasting method that adjusts schedule-based estimates using historical discrepancies between scheduled and completed cases. This approach improves fill rates compared to schedule-only methods, though with some trade-offs related to increased overstock risk. We also compare requested versus actual supply usage to identify consistent patterns of under- and over-picking in surgical case carts.

An interactive dashboard was developed to visualize predicted demand, usage trends, preference card performance, and scheduling dynamics. This decision-support tool enables more proactive inventory planning, with the goal of reducing OR delays, improving resource utilization, and supporting more reliable perioperative operations.

31 - 32- The Demand for Mixed-Methods Modeling in Health Systems

Cullen Jackson, Clemson University, Clemson, SC, United States, Steven Foster

The US is facing unprecedented demands for health services across domains. The natural complexity of challenges like an increasing elderly population, staffing shortages, and increasing cost of care are intrinsically coupled with challenges in designing infrastructure to meet this complex, increasing demand.

The Center for Human Factors at Prisma Health conducts health system research concerning health care operations challenges with human factors principles and methods. However, health system research and design require multidisciplinary teams to consider not only clinician workflows, but also complex patient flow dynamics, multi-level scheduling, rigorous secondary data analyses, and integration of hospitals' finances.

Operations researchers provide modeling expertise and data-driven approaches to address challenges associated with increasing health service demands. The Center for Human Factors in Health Care is engaged with projects evaluating drivers of systemic patient flow at a large hospital, emergency room front end process efficiency, and the relationship between pediatric hypertension diagnosis and blood pressure measurement workflows. Results from these studies identify complex interactions between policy, human behavior, patient flow, and other nested complex systems that necessitate the creation of model-based decision-support systems.

Given the inherent complexity of systems research in large sociotechnical systems, these projects are either proposed in collaboration with experts in mathematical modeling or will require the expertise of operations researchers for future intervention design and implementation. This work seeks to highlight the important need for mixed methods modeling in health systems, as exposed by existing human factors interventions in the largest health system in the state of South Carolina.

32 - 29- *CANCELED Confidence-Constrained UAV Routing for Adaptive Search Operations

Stephen Sunday, University of Louisville, Louisville, KY, United States, Monica Gentili

In disaster and emergency response scenarios, search and assistance operations are often hindered by damaged infrastructure, limited accessibility, and significant uncertainty in target locations. Traditional search approaches are inefficient in such environments, while Unmanned Aerial Vehicles (UAVs) provide a promising alternative for rapid and flexible deployment.

This work develops a confidence-constrained vehicle routing model for coordinating a fleet of UAVs in search operations over a target region. The proposed model explicitly integrates detection uncertainty into routing decisions by linking the probability of target presence with the number of required observations needed to achieve a prescribed confidence level. The resulting optimization problem jointly accounts for resource availability, flight time constraints, observation time, sensing capabilities, and spatial prior information on target locations.

To capture the dynamic nature of search operations, we embed the routing model within an iterative framework that alternates between (i) updating the spatial probability distribution of target locations using drone-collected data and (ii) re-optimizing UAV trajectories based on the updated information. This closed-loop structure enables adaptive allocation of search effort toward high-probability regions, improving detection efficiency over time.

Preliminary computational results on synthetic instances illustrate the value of integrating uncertainty and confidence requirements directly into routing decisions.

33 - 33- A framework for customizing cancer post-treatment follow-up: case of breast cancer

Abdelaziz Berrado, EMI, Mohammed V University in Rabat, Rabat, Morocco, Hanane Amroune, Ismail Chitaouy, Maissae Haddouchi, Loubna Benabbou

Clinical surveillance of cancer patients is necessary to ensure early detection of recurrence after curative treatment and to monitor patient progression. Although clinical guidelines commonly recommend fixed surveillance schedules for all patients, static intervals between follow-up visits may not be compatible with individual disease progression, increasing the risk of delayed recurrence detection for some cases as it maybe burdensome for healthcare ressources and patients for others.

Motivated by the idea of personalizing follow-up schedules based on each patient’s prognosis, we introduce a generic framework for an intelligent cancer post-treatment follow-up. The framework enables the prediction of time to recurrence using patient-specific clinical longitudinal data, which can be used to customize follow-up schedules for each patient.

The framework can enhance medical visit planning by incorporating evidence-based follow-up suggestions aligned with each patient's clinical development, thereby reducing under-surveillance risk, optimizing resource use and easing patient burden.

We illustrate the application of the proposed framework to a cohort of patients with locally advanced breast cancer undergoing neoadjuvant chemotherapy from the I-SPY 1 trial, a multicenter prospective study conducted across nine institutions and publicly available.

The approach, while validated for breast cancer, applies to other cancer types and diverse data modalities. Its application can support doctors in predicting relapse time, establishing optimal follow-up intervals, and enabling prompt treatment decisions, so improving the efficiency, efficacy, and tailoring of post-treatment care.

34 - 34- Health Economics and Cost Effectiveness Research at the National Institutes of Health

Jeremy Green, National Institutes of Health, Bethesda, MD, United States

The National Institutes of Health (NIH) continues to express its interest in health economics including analyses of cost efficiency (Simon 2026). To facilitate evaluation of NIH sponsored health economics research, and to encourage the planning of continued efforts in this area, we created a new dataset. For 2025, this new dataset includes detailed information on 621 NIH awards for cost effectiveness research, and $350 million in total funding. The states with the largest number of cost effectiveness research projects were California, Massachusetts, and New York, which accounted for approximately 35% of total cost effectiveness projects, and 31% of total cost effectiveness project funding. Most of the cost effectiveness projects were direct research projects, while 5 projects were training projects, and 5 projects were for Small Business Innovation Research and Small Business Technology Transfer programs. Nearly 80% of cost effectiveness research projects were conducted by institutions of higher education, with the remaining 20% of cost effectiveness research projects conducted by independent hospitals, research organizations, and other organizations. About 60% of NIH cost effectiveness research projects were led by departments of internal medicine, public health, and preventive medicine. The organizations with the largest number of NIH cost effectiveness research projects were University of California at San Francisco, Johns Hopkins University, University of Washington, and Massachusetts General Hospital which accounted for 20% of NIH cost effectiveness research projects, and 18% of NIH cost effectiveness research funding, for 2025.

35 - 35- *CANCELLED Optimizing Partial AUROC for Risk Score Systems

Yunxiao Liang, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen

We evaluate our framework on the SWYC dataset. Numerical experiments demonstrate that our scalable approach solves the pAUC optimization problem in a fraction of the time required by exact solvers. Compared to traditional baseline scoring systems, the proposed model significantly improves local sensitivity and pAUC at the target $\alpha=0.25$ threshold. Furthermore, the resulting integer-weighted scorecard remains completely transparent and easy to compute manually in fast-paced clinical settings.

Wednesday, July 29, 2026, 1:30 PM - 2:45 PM

WC01

301A

Emerging Topics in Healthcare AI

Invited Session

AI in Healthcare

Chair: Mehmet Ahsen, University of Illinois at Urbana-Champaign, Champaign, IL, United States

1 - Enhancing Fairness and Accuracy in Mammography Screening Using a Custom Ensemble

Mehmet Ahsen, University of Illinois at Urbana-Champaign, Champaign, IL, United States, Mehmet Ayvaci

As AI algorithms become increasingly prevalent in healthcare, ensuring both accuracy and fairness in decision-making poses a dual challenge. Breast cancer screening is a particularly high-stakes example: FDA-cleared AI tools are already in clinical use, and nearly 40 million mammograms are performed annually in the United States, yet many of these systems have been trained on limited subpopulations, raising equity concerns. We address this challenge by leveraging the diversity of predictive algorithms developed for the same task and forming a linear ensemble. We develop a statistical model that establishes conditions under which such an ensemble can satisfy equal-opportunity fairness and then demonstrate its application using simulated data calibrated from a real-world AI mammography competition. Our analysis shows that ensembles can improve both accuracy and fairness, especially when constituent algorithms differ in subgroup performance. These findings provide actionable guidance for health.

2 - Identifying Racial Disparities with Machine Learning

Margret Bjarnadottir, University of Maryland, College Park, MD, United States, Ritu Agarwal, Daniel Smolyak, Daichi Shimbo

Health disparities across demographic and socioeconomic groups are well-established. Disparities may be exacerbated at the intersection of multiple demographic or social dimensions, however, studying intersectional disparities has typically required ex ante selection of a small number of factors for inclusion in an interaction regression model. We utilize Causal Trees to automatically identify the demographic or social factors that, in combination, can mitigate or exacerbate racial hypertension disparities between Black and White adults, highlighting patient groups with higher or lower disparities. Using data from the All of Us Research Program we highlight the approach using hypertension as a case study. The use of causal trees offers new insights by discovering multifaceted groups of high disparities. Within these groups were adults with similar demographics and social determinants, but who nevertheless had racial disparities in the prevalence of hypertension unawareness and systolic BP. These insights can generate new hypotheses and motivate additional studies and support the targeting of policy interventions to populations where there is the most opportunity to correct for disparities in hypertension

WC02

301B

AI-facilitated Diagnosis in Healthcare

Invited Session

AI in Healthcare

Chair: Vishal Ahuja, Southern Methodist University Cox School of Business, Dallas, TX, United States

Co-Chair: Yingchao Lan, University of Nebraska-Lincoln, Lincoln, NE, United States

1 - Scalable Family Genetic Testing

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, Daniel Adelman

Hereditary genetic testing increasingly relies on multi-gene panels, but sequential family testing becomes computationally intractable as pedigree size and panel width grow. We develop an approximate dynamic programming framework for family panel testing that reformulates the sequential problem as a linear program, decomposes the value function additively by gene, and solves the resulting approximation by row generation with embedded Bayesian-network inference. This reduces the burden of exact dynamic programming while preserving a transparent accuracy–tractability tradeoff.In exact-solvable single-gene benchmarks, the proposed ADP achieves a mean policy gap of 1.85% across 33 test instances. As pedigree complexity and panel width grow, the method delivers substantial computational savings while remaining computationally usable in regimes where exact dynamic programming becomes burdensome. In large-family multi-gene benchmarks, the current best scalable specification attains an 11.72% mean policy gap, indicating that the method already makes these instances computationally tractable, while highlighting a remaining accuracy–tractability tradeoff at larger scales.

2 - Robust and Safe use of AI in Radiotherapy

Arka Roy, The University of Texas at San Antonio, San Antonio, TX, United States, Ruiqi Li, Nikos Papanikolaou

Each year, over two million people in the United States are diagnosed with cancer, and many require radiotherapy. This treatment demands precise delineation of tumors and nearby organs on CT scans so that high radiation doses reach the tumor while sparing healthy tissue. However, contouring is labor-intensive and requires specialized expertise, creating significant barriers for rural and underserved clinics with limited staffing. Artificial intelligence (AI) tools can help automate this process, but models trained on small or low-quality datasets, common in these settings, often produce errors that compromise treatment accuracy and safety. We develop new AI models that not only generate high-quality contours but also quantify their own uncertainty. We then design treatment-planning methods that explicitly incorporate these uncertainty estimates to ensure plans remain safe even when contours are imperfect. Using head-and-neck cancer cases, we evaluate our approach under varied uncertainty scenarios. Our results demonstrate that uncertainty-aware AI ad robust optimization can enhance the reliability and clinical robustness of radiotherapy, particularly in resource-constrained environments.

3 - Clustered Multi-Task Learning for Prediction of Adverse Pregnancy Outcomes

Julie Lee, Bucknell University, Lewisberg, PA, United States, Gian-Gabriel Garcia

In clinical prediction models, combining individual indicators into a composite outcome can help overcome issues associated with single outcomes such as low predictability and low prevalence, but their clinical utility may be limited. We develop an optimization framework to simultaneously cluster related outcomes and learn model parameters for each cluster. We apply our formulation to indicators comprising maternal and neonatal morbidity and demonstrate that our approach can aid interpretability by finding underlying groups of related tasks and deriving an interpretable set of predictors.

WC03

302A

Operations Research in Action: AI and Optimization in Hospital Operations

Invited Session

Healthcare Operations Management

Chair: A. Cecilia Zenteno, Massachusetts General Hospital, Boston, MA, United States

Co-Chair: Martin Copenhaver, Johns Hopkins Hospital, Baltimore, MD, United States

1 - A Causal Framework for Latent Treatments: Evaluating Supporter Communications in Digital Mental Health Services

Joyce Luo, Massachusetts Institute of Technology, Cambridge, MA, United States, Dessislava Pachamanova, Georgia Perakis

Digital mental health services have expanded access to mental health treatment and have been shown to be as effective as in-person treatment. User disengagement nevertheless remains a persistent challenge. Programs have aimed to combat this by integrating support from therapists through regular communications (text messages or emails), which has shown promise. However, there are few studies into how this support should be administered and whether optimizing these communications can further improve engagement. In collaboration with a leading provider of digital mental health solutions, we study how supporter communications can be better targeted for particular types of users to improve outcomes. We propose a novel causal inference framework that estimates heterogeneous treatment effects for high-dimensional treatments (such as text) from observational data. Our algorithm integrates causal machine learning and optimization to jointly discover latent clusters based on both patient and treatment characteristics and estimate treatment effects. We also show how large language models can be used to explain the distinct characteristics of the identified latent treatment-patient clusters, ensuring the clusters are interpretable for practitioners. We provide convergence guarantees for our algorithm and show analytical bounds on the error of the treatment effect estimates. In computational experiments, we first demonstrate strong performance on semi-synthetic data, where our approach identifies treatment clusters that are closer to the "true" clusters than benchmarks and estimates more accurate treatment effects. Then, using data provided by our collaborators, we apply our methodology to analyze and recommend personalized communications from supporters to patients that aim to improve engagement.

2 - Proactive Transfer Admission Control for Emergency Departments

Jing Dong, Columbia University, New York, NY, United States, Ruicheng Ao, Alyssa Liu, Martin Copenhaver

Emergency Department (ED) crowding remains a persistent challenge, undermining timely care, elevating clinical risk, and placing substantial strain on hospital resources. While hospitals cannot easily control the arrival of walk-in emergency patients, they often have some flexibility in managing transfer admissions from other hospitals. Temporarily restricting such transfers during periods of high congestion can provide an effective operational lever to alleviate crowding. In this work, we investigate how predictive information, namely forecasts of future demand and inpatient admission delays, can be systematically incorporated into ED admission control decisions. We develop a queueing model that captures key patient flow dynamics, accounting for a time-varying operating environment, two distinct sources of ED arrivals, and delayed inpatient admissions (i.e., ED boarding). Building on this model, we derive the optimal admission control policy under a fluid approximation. The fluid-based policy can be seamlessly adapted to the stochastic system. It explicitly leverages predictive information, offers clear interpretability and practical implementability, and achieves performance close to the stochastic optimum. Through extensive numerical experiments and a case study based on real hospital data, we show that by proactively regulating the acceptance of transfer patients based on real-time congestion levels and anticipated future conditions, the proposed approach achieves a superior efficiency frontier balancing ED congestion and transfer acceptance than reactive admission policies.

3 - Hospital Admission and Patient Transfer Control: An Offline Doubly-Robust Learning Approach

Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, United States, Esmaeil Keyvanshokooh, Arlen Dean, Mark P. Van Oyen

Hospitals frequently face congestion in high-acuity units, where admission and transfer decisions must be made under limited capacity and significant clinical consequences. We study a patient flow optimization problem focused on admission-induced intra-hospital transfer decisions across care units. In this setting, accommodating a new arrival may require transferring an existing patient to a lower-acuity unit, creating a tradeoff between capacity management, access, and potential safety risks.

We model this problem as a finite-horizon event-based Markov decision process in which decisions are made at patient arrival epochs and system evolution is driven by admissions, discharges, and bed availability. To address the practical challenge that online experimentation in hospitals is often infeasible, we develop an offline reinforcement learning framework that learns transfer policies from historical patient flow data. Our approach incorporates pessimism and constraints to account for uncertainty in under-observed state-action pairs and to discourage policies that may violate operational capacity limits.

The framework is designed to identify when it is preferable to admit directly, transfer a patient to a lower level of care in order to accommodate new demand while balancing operational efficiency and patient risk. This work contributes a data-driven decision framework for real-time hospital flow control and highlights the potential of safe offline reinforcement learning for high-stakes healthcare operations.

4 - Interventions for Patients with Complex Medical and Social Needs

Hari Balasubramanian, University of Massachusetts, Amherst, Amherst, MA, United States, Sindhoora Prakash

Patients with multiple chronic conditions and social needs represent a small percentage of the population but have a disproportionate impact on healthcare costs and utilization. Organizations around the United States have created programs, often referred to as complex care interventions, to improve the health and well-being of such patients and reduce avoidable hospital and emergency department use. In this talk, we focus on an ongoing collaboration with the Camden Coalition, a complex care organization based in New Jersey, that has tested its care models using a randomized control trial (RCT). We discuss key operational features of the the care model as well as results from a secondary analysis of the RCT.

WC04

302B

Innovations in Healthcare Delivery

Invited Session

Healthcare Operations Management

Chair: Masoud Kamalahmadi, University of Miami, Coral Gables, FL, United States

1 - Bridging the Nursing Gap: How Virtual Nurse Adoption Impacts Patient Outcomes

Blair Liu, UNC-Chapel Hill Kenan-Flagler Business School, Chapel Hill, NC, United States, Yuqian Xu, Brad Staats

The growing shortage of registered nurses (RNs) has placed unprecedented strain on U.S. hospitals, prompting healthcare systems to seek scalable and innovative staffing solutions. One such intervention is the Virtual Nurse (VN) model, which allows RNs to deliver non-clinical inpatient services—such as admissions and discharges—through synchronous video technology. Yet, despite its increasing adoption, the model's impact on patient outcomes remains largely unexplored. We leverage encounter-level data from a large East Coast healthcare system and apply a difference-in-differences (DiD) design, analyzing nearly 28,000 inpatient encounters across two hospitals, one of which implemented the VN program. This study provides the first causal empirical evidence of VN's impact on patient outcomes and offers actionable insights for optimizing virtual workforce integration in hospital operations.

2 - Hospital-at-Home Care: An Analytical Framework for Payment Design

Yinuo Li, Washington University in Saint Louis, st louis, MO, United States, Arlen Dean

Hospital-at-Home (HaH) programs deliver acute-level care in patients' residences, offering a promising alternative to traditional inpatient admission. Yet the design of reimbursement policies that incentivize high-quality HaH adoption remains poorly understood.

We develop **an analytical framework** for optimal HaH payment design, modeling the interaction between a policymaker (insurer) and hospitals as a **two-stage Stackelberg game**. In the first stage, the policymaker sets HaH reimbursement rates; in the second, hospitals choose two key operational decisions — eligibility screening

intensity and staffing intensity per HaH patient. We also incorporate **queuing theory** to capture inpatient congestion effects, modeling the hospital's bed system as an M/M/K/N queue whose load depends endogenously on HaH diversion. Readmission rates serve as the central quality measure, specified abstractly with properties linking them to both staffing and eligibility decisions: higher staffing reduces readmissions with diminishing returns, while broader eligibility screening admits marginally sicker patients and raises readmission risk.

We characterize the hospital's best-response functions and show how the HaH payment rate governs the tradeoff between volume expansion and quality investment. The framework extends to a multi-hospital setting where facilities compete for eligible patients through a logit demand model, generating a Nash equilibrium in quality. We derive comparative statics on how reimbursement generosity, competition intensity, and inpatient capacity jointly shape HaH adoption, care quality, and social welfare. Our results provide policymakers with actionable guidance on setting payment levels that balance cost savings from inpatient diversion against the quality risks of rapid HaH expansion.

3 - Telehealth's Double-Edged Sword: Hidden Costs of Expanding Virtual Access to Care

Masoud Kamalahmadi, University of Miami, Coral Gables, FL, United States, Jong Myeong Lim

Leveraging variation in telehealth adoption rates across physicians from a major health system, we show that telehealth's potential to expand access may come at the cost increased workload for physicians and revenue losses for healthcare organizations.

4 - The Impact of Additional Capacity on the Cancer Waiting Lists

Xinran Hao, Judge Business School, University of Cambridge, Cambridge, United Kingdom, Houyuan Jiang, Feryal Erhun

Cancer is the second leading cause of death worldwide, yet many cancers are highly treatable when detected early. In practice, limited diagnostic and treatment capacity, together with long waits across the pathway, undermine early detection and timely care. Motivated by these challenges, we develop a Markovian queueing network model of the cancer pathway that captures three key phases, screening, diagnosis, and treatment, and distinguishes between early- and late-stage patients. The pathway includes tandem structures with one-way complementarity between upstream and downstream capacity, and parallel early- and late-treatment queues with one-way substitution, where expanding early-treatment capacity reduces demand for late-stage treatment.

Using a fluid approximation, we derive closed-form expressions for the number of patients at each station over time and at equilibrium, and prove that the equilibria are globally asymptotically stable. These analytical characterizations reveal how arrival rates, service capacities, and waiting-related attrition jointly shape congestion across the pathway.

We then evaluate capacity expansion policies. Our analysis shows that increasing resources locally, for example at a single diagnostic or treatment station, can generate unintended consequences, including bottleneck shifting, longer waiting lists, and longer waiting times elsewhere. To avoid these effects, we derive analytical thresholds and demonstrate that proportionally allocating additional resources across multiple stations is typically more effective than isolated expansions.

Finally, we study how limited capacity should be prioritized when minimizing either waiting lists or waiting times, highlighting the roles of diagnosis departure rates, early detection rates, and station-level utilization in shaping optimal allocation.

WC05

302C

Emerging Topics in Deceased Donor Organ Transplantation

Invited Session

Public Health Policy

Chair: Diwakar Gupta, University of Texas, Austin, TX, United States

Co-Chair: Paola Martin, Indiana University, Bloomington, IN, United States

1 - The Promise of Uncertainty-Aware Ai in Reducing Deceased-Donor Kidney Non-Utilization Rates

Madeleine Pollack, Massachusetts Institute of Technology, Cambridge, MA, United States, Thomas Daillak, Korkut Uygan, Mehmet Toner, Heidi Yeh, David Cron, Swati Gupta

Deciding how and where to allocate a single deceased-donor organ among possibly hundreds of candidates on the waitlist is a complex multi-criteria decision. The United Network for Organ Sharing (UNOS) has developed a *match run:* a fair and efficient ordering of biologically compatible patients through which an available organ is sequentially offered. However, this system can be ineffective for kidneys unlikely to be quickly matched. Because organ quality deteriorates over time, there is an increased risk of nonutilization; indeed, up to 30% of kidneys offered through the match run are never transplanted. To avoid nonutilization, ad-hoc placement decisions are often made; however, without a principled procedure for determining when a kidney should be offered off-sequence and to whom, there is no systematic way to balance non-utilization risk, equitable access across transplant centers, and a fair allocation rule among patients.

Our work addresses this gap through predictive modeling and uncertainty-aware rerouting. We first develop two complementary machine learning models: (1) a global model that predicts whether a donor organ will be accepted on the match run, and (2) city-specific models that estimate whether transplant centers in that region will accept a particular organ, capturing differences in demography, resources, and clinical practice. Using training data, we then partition the donor–patient feature space into clusters based on predictive fidelity to quantify model confidence. We incorporate both model predictions and feature-specific confidence levels to develop uncertainty-aware routing policies across three large American cities, demonstrating significant reductions in estimated organ wastage.

2 - Modeling Heart Transplant Offer Acceptance with Machine Learning

Najiya Fatma, NYU Langone Health, New York, NY, United States, Jiayi Cheng, Brian Wayda, Nader Moazami, Sommer Gentry

Proposed organ allocation changes are tested using simulations that depend on models of organ offer acceptance patterns. We developed machine learning models of heart offer acceptance using the Scientific Registry of Transplant Recipients dataset (2019-2024). We constructed a longitudinal candidate dataset and linked it with potential transplant recipient and deceased donor datasets. To address the imbalance of the vast majority of “no” responses, we applied weighted penalization giving higher cost to misclassifying the accepted offers. The dataset was split 3:2:1 into training, testing, and validation cohorts before training with tree-based models, boosting ensembles, and linear classifiers. We evaluated model performance using recall, precision-recall curves, and false acceptance and false rejection percentages, defined as percentage of candidates incorrectly predicted to accept or decline an organ offer. The Shapley Additive Explanations identified center number and rank, offer number, donor-recipient distance, candidate device use (ECMO, IABP, BiVAD), months since listing, blood type, and registration at certain centers as predictors of offer acceptance. Across models, predicting accepted offers remained challenging. HistGradient Boosting (HGB) and Extra Trees (ET) outperformed Logistic regression (LR) (HGB: recall: 88%, false acceptance: 17%, ET: 85%, 19%; LR: 83%, 22%), though all models had low average precision (22% vs. 26% vs. 11%) despite high ROC-AUC. Offer acceptance models can provide insights to clinicians by identifying key candidate-, donor-, or match run-related variables to consider during the offering process. HGB and ET models can be integrated into heart allocation simulators to evaluate alternative heart allocation policies.

3 - Matching with Deteriorating Quality: the Value of Simultaneous Offers

Xin Chen, Stanford University, Stanford, CA, United States, Itai Ashlagi, Angela Kohlenberg

Motivated by applications such as deceased-donor organ allocation, we study a stochastic matching model in which strategic, impatient agents receive sequential, priority-ordered offers for heterogeneous items whose quality deteriorates over time. We show that performance is governed by the acceptance threshold -- the first position in the priority list at which an agent is willing to accept a lower-quality item -- and that match loss scales directly with this threshold. Deterioration pushes this position lower down the queue, increasing

abandonment and reducing utilization. We characterize how match loss depends on the deterioration function, priority rule, and market parameters, and derive explicit results for common decay distributions (including exponential and stepwise) and priority rules (including first-come-first-served and last-come-first-served).

We then study simultaneous offers, in which each offer round is broadcast to multiple agents and the highest-priority agent to accept is matched. We quantify the marginal impact of each simultaneous offer and show that broadcasting compresses the offer process in time, effectively lowering the acceptance threshold and substantially improving outcomes. Even small increases in the number of simultaneous offers can yield large gains in utilization and match quality, while improving fairness by enabling higher-priority agents to be matched sooner.

When simultaneous offers are costly, we characterize the optimal broadcast size that balances these costs against losses from deterioration and unmatched agents. Simulations using kidney allocation data further illustrate and support our results.

4 - Optimal Threshold Policies for Opo Out-of-Sequence Placement: An Mdp Approach

Zihan Chai, University of Texas at Austin, Austin, TX, United States, Diwakar Gupta

Out-of-Sequence (OOS) placement---whereby Organ Procurement Organizations (OPOs) bypass the match-run priority list to expedite organ placement---grew from 4.3% to 23.4% of deceased donor kidney placements between 2020 and 2024, raising concerns over equity and procedural adherence. We hypothesize that recent policy reforms---the 2021 Kidney Allocation System (KAS250) and the 2020 OPO Conditions for Coverage---have increased congestion, driving the adoption of OOS as a strategic operational response. To investigate this, we construct a novel measure of operational congestion---the peak number of concurrent match runs during the lifetime of each match run---and document a strong positive association between congestion and OOS usage across 129,807 OPTN match-run records from 56 OPOs (2015--2024). Motivated by this evidence, we model the OPO as a capacity-constrained processor managing a stochastic stream of time-sensitive placement tasks and formulate its allocation decision as a discrete-time Markov Decision Process. We prove that the optimal value function is concave in queue length and show the optimal policy follows a threshold structure: the OPO follows standard allocation below a critical congestion level and move excess arrivals one-for-one to OOS placement above it, imposing a finite effective queue capacity. The threshold is monotone in the length of the planning horizon and converges to a unique stationary policy in the limit. These results provide a structural basis for characterizing OOS as a rational response to operational frictions rather than an arbitrary deviation and offer a foundation for designing mechanisms that balance placement efficiency against equity.

WC06

303

ML and AI for Medical Decision Making

Invited Session

Medical Decision Making

Chair: Daniel Otero Leon, University of Virginia, Charlottesville, VA, United States

1 - Informing Inference on Adolescent Substance use: An Entropy-Optimized Exploratory Graph Analysis of the Abcd Dataset

Gabriel Lawrence, University of Virginia, CHARLOTTESVILLE, VA, United States

Adolescent substance use is a critical public health challenge, with 15.1% reporting past-year illicit use. This widespread use intertwines with psychiatric conditions; youth with depression or anxiety exhibit higher substance use rates. Consequently, adolescent substance use etiology is highly complex, involving interconnected socioeconomic, genetic, and neurobiological factors. The Adolescent Brain Cognitive Development (ABCD) study provides data to explore these dynamics using longitudinal substance use, demographics, mental health, and biometric data. However, extracting mathematically valid inferences from high-dimensionality datasets such as the more than 30,000 variables in the ABCD dataset poses a methodological challenge.

In this presentation, we present an analytical framework leveraging Exploratory Graph Analysis (EGA) to map the multidimensional risks of drug use indicators within the ABCD cohort. We represent complex interdependencies among socio-economic status (SES), prior drug use, and region as a Gaussian Graphical Model (GGM) using data from 11,800 children. To validate network stability, we apply an information-theoretic approach. By calculating structural entropy, we isolate the optimal topological configuration, minimizing network uncertainty to robustly verify underlying structural dimensions.

Our findings demonstrate how combining EGA with entropy-based structural verification enhances feature selection, reduces noise in high-dimensional multimodal data, and strengthens predictive inference by using network-wide entropy and universal partial correlations for all variables regarding adolescent substance use onset. Ultimately, this framework equips researchers and policymakers with a tool for mapping vulnerability pathways, facilitating targeted, data-driven preventative interventions.

2 - Characterizing Individualized Physical Activity Trajectories in Cardiac Rehabilitation

Kuofu Liu, University of Michigan, Ann Arbor, MI, United States, Amirhossein Moosavi, Mariel Lavieri, Brahmajee Nallamothu, Mike Thompson, Jessica Golbus

Cardiac rehabilitation (CR) is a supervised exercise and lifestyle program for patients with cardiovascular disease that reduces morbidity and mortality. Despite substantial heterogeneity in recovery trajectories, CR is typically delivered as a standardized 36-session program. This uniform structure can lead to inefficient resource utilization, particularly given long waitlists for CR services. To support more personalized discharge decisions, we developed a data-driven framework to characterize patient recovery trajectories using wearable-derived physical activity data. We used a discrete-time Kalman filter to estimate underlying weekly physical activity levels and to identify times at which physical activity levels reached a plateau. We fit two generalized linear mixed-effects models: a nowcast model estimating the weekly probability of entering a plateau and a forecast model predicting the probability of reaching a plateau over the subsequent 6 weeks. A total of 175 participants were included in the analysis (mean age 60.6 (IQR 15.2) years, 31% female). Fifty-five percent of participants reached a physical activity plateau prior to CR discharge after a median of 15 CR (IQR, 10–24) sessions. The nowcast model achieved an AUC of 0.776 (95% CI, 0.706–0.844) and the forecast model an AUC of 0.829 (95% CI, 0.755–0.889). Adaptive discharge policies that align CR delivery with patient-specific recovery dynamics have the potential to improve CR access though large-scale prospective validation is needed.

3 - Integrated Chronic Diseases Management for Atherosclerotic Cardiovascular Disease Prevention

Panus Sawetpiyakul, University of Washington, Seattle, WA, United States, Yixuan Long, Sun Ju Lee, Jovan Julien, Juba Ziani, Jithin Varghese, Gian-Gabriel Garcia

Chronic diseases such as Atherosclerotic Cardiovascular Disease (ASCVD) and diabetes are among the leading causes of death both in the United States and worldwide. Nonetheless, many chronic diseases are considered to be preventable through timely screening and subsequent initiation of the appropriate medication (e.g., statins for hypercholesterolemia). As such, prior research has considered optimal screening and treatment models for chronic diseases such as diabetes, chronic kidney disease, hypertension, and hyperlipidemia. However, there is a dearth of prior work considering the explicit integration of treatment decisions and their impact on optimal screening decisions. In this project, we propose a framework for optimizing the frequency of routine screening for (potentially multiple) chronic diseases with consideration of downstream treatment decisions. Specifically, we develop a finite-horizon Markov decision process (MDP) for chronic disease screening with explicit consideration of downstream treatment decisions to mitigate risk of ASCVD. In the screening phase of our model, we optimize the screening interval to maximize quality-adjusted life years (QALYs) a function of health status and treatment disutility. Concurrently, patients whose clinical measurements (e.g., blood pressure, cholesterol, HbA1c) are learned from screening will receive treatment for their chronic conditions if recommended by existing clinical guidelines. We evaluate our model with data from the US National Health and Nutrition Examination Survey (NHANES) and perform probabilistic sensitivity analysis. Our findings shed light on the value of integrating screening and treatment decisions, thereby leading to a more comprehensive approach in chronic disease management.

4 - Beyond the Clock: Improving Elective Surgery Duration Predictions with Machine Learning

Daniel Otero Leon, University of Virginia, Charlottesville, VA, United States, Constanza Lorca, Elise Williamson, Michael Dertke, Mackenzie Craig, Jacob Singer, Nathaniel Kusic, Mia deLadurantaye, Robert Riggs, Thomas Shin

Surgery is a cornerstone of modern healthcare, with the operating room (OR) as one of the most resource-intensive units within a hospital. Since surgeries account for a substantial portion of hospital revenue, accurate prediction of surgery durations and efficient OR scheduling are critical to minimizing delays for patients as well as reducing overtime for staff. However, OR scheduling is complex due to variability, including patient history, unforeseen complications, and emergency cases that can disrupt planned schedules. This complexity is compounded by the current OR scheduling approach, which remains a highly manual process driven largely by human decision-making and crude statistical averages. The goal of this project is to develop a model that improves elective surgery duration predictions by identifying patterns from historical data from the UVA Department of Surgery. Given the large number of unique surgeries, we clustered the data to gain greater insights into underlying patterns. Using this processed data, we developed XGBoost and Random Forest models that identify the most influential variables affecting surgery duration and improve time estimates for future procedures. Preliminary models show a 10.65-minute improvement in root-mean-square error for time prediction, a 14% decrease from the current prediction method.

5 - Conformal Inference for Parameter Uncertainty in Optimal Liver Acceptance Policies

Thadryan Sweeney, Dartmouth College, Hanover, NH, United States, Wesley Marrero

Conformal Inference for Parameter Uncertainty in Optimal Liver Acceptance Policies

Robust Markov Decision Processes (RMDPs) and Interval Linear Programs (ILPs) are routine for optimizing over parameter ranges. Rather than specifying a single value for a parameter, these models allow users to provide a range of possible parameter values and the model optimizes decisions accounting for this range. RMDPs produce decisions that are optimal under the worst-case reward, whereas ILPs provide a set of decision variables. These features make these models especially suited to high-stakes healthcare decisions where underlying processes may be ambiguous, unknown, or poorly characterized. A key challenge in the use of these models is establishing non-arbitrary parameter bounds. Venn-ABERS classifiers (VACs) offer and avenue to address this by generating calibrated prediction sets. Rather than single labels, VACs estimate class probabilities using counterfactual training — assigning a new observation each possible outcome and appending it to the training data for re-fitting. This reveals how each outcome affects predictions. We propose using VACs to create calibrated, statistically-motivated bounds on parameters for these models. In solving the resulting optimization problems, we present and solve an MDP reformulated as an interval linear program and demonstrate the method with an application to optimizing "treat" or "defer" decisions in the context of liver transplantation organ acceptance.

WC07

304

Bridging Analytics and Clinical Decision-Making in Dynamic Disease Management

Invited Session

Health Applications Society

Chair: Gizem Nemutlu, Brandeis University, Waltham, MA, United States

1 - Context-Aware Optimization of Follow-up Intervals for Type 2 Diabetes Care Using Markov Decision Processes

Muge Capan, University of Massachusetts Amherst, Amherst, MA, United States, Parisa Lotfibagha, William J. Gallagher, Elizabeth B. Selden, Kristen Miller

Chronic disease management relies on regular patient-provider interactions to follow-up on disease progression and control. For Type 2 Diabetes (T2D), current guidelines prescribe fixed time intervals between subsequent primary care visits for all patients, overlooking heterogeneity in clinical trajectories and patient characteristics. This study introduces a Contextual Markov Decision Process (CMDP) model to optimize subpopulation-specific follow-up interval decisions using Electronic Health Record (EHR) data from 22,154 T2D patients across 10 primary care clinics. Contexts are identified by: i) dimensionality reduction of variables representing the individual health trajectories utilizing Principal Component Analysis, and ii) assigning patients to contexts via principal components and additional patient-level features using clustering. Two distinct contexts emerged, representing a lower- and a higher-risk subpopulation. CMDP-derived policies recommend: (i) follow-up within 1 month if lab value at current visit is unmeasured; (ii) up to 3 months for elevated lab values or recent hospitalizations; and (iii) 6 to 12 months for sustained glycemic control, with shorter follow-up intervals for patients in high-risk context. The optimal policies achieved lower expected cumulative cost than benchmarks (e.g., in the higher-comorbidity context, the CMDP policy reduced cost by about 34.8\%, and in the lower-comorbidity context by about 6.4\%, relative to an American Diabetes Association-like fixed interval follow-up policy. These findings demonstrate how context-aware approaches can inform adaptive follow-up strategies, and have the potential to advance chronic care management in primary care by synthesizing machine learning and probabilistic decision models.

2 - Optimizing Treatment Allocation to Maximize the Health of a Population

Miaolan Xie, Purdue University, Edwardson School of Industrial Engineering, West Lafayette, IN, United States, Daniel Adelman, Alba Olivares-Nadal

Population Health Management (PHM) is a central area of growing focus. Optimizing PHM strategies presents several challenges: managing high-dimensional patient covariates, tracking their evolution and long-term treatment response, and accounting for population inflow and outflow. We propose an approach based on measurized Markov Decision Processes (MDPs) that integrates all of these components. We model an MDP in which a scarce treatment optimizes the long-term distribution of the patient population subject to capacity constraints. This bypasses both the dimensionality and practical challenges of tracking individual patient covariates. To solve the resulting infinite-dimensional problem, we develop an approximation that reduces the task to identifying a finite set of high-performing treated and untreated patients. Despite this complexity, our approach yields a simple, clinically implementable index policy: treat a patient if their adjusted impactability exceeds a threshold. The adjusted impactability captures the long-term consequences of receiving, or not receiving, treatment. While straightforward to apply, the policy remains flexible and can incorporate general machine learning models. Using CMS data and machine learning to estimate population dynamics, we demonstrate statistically significant improvement over a myopic benchmark. This advantage grows with the time horizon, consistent with our policy's forward-looking nature. At the longest horizon tested, this corresponds to over 1,500 additional home days annually per 1,000 patients.

3 - Improving Cystic Fibrosis Screening Through a Novel Testing Design

Hussein El Hajj, Santa Clara University, Santa Clara, CA, United States, Douglas Bish, Ebru Bish

Newborn screening (NBS) plays a key role in detecting life-threatening genetic disorders, with cystic fibrosis (CF) being a prominent example. With over 1,000 CF-causing variant types, a genetic test that can detect enough variant types for sufficient screening accuracy is expensive and capacity-constrained. Consequently, in the US, the first tier of testing uses an inexpensive biochemical test that measures immunoreactive trypsinogen (IRT) levels, which are often elevated in CF cases. Genetic testing is reserved for a small fraction of newborns with elevated IRT levels, based on an IRT threshold. However, setting an optimal threshold presents challenges, particularly due to variations in IRT levels and CF prevalence among different racial and ethnic groups. As a result, the current practice of using a single IRT threshold for all newborns is not necessarily the most efficient or equitable strategy.

We introduce a novel screening design, integrating an inexpensive small-panel genetic test (DNA) before the IRT test. This strategy customizes IRT thresholds based on the number of variants detected by the small-panel DNA. This design leads to novel optimization problems, including variant type selection for the small-panel

DNA and IRT threshold customization. We establish key structural properties of optimal IRT thresholds and develop Pareto frontiers that inform decision-makers about the efficiency versus equity trade-off, considering a minimax type equity measure. The case study, utilizing CF NBS data from the states of New York and California, demonstrates the potential efficiency and equity gains of this novel approach compared to current practice and race-based benchmarks.

4 - Diagnosis Patterns and Disparities in Adhd Among Children with and Without Down Syndrome: Evidence from Nsch 2016 & 2023

Gizem Nemutlu, Brandeis University, Waltham, MA, United States

Attention-deficit/hyperactivity disorder (ADHD) is one of the most prevalent neurodevelopmental conditions in children, yet diagnosis is shaped by both clinical complexity and access to care. Children with Down syndrome (DS) may be particularly vulnerable to underdiagnosis due to overlapping behavioral phenotypes and competing clinical priorities.

We use nationally representative data from the National Survey of Children's Health (NSCH), 2016–2023, to examine ADHD diagnosis patterns among children with and without DS. Using survey-weighted estimators, we construct cohorts based on parent-reported diagnoses and estimate ADHD prevalence across groups. We then model the likelihood of ADHD diagnosis as a function of demographic and socioeconomic factors using multivariable logistic regression, accounting for the complex survey design.

Preliminary results indicate that children with DS are less likely to have a reported ADHD diagnosis compared to children without DS, conditional on behavioral indicators, suggesting potential under-recognition. Across both populations, diagnosis likelihood varies systematically with insurance status, household income, parental education, and race/ethnicity. In particular, publicly insured, lower-income, and minority children exhibit lower odds of diagnosis, consistent with disparities in access to evaluation and care. Interaction analyses suggest that these disparities may be amplified within the DS subgroup.

These findings highlight gaps in diagnostic pathways for ADHD, especially among clinically complex populations. From a health systems perspective, the results point to missed opportunities for early identification and intervention, and motivate the need for more consistent screening and equitable access to diagnostic services. Ongoing work will integrate clinical datasets to validate findings and inform targeted screening strategies.

5 - An Interpretable, Robust, Risk-Averse Reinforcement Learning Framework for Clinical Decision-Making Applied to Sepsis Treatment

Angela Lin, MIT ORC, Cambridge, MA, United States, Dessislava Pachamanova, Georgia Perakis

Every hour of every day, clinicians make time-sensitive, potentially life-or-death treatment decisions for critically ill patients. This work addresses decisions faced by clinicians treating patients with sepsis, a dysregulated response to infection that is the third leading cause of death in the US. Patients' responses to treatments are highly heterogeneous, and some of the resources needed to treat severe sepsis patients, such as ventilators, dialysis machines, and specialized personnel, are limited in quantity and therefore require difficult decisions to be made on how to allocate these resources. In this work, we formulate a joint treatment and resource allocation problem: we use an algorithm that learns a concise set of states for a Markov Decision Process (MDP) model of patient dynamics; propose risk-averse and robust solution methods that consider worst-case outcomes when solving the weakly coupled MDP under shared resource constraints; and provide interpretable, decision tree-based rules for explaining the recommendations for each patient. Our simulations show that our proposed framework is robust to variations in resource availability, achieving the same number of patient discharges as observed in the data while using only 40% of the resources. To facilitate implementation, we have been working closely with a team of clinicians at Tampa General Hospital to co-design a user interface that integrates recommendations and supporting information in a format conducive to trust, transparency, and adoption by clinicians.

WC08

305A

Improvement, Systems, and Equity: Quality Healthcare for All

Invited Session

Healthcare Quality and Safety

Chair: Kimberly Harry

1 - Employee-Experienced Culture as a Leading Indicator of Patient Experience

Sarah Bonzo, State University of New York at Oswego, Honeoye Falls, NY, United States

Patient experience performance remains uneven and difficult to sustain despite extensive investment in training, dashboards, and incentives. Because patient experience metrics are retrospective, leaders often lack timely visibility into the organizational conditions shaping those outcomes.

We assessed whether employee-expressed workplace culture, captured in publicly available online reviews, aligns with patient experience and quality benchmarks. Linking 2017–2019 Glassdoor reviews with HCAHPS and CMS Overall Hospital Quality Star Ratings for 337 U.S. hospitals, we examined how multiple employee-experience dimensions relate to patient experience and quality performance. Culture & Values was the most consistent correlate across outcomes and outperformed compensation, leadership approval, and business outlook in adjusted models.

Narrative patterns reinforced the quantitative findings: lower-culture hospitals were described using terms such as blame, hostility, and poor treatment, while higher-culture hospitals were characterized by language reflecting inclusion, supportive leadership, and pride. Together, these results suggest that culture is a measurable, externally visible signal of the conditions under which patient-centered care is delivered, and may provide earlier visibility into patient experience risk than patient surveys alone.

2 - A Systems-Oriented Data Analytics Framework for Improving Healthcare Quality and Patient Safety

Omar Faruq Osama, SUNY at Binghamton, Binghamton, NY, United States, Kimberly Harry

In contemporary healthcare systems, improving treatment quality and patient safety remains a major challenge due to increasing care complexity, rising service demand, fragmented workflows, and expanding digital health infrastructures. Although electronic health records and related technologies generate large volumes of clinical, operational, and administrative data, many healthcare organizations still struggle to convert these data into timely and actionable insights for quality improvement.

This presentation proposes a systems-oriented, equity-aware data analytics framework for improving healthcare quality and patient safety. Unlike traditional quality improvement approaches that rely heavily on retrospective reporting, the proposed framework emphasizes proactive and continuous learning by integrating predictive analytics, AI-enabled monitoring, and decision-support tools across clinical and operational settings. By combining multiple data sources, the framework supports earlier identification of patient risk, workflow bottlenecks, and potential safety events while helping organizations design more timely and targeted interventions.

The novelty of this framework lies in its ability to connect data analytics, systems thinking, and quality improvement in a closed-loop model of healthcare improvement. Rather than examining safety, efficiency, and equity as isolated outcomes, the framework treats them as interdependent results shaped by workflows, information systems, organizational routines, and human decision-making. This systems-oriented perspective contributes to healthcare operations research by offering a more proactive, coordinated, and scalable approach to strengthening patient safety, improving care quality, and supporting equitable healthcare delivery.

3 - Prospective Prediction of Length of Stay at Admission for Children with Complex Chronic Conditions

Ajit Appari, Northeastern University, Boston, MA, United States

Prevalence of complex chronic conditions [CCCs] (coexistence of chronic conditions affecting three or more body-systems) among children has risen to nearly 30% causing unprecedented growth in hospital care expenditure. Despite the rich literature on hospitalization-related outcomes for adults with CCCs, research is sparse for adolescent population.

In this study, I implement machine learning classifiers for predicting count of in-hospital complications, and length of stay (LOS) for adolescents with multimorbidity (patient age: 10-19 years) using 17 quarterly releases of de-identified all-payers inpatient claims data for Texas during Oct2015-Dec2019. The analytic sample includes 281,472 hospitalizations for adolescents with CCCs across 400+ hospitals.

I adopted multi-stage modeling process. First, patients were segmented into multiple CCC patterns using unsupervised learning by considering 150+ clinically meaningful chronic condition categories. Next, for each CCC pattern, multi-class Random Forest Classifiers are established to predict in-hospital complications (categories: none, 1-2; 3-5; 6-10; >10). Finally, for each CCC pattern, I evaluate multiple machine learning models (multinomial logistic regression, Random Forest, and Support Vector Machine) to predict LOS [three categories: Short LoS (<3days), Moderate LoS (=3-5days), Prolonged LoS (>5days)] conditional on the source/type of admission, major diagnostic categories, estimated in-hospital complications risk profile, co-occurring CCC patterns burden, demographic factors (age, race/ethnicity, insurance), and neighborhood structural disparity factors measured at zip-code tabulation areas (e.g., socioeconomic disadvantage, affluence, and ethnic/immigrant concentration), and provider factors.

WC09

305B

*CANCELED Improving Case Management and Medical Decision Making

Contributed Session

Medical Decision Making

Chair: david wierz, the oci group, Malvern, PA, United States

1 - *CANCLED Enhancing Outcomes - Operationalizing Value-Based Care - Integrating Home-Based Delivery

david wierz, the oci group, Exton, PA, United States

Value-based care and Home-based care each support practices that enhance outcomes. Those enhancements encompass - 1) Clinical, 2) Financial, 3) Operational, 4) Quality of Care and 5) Quality of Life metrics. Results accrue to the patient and family together with clinical personnel and payers. Integrating home-based care into structures for value-based care creates a means for offering care integrated across a continuum. That is unifying care for the patient as a person thru delivery, location and time. Effect provides opportunities to refine and align incentives for care in common addressing all outcome metrics. The refinement and alignment further supports care that acknowledges the interests of the patient and family together with the care team. There is additional opportunity to integrate social needs - social determinants of health (SDOH) - into care delivery creating a unified web supporting overall health of the person while a patient.

Presentation uses case studies in three areas - oncology, orthopedics, diabetes - to highlight pathways for integrated values-based and home-based care into an adaptive, dynamic and supportive system of delivery. Focus in on practical aspects of care, integration and outcomes - what works - what doesn't. Development considers the organizational, operational and workflow elements with financial and clinical requirements - and - means to use technology in creating leverage across the aspects of care.

Discussion illustrates the use of information and technology in highlighting means to build communication and engagement between the patient, family and care team. Analytics - simple and advanced - are enablers to optimize wholistic performance across metrics

WC10

306A

Explainable AI and Causal Inference in Healthcare

Contributed Session

Data Analytics

Chair: Keyu Deng, 430 E Walnut, 65806

1 - Drug Shortage and Drug Acquisition Cost Increase: Heterogeneous Evidence from Fixed Effects Model and Casual ML in the U.S.

Qiaoyuan Li, University of California, Berkeley, Berkeley, CA, United States, Geetha Sreenivasa Rao Repalle

Drug shortages are widely believed to increase drug costs, yet average estimates may obscure substantial variation across shortage episodes. This study examines the causal impact of drug shortages on U.S. generic drug acquisition costs and identifies the conditions under which shortages meaningfully affect prices.

We link weekly National Average Drug Acquisition Cost data with the FDA drug shortage list and construct a quasi-experimental framework comparing shortage and non-shortage periods within the same product. To address selection bias and limited overlap between affected and unaffected drugs, we apply propensity-score matching and estimate effects using two-way fixed effects models alongside Double Machine Learning to characterize heterogeneous treatment effects.

Results show that the average drug shortage effect on drug acquisition costs is small at approximately 2.4% in log price terms and statistically insignificant. These results suggest that drug shortages do not uniformly raise prices across products and episodes. However, substantial heterogeneity emerges across episodes. Short-duration shortages exhibit little measurable impact, whereas shortages lasting longer than four weeks increase acquisition costs by approximately 3.6% relative to non-shortage periods. Feature importance analysis indicates that more than 99% of heterogeneity is explained by shortage duration measurements, suggesting that persistence rather than incidence drives price escalation.

These findings highlight drug shortage duration as an important factor for policymakers seeking to manage pharmaceutical costs. The proposed framework provides healthcare policymakers and supply-chain decision makers with a data-driven approach to anticipate and mitigate financial risks arising from prolonged supply disruptions.

2 - From Regulatory Text to Shortage Risk: An Nlp and LLM Framework for Pharmaceutical Quality Risk Prediction

Amirreza Sahebi, North Carolina State University, Raleigh, NC, United States, Shailesh Divey, Amir Sadeghi, Robert Handfield, John Gray

Drug shortages driven by manufacturing quality failures impose severe costs on patients, hospitals, and supply chains. Yet the rich unstructured regulatory text that FDA generates like inspection observations, warning letters, recall narratives, and import refusal records, remains largely untapped as a predictive resource. This paper presents a multi-source regulatory intelligence framework that integrates five FDA databases covering 129 pharmaceutical manufacturing facilities across 14 high-priority generic Active Pharmaceutical Ingredients (APIs).

We construct a facility-level feature matrix that integrates structured regulatory outcomes with NLP-derived text signals, enabling systematic validation and prediction of downstream supply disruptions. An LLM pipeline further extracts structured features from Form 483 PDFs, capturing violation types, escalation patterns, and latent firm behavior through the lens of Agency Theory and Remediation Economics. We then validate these features against independent ground-truth measures, including analytical drug quality testing and pharmacy claims data capturing real-world dispensing disruptions.

Preliminary results reveal actionable early-warning signals: specific regulatory citation patterns predict the most severe inspection outcomes (Pearson correlation = 0.51), contamination and manufacturing practice failures account for 94% of drug recalls linked to our study compounds, and import shipments rejected after laboratory testing concentrate in a small set of high-risk facilities. These signals can be detected month before shortages materialize, enabling proactive supply chain risk management.

The resulting index provides a tool for proactive supplier risk stratification and informing sourcing strategies in pharmaceutical supply chains. To our knowledge, this is the first study linking FDA regulatory inspection text to pharmacy claims data for supply disruption risk.

3 - Regulatory Trajectories as Signals of Nursing Home Quality Deterioration

Keyu Deng, Missouri State University, Springfield, MO, United States, Zhiguo Yang, Weiyi Duan, Xiang Guo

Regulatory penalties in U.S. nursing homes, including monetary fines and payment denials, signal serious quality and compliance failures and highlight the need for early-warning models that support proactive regulatory oversight. Most existing prediction approaches rely on static facility characteristics or aggregated historical summaries, which may not fully capture how regulatory signals accumulate and evolve as facility quality changes over time. This study examines whether modeling regulatory trajectories provides a more informative representation of facility quality dynamics.

Using publicly available data from the Centers for Medicare & Medicaid Services, including Health Deficiencies, Fire Safety Deficiencies, Penalties, and provider-level facility characteristics, we construct facility–time samples and predict whether a nursing home will receive a monetary fine within the next 180 days. As benchmarks, we develop aggregated tabular baselines using historical summaries from 365-day and 90-day observation windows together with contemporaneous facility covariates. Under a strict chronological split, the strongest baseline (CatBoost) achieves a validation AUPRC of 0.314 and a test AUPRC of 0.259.

We then introduce a hierarchical regulatory trajectory representation, grouping deficiency events into inspection episodes and ordering them over time to form facility-level regulatory histories. These trajectories are encoded using a transformer-based sequence model and combined with facility covariates for prediction. The resulting model improves performance to a validation AUPRC of 0.364 and a test AUPRC of 0.323.

These findings suggest that regulatory trajectories provide richer signals of evolving facility quality than flat historical summaries and improve early detection of enforcement risk.

WC11

306B

Disparities to Decisions: Tools for Action and Training for Research in Health Disparities

Panel Session

Minority Issues Forum (MIF)

Co-Chair: Karen Hicklin, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY, 10029, United States

1 - Moderator Panelist

Karen Hicklin, Icahn School of Medicine at Mount Sinai, New York, NY, United States

Despite substantial advances in documenting health disparities, significant challenges remain in translating evidence into actionable decisions and preparing researchers to conduct rigorous, impactful disparities-focused work. This panel brings together scholars and practitioners to discuss analytic tools and decision-support approaches that move disparities research beyond description toward intervention design, implementation, and evaluation. Panelists will highlight methods such as decision modeling, optimization, implementation science,

and applied analytics that inform when, where, and for whom interventions should be deployed under real-world constraints.

In addition, the panel will address the training and mentorship needed to sustain high-quality health disparities research in a changing research climate. Discussion topics will include developing methodological expertise, operationalizing disparities measures, navigating data and funding constraints, and supporting early-career investigators and trainees. Through an interactive discussion, the panel aims to identify strategies for translating disparities research into action while building a robust and prepared research workforce capable of advancing disparities reduction through rigorous, decision-focused science.

2 - Panelist

Julie Swann, NC State University, Raleigh, NC, United States

3 - Panelist

Chaitra Gopalappa, University of Massachusetts, Amherst, Amherst, MA, United States

4 - Panelist

Jennifer Lobo, University of Virginia, Charlottesville, VA, United States

5 - Panelist

Julie Ivy, University of Michigan, Ann Arbor, MI, United States

WC12

306C

Emerging Challenges in Healthcare Operations Management

Invited Session

MSOM: Healthcare

Chair: Yu Ma, University of Wisconsin Madison, Madison, WI, United States

1 - Online Learning with Survival Data

Arielle Anderer, Cornell University, Johnson School, Ithaca, NY, United States, Hamsa Bastani, John Silberholz

Decision-makers frequently utilize adaptive experiments to optimize time-to-event outcomes, such as accelerating healthcare screenings or delaying customer churn. Traditional multi-armed bandit algorithms fail in these settings because they assume outcome delays are non-informative, leading practitioners to rely on dichotomization, a heuristic that collapses continuous timing into binary outcomes at a fixed threshold. We introduce "survival bandits," a principled class of algorithms that integrate the Cox proportional hazards model to utilize the full temporal signal of every event. We analytically contrast the regret of both approaches under a large-n limit with Weibull-distributed survival times satisfying the proportional hazards assumption. Our analytical results prove that dichotomization imposes substantial inefficiency, increasing regret by 41% to 54% even under optimal parameterization. We further demonstrate that survival bandits are uniquely robust to event rate uncertainty, whereas dichotomized approaches suffer significant performance degradation due to fragile threshold dependencies. Even when the proportional hazards assumption is violated, we show that survival bandits effectively identify the best arm under realistic scenarios. Simulations using real-world cervical cancer screening data validate our findings, demonstrating that survival bandits consistently reduce regret relative to the best-performing dichotomized algorithms.

2 - From Implementation to Impact: Human-Algorithm Collaboration in Hospital Bed Placement

Arlen Dean, Washington University in St. Louis, St. Louis, MO, United States, Mohammad Zhalechian, Iman Attari, Mark Van Oyen

We present our work with a large hospital to improve coordination among bed managers for patient-to-bed placements. Hospital bed managers are responsible for managing patient flow, a critical aspect of hospital operations and patient care. Allocating beds can be complex and time-consuming due to the shared resource nature of beds and the varying care needs of patients. To address these challenges, we developed and

implemented an algorithm based on an optimization model that uses operational practices and real-time patient and system information. We examine the effects of our algorithm on our partner hospital's operations post-implementation using an empirical analysis based on several key outcome metrics related to care quality.

3 - When AI Listens First: Crisis Triage and Emotional Support in Mental Health Helplines

Xiaoquan Gao, Singapore Management University, Singapore, Singapore, Yanhan (Savannah) Tang

We study the operational impacts of large language model (LLM) adoption in Singapore's national mental health crisis helpline. The system allows an LLM to engage with callers first, providing triage of suicidal risk, detection of non-genuine calls, and conversational support for calming caller emotions before escalation to a human counselor. We develop a queueing model to captures the joint dynamics of AI support and human counseling under heterogeneous caller severity levels. Our work investigates how LLM performance in terms of triage accuracy and emotion support affect the service quality measured by high-risk abandonment, staff workload, and caller experience (e.g., wait time, care delivery). Our analysis identifies the operational conditions under which AI can safely “listen first,” guiding the design of triage thresholds, staffing policies, and human-AI coordination rules in mental health crisis systems.

4 - Beating the Winner's Curse via Inference-Aware Policy Optimization

Bryce Mclaughlin, University of Pennsylvania, Philadelphia, PA, United States, Hamsa Bastani, Osbert Bastani

There has been a surge of recent interest in automatically learning policies to target treatment decisions based on rich individual covariates. In addition, practitioners want confidence that the learned policy has better performance than the incumbent policy according to downstream policy evaluation. However, due to the winner's curse---an issue where the policy optimization procedure exploits prediction errors rather than finding actual improvements---predicted performance improvements are often not substantiated by downstream policy evaluation. To address this challenge, we propose a novel strategy called inference-aware policy optimization, which modifies policy optimization to account for how the policy will be evaluated downstream. Specifically, it optimizes not only for the estimated objective value, but also for the chances that the estimate of the policy's improvement passes a significance test during downstream policy evaluation. We mathematically characterize the Pareto frontier of policies according to the tradeoff of these two goals. Based on our characterization, we design a policy optimization algorithm that estimates the Pareto frontier using machine learning models; then, the decision-maker can select the policy that optimizes their desired tradeoff, after which policy evaluation can be performed on the test set as usual. Finally, we perform simulations to illustrate the effectiveness of our methodology.

5 - Decision-Relevant Uncertainty: Auditing LLMs for Sequential Ivf Treatment Planning

Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States, Agni Orfanoudaki

As AI-assisted prognostic tools increasingly inform patient decisions, their calibration and reliability become a first-order clinical concern. Patients now routinely rely on LLMs to reason through treatment options, yet the uncertainty in these predictions is rarely surfaced, and their downstream implications for decision quality are not well-captured in existing models. In vitro fertilization (IVF) serves as an exemplar: it presents a challenging sequential decision setting where a patient faces a sequence of costly, high-stakes choices governed by a heterogeneous reward function trading off time to pregnancy, financial cost, and clinical risk.

We present an empirical study of LLM calibration for IVF prognosis across a structured prompt space designed to isolate variation axes that are both decision-relevant and those on which estimates should be invariant. For example, embryo yield and per-transfer live birth probability should vary with clinical features yet remain invariant to patient demeanor. The heterogeneity of patient interactions with such systems poses a distinct evaluation challenge, which we address through three uncertainty signals: prompt sensitivity, spontaneous uncertainty expression, and self-reported interval width. We connect these to downstream decision quality in a sequential IVF planning framework, and introduce a general framework for assessing decision-relevance and input sensitivity in optimization models with LLM-derived inputs.

WC13

307

Healthcare Operations Analytics and Decision Support

Invited Session

Location Analysis

Chair: Ali Dogru, University of Southern Mississippi, Hattiesburg, United States

1 - Role of Patient Attendance Characteristics and Service Policies in Scheduling Telehealth Services

Ibrahim Capar, University of North Carolina Wilmington, Wilmington, NC, United States, Ali Dogru, Ozgur Araz

Motivated by the growing adoption of omni-channel healthcare services, we address the stochastic appointment scheduling problem in an open-access clinic offering a hybrid service: regular office visits and virtual consultations. Patients are unpunctual and heterogeneous regarding their no-show probabilities and service times. The objective is to determine an intra-day appointment scheduling template used to schedule in-office and virtual appointments while anticipating same-day appointment requests and service preferences of patients, minimizing the weighted average of waiting time for patients, physician idle time, and overtime. We propose two solution approaches: a modified backtracking depth-first search algorithm, which systematically explores tractable instances, and a genetic algorithm, which efficiently solves larger problem instances. Both algorithms rely on sample average approximation for objective function evaluation. In collaboration with hospital administrators from a not-for-profit healthcare organization, we conducted a computational study to evaluate how schedule characteristics, patient attendance, and clinic preferences affect performance. We find that sequencing high-variability appointments at the end may underperform with uneven capacity. Operational preferences often have a greater impact than patient behavior. Clinics can boost efficiency by allocating more follow-ups to telemedicine and improve access by adopting open scheduling when same-day demand stays below a critical threshold.

2 - Predicting Appointment Disruptions in Specialty Care: A Case for Neurological Services

Pedro Gomes, University of Nebraska-Lincoln, Lincoln, NE, United States, Yanxin Liu, Ozgur Araz

Missed and overran appointments reduce clinic productivity and disrupt capacity planning, particularly as telemedicine becomes integrated into specialty care. We analyzed 55,078 scheduled appointments from 20,634 patients across multi-site neurology clinics between 2022 and 2024 to examine differences in no-show behavior across appointment modalities. The overall no-show rate is 0.1017%, with in-person visits exhibiting a 0.0545% rate and telehealth visits a 0.0472% rate.

We develop a predictive framework to estimate patient-level no-show and visit overrun probabilities as a function of scheduling and patient characteristics. Covariates include appointment type, patient age and its interaction with appointment type, rurality (RUCA), neighborhood deprivation index (ADI), sex, and insurance status, while controlling for time and physician fixed effects.

Our approach allows probabilistic estimation of heterogeneous no-show and visits overrun risk while accounting for physician-level and temporal variation. By quantifying how telemedicine interacts with patient characteristics to affect attendance behavior, the model provides relevant decision inputs for overbooking, reminder targeting, and capacity allocation for specific modalities.

This study contributes to healthcare operations analytics by offering a predictive framework to support data-driven scheduling decisions in specialty telehealth settings. The results have implications for managing appointment type mix, improving resource utilization, and designing access strategies as telemedicine adoption continues to expand.

3 - Data-Driven Optimization of Hybrid Appointment Systems with Telemedicine Consultations

Emirhan Uzunlar, University of Nebraska-Lincoln, Lincoln, NE, United States, Ozgur Araz, Ali Dogru, Ibrahim Capar

Problem definition: This study addresses appointment scheduling informed by a predictive model for a psychiatry clinic operating under a hybrid service model, offering both in-person and telepsychiatry visits. The clinic schedules various appointment types: an initial intake session followed by multiple follow-up visits whose number and duration depend on the patient's diagnosed condition at intake.

Methodology/results: We propose an integrated predictive-prescriptive analytics framework. In the predictive phase, multiple machine learning classifiers are trained on the 2021–2022 Medical Expenditure Panel Survey (MEPS) dataset to estimate patient-specific probabilities of preferring telepsychiatry appointments using demographics, health conditions, and insurance variables. These predictions are incorporated into a simulation optimization model capturing both intra-day and inter-day scheduling dynamics. The model accounts for appointment mode- and type-specific stochastic service durations, mode-specific patient no-shows and unpunctuality, random same-day appointment requests, and rescheduling. Scheduling templates are evaluated using sample average approximation over a one-year horizon. The objective function minimizes a weighted composite metric balancing indirect and direct waiting times for in-person and telepsychiatry patients, physician idle time, and overtime, normalized by patient throughput. A case study using data from a psychiatry clinic demonstrates the framework's practical applicability.

Managerial implications: Relying only on a weighted-average objective can produce excessive indirect waiting times for certain appointment types and service modes, leading to access imbalances. Operational adjustments such as introducing 15-minute follow-ups alongside 30-minute sessions, refining appointment slots, incorporating patient-specific telemedicine preferences, and optimizing schedules over a long-term horizon can expand scheduling capacity by about 40% without increasing psychiatrist work hours.

Wednesday, July 29, 2026, 3:15 PM - 4:30 PM

WD01

301A

Learning and Optimization for Decision-Making in Healthcare

Invited Session

AI in Healthcare

Chair: Holly Wiberg, Carnegie Mellon University, Pittsburgh, PA, United States

1 - Making Multiclass Prediction Actionable to Reduce Missed Appointments via Optimization-Based Classification

Yifang Yan, The Pennsylvania State University, University Park, PA, United States, Qiushi Chen

Missed primary care appointments disrupt continuity of care, leading to poorer health outcomes, higher healthcare costs, and wasted clinical resources. A substantial literature has focused on predicting the missed appointments to identify influencing factors and guide outreach interventions. However, most studies focus on binary outcomes and rarely distinguish late cancellations from no-shows, despite their different implications for workflow and intervention design. For multiclass prediction problems, results are typically presented as predicted probabilities, which can be difficult to translate directly into action. To bridge these gaps, we present a novel classification with inconclusiveness framework that assigns multiclass predicted probabilities into a deterministic class while also identifying cases with low prediction confidence (i.e., inconclusive cases). We formulated optimization problem with a mixed integer programming model, and developed a tailored spatial branch-and-bound algorithm by exploiting the problem structure to solve the large-scale optimization problem efficiently. We demonstrated our approach in an application for primary care appointment scheduling among adult patients with type 2 diabetes, based on real-world health records from 15 family medicine clinics at a regional academic health center in central Pennsylvania. In a simulated, capacity-constrained patient navigation center, we compared our rule with strategies prioritizing no-show risk alone or potential health impact alone, providing a transparent bridge from prediction to scheduling decisions. Under the same navigation-center capacity and moderate-pressure scheduling regime, our rule yielded more completed appointments, fewer no-shows and late cancellations, and modestly better long-term HbA1c control.

2 - Scaling Distributed Healthcare: Principled Predict-Then-Optimize via Discrete Optimization

Hao Hao, Carnegie Mellon University, Pittsburgh, PA, United States, Peiyu Wang, Holly Wiberg, Peter Zhang

As health systems become increasingly geographically distributed, they face the challenge of delivering hospital-level care efficiently to a dispersed patient population. Hospital-at-home (HaH) offers a promising alternative by allowing eligible patients to receive acute care services at home. Clinical evidence suggests that HaH can improve patient outcomes and satisfaction. Operationally, HaH can relieve pressure on in-hospital bed capacity. However, effectively scaling HaH poses major operational challenges. The key trade-off is to balance the benefit of additional service capacity and patient demand with the increased service cost and demand uncertainty. We address: Holding service-level fixed, when is HaH cost-effective relative to traditional in-hospital care, and under what medical and logistical conditions? We provide an integrated learning-and-optimization framework to jointly optimize patient selection and supply chain coordination while hedging against demand uncertainty. First, hospitalized patients are assigned either to remain in-hospital or enter HaH care. HaH patients are then assigned to service depots, where demand is pooled and operational decisions are coordinated. On the learning problem, we generate patient-level demand predictions and construct demand uncertainty sets via conformal prediction. These uncertainty sets are input to a robust optimization model that determines staffing, procurement, inventory positioning, and transportation policies. We further show that the overall learning-and-optimization problem can be formulated as a bilevel discrete optimization model with a submodular structure, susceptible to efficient approximation algorithms. Finally, numerical experiments calibrated on MIMIC-IV data identify when HaH is cost-effective, how demand pooling drives savings, and how uncertainty impacts selection decisions.

3 - Dynamic Overflow Capacity Management for Emergency Department Boarding

David Kaufman, University of Michigan - Dearborn, Dearborn, MI, United States, Jairo Martínez-Correa, Gabriel Zayas-Caban

Emergency department (ED) boarding—when admitted patients remain in the ED due to insufficient inpatient capacity—continues to drive ED overcrowding and degrade care delivery, underscoring the need for coordinated, system-level interventions. In this talk, we address the problem of dynamically adjusting buffer capacity after phase-one service in a two-stage tandem queueing model. The goal is to temporarily alleviate congestion in the first stage while minimizing total holding costs from jobs waiting or receiving service across both stages. We model this scenario using a finite-horizon continuous-time Markov decision process, whose optimality equations yield differential equations—distinct from the algebraic/recursive equations of infinite-horizon problems. After establishing the existence of a solution to these equations, we implement a time-discretization approach to approximate the optimal value function and corresponding policy. Numerical examples illustrate the structure of the discretized policy.

WD02

301B

Cost Effectiveness and Financial Concerns in Healthcare

Contributed Session

Health Economics and Cost Effectiveness

Chair: Onkar Malgonde, North Carolina State University

1 - Using GPS to Navigate the Connection Between Physician Workload and Medical Malpractice Risk

Denise Bergdolt, University of Pittsburgh, Pittsburgh, PA, United States, Prakash Mirchandani

Medical malpractice—professional negligence by a practitioner resulting in patient harm—is a critical issue in the U.S. healthcare system, affecting thousands of stakeholders and carrying a financial impact of billions of dollars annually. We construct an extensive, national, retrospective database to examine how physician workload

intensity impacts paid medical malpractice claim intensity at a state-year level. To investigate the relationship, we implement a generalized propensity score stratification method that is suited for continuous treatment levels (in our case, workload intensities). We find an inverted-U relationship between the natural log of the workload intensity and paid medical malpractice claim intensity. We attribute the concave pattern to a shift in the relative dominance of the work stress and defensive medicine effects. As workload rises, physicians may experience increased stress, mental workload and burnout, all of which negatively correlate with medical job performance in the literature. Meanwhile, a different stream of literature illustrates that if physicians do not have sufficient time to examine their patients, they are more likely to use defensive medicine, which leads to increased treatment documentation, ultimately decreasing the odds of medical malpractice conviction in court settings. We employ imaging utilization as a proxy for defensive medicine to confirm its mediating effect between workload and paid medical malpractice claim intensity, and investigating the moderating effect of the tort-law environment in different states. We also conduct several robustness checks to check for the validity of our main findings.

2 - Evaluating the Impact of the Swiss Hospital Capacity Planning Reform on Hospital Costs: Evidence from a Staggered Difference-in-Differences Analysis

Daria Bukanova-Berend, School of Medicine, University of St. Gallen, St. Gallen, Switzerland, Charlotte Lang, Justus Vogel, David Ehlig, Alexander Geissler

Objectives: This study evaluates the impact of Switzerland's 2012 hospital capacity planning reform on the economic efficiency of inpatient care. The reform introduced service group–specific mandates (SPLGs), allowing hospitals to provide selected services if they meet defined quality standards. By consolidating service provision, the policy aimed to achieve economies of scale and reduce rising hospital expenditures. We assess whether the reform lowered inpatient costs per case overall and across service groups.

Methods: We analyze nationwide inpatient data from the Swiss Federal Statistical Office (2010–2023), covering over 10 million cases. A staggered Difference-in-Differences design exploits variation in SPLG implementation timing across cantons. Costs per case are adjusted using the case mix index for comparability. We estimate separate models for 30 service groups and use event studies to assess dynamic effects and test parallel trends.

Results: The reform led to an average 11.4% reduction in cost per case per 100,000 inhabitants. Event studies show increasing cost reductions in the first three years post-implementation, with no evidence of pre-trends. Effects vary across service groups: 11 SPLGs (e.g., cardiovascular, endocrinology) show significant cost decreases, while rheumatology shows an increase and others remain unaffected.

Discussion: The findings provide early evidence on long-term effects of Swiss hospital planning reform. While cost reductions support service concentration, heterogeneous effects highlight the role of clinical and regional factors.

3 - Opening the Black Box of Hospital M&A: How Post-Acquisition It Integration Shapes Service Expansion and Consolidation at Target Hospitals

Onkar Malgonde, North Carolina State University, Raleigh, NC, United States, Huseyin Tanriverdi, Xiaoxuan Yang, Sarah Lalani

Mergers and acquisitions (M&A) in the U.S. hospital industry have accelerated in recent years, often positioning large multihospital systems (MHSs) as lifelines for financially distressed hospitals. While acquiring systems frequently promise to restore financial stability, a widespread concern is that they do so by eliminating services, thereby reducing patient access. Yet, empirical evidence on whether and why service reductions versus expansions occur remains limited. This study examines the mechanisms underlying post-acquisition service line changes by focusing on post-M&A IT integration as a key driver. Moving beyond prior research that treats M&A as a black box, we theorize that integration operates at multiple geographic levels—local (hospital service area), state, and national—depending on the acquirer's pre-existing footprint. These integration pathways shape whether services are expanded, eliminated, or reallocated across system hospitals. We test our hypotheses using a sample of 688 target hospitals acquired by 209 MHSs between 2007 and 2018. The results reveal heterogeneous effects of IT integration across geographic levels. Greater integration at the local and state levels

is associated with increased service expansion at target hospitals, alongside a reorganization of service delivery: targets are more likely to outsource certain services to nearby system affiliates while discontinuing in-house provision of those services. Contrary to prevailing perceptions, we find that, acquisitions by large MHSs lead to net service expansion at target hospitals. We reconcile this apparent paradox by showing that perceived service reductions often reflect a shift in governance mode—insourcing to intra-system outsourcing—rather than true loss of access.

WD03

302A

Healthcare Access, Equity, and Operations

Invited Session

Healthcare Operations Management

Chair: Dilay Ozkan, University of North Carolina at Chapel Hill

Co-Chair: Serhan Ziya, University of North Carolina, Chapel Hill, NC, United States

1 - Patient Demographics in Emergency Department Boarding: Differences in Transfer to Inpatient Care

Sukriye Nilay Argon, University of North Carolina, Chapel Hill, NC, United States, Dilay Ozkan, Serhan Ziya

Emergency department (ED) boarding is a major contributor to prolonged lengths of stay, ED crowding, and reduced patient satisfaction. Boarding begins when an ED provider makes the decision to admit a patient and ends when the patient is physically transferred to an inpatient unit. Prior to transfer, patients are assigned to an inpatient care team. In this study, we examine these post-admission decisions using electronic health record data from a large academic medical center. We find that not all patients with an admission decision are assigned to an inpatient care team. In addition, some patients, who are assigned a care team are never transferred to an inpatient unit. We investigate how patient demographics are associated with these post-admission outcomes using weighted logistic regression and log-transformed linear regression models. Our results show that patients insured with Medicaid wait longer for inpatient care team assignment but have higher odds of receiving inpatient care than commercially insured patients. In contrast, uninsured patients have lower odds of receiving inpatient care. These findings highlight the role of government-supported insurance in access to inpatient care.

2 - Balancing Fairness and Efficiency in Scheduling for Maternity Care

Yuhang Du, London Business School, London, United Kingdom, Yueyang Zhong

We study maternity care delivery as a resource-constrained service system in which heterogeneous service requests compete for limited capacity under hard completion deadlines. Each request must be completed before its deadline, after which the service loses its value. In this setting, we distinguish between early and late engagers, that is, expectant mothers who seek care earlier or later relative to their clinical deadlines. Under the Earliest-Deadline-First (EDF) scheduling policy in National Health Service (NHS), late engagers tend to be prioritised, leaving early engagers systematically disadvantaged through longer delays and higher risk of being left without care. To mitigate this inequity, we propose a reservation-based scheduling policy that allocates a dedicated fraction of capacity to early engagers. Our analytical and simulation results show that such policy can substantially improve fairness in access to maternity care while maintaining high overall efficiency.

3 - Updating Policies for Healthcare Appointment Scheduling

Taha Huzeyfe Aktas, University of Amsterdam, Amsterdam, Netherlands, Alex Kuiper, Chintan Amrit

As is common in healthcare, appointment scheduling serves as a tool for managing the limited availability of key resources such as specialists and high-cost diagnostic equipment. The primary goal of these scheduling systems is to maximize resource utilization while minimizing patient waiting times. Healthcare providers generally employ equidistant scheduling with uniform appointment slot sizes for all patients. Despite its widespread adoption, this approach is known to be suboptimal in the operations management literature. Moreover, dynamic intra-day aspects, such as a growing waiting room, have received relatively little attention.

In this paper, we propose an alternative paradigm based on a simple, easy-to-apply update rule during session runs. The strategy involves postponing future appointments when patient backlog increases. This approach can be readily integrated into conventional appointment schedules and yields significant performance improvements. We demonstrate that such updating policies can outperform optimized static schedules reported in the literature. Effective strategies rely on tightening inter-arrival times to mitigate idling while maintaining time buffers that allow flexibility to postpone appointments in response to overcrowding. This prevents excessive waiting times, improves fairness, and ensures high server utilization. Building on these insights, we propose a simulation-based hill-climbing algorithm to tune the update rule under changing environmental conditions and provide an accompanying web-based interface. Across benchmark instances, the resulting updating policies consistently outperform optimized static appointment schedules.

4 - A Bi-Level Optimization Model for Collaborating Hospitals in A Healthcare Network

Samuel Bakker, Ghent University, Gent, Belgium, Broos Maenhout, Premysl Sucha

This study envelopes a bilevel optimization problem derived from the Belgian healthcare landscape. The modeled problem depicts a region with multiple hospitals that can increase their efficiency by treating more patients of the same diagnosis-related group (DRG). These hospitals (i.e. followers) compete with each other for the same patients whilst maximizing their profits. The leader (i.e. network manager or government) on the other hand, tries to maximize the total number of treated patients by assigning subsidies to the respective follower hospitals. Economies of scale are modeled by decreasing the per-patient resource requirements for a DRG as the total volume of treated cases rises, using a piecewise linear approach. This allows us to model the realistic case in which focused factories achieve higher levels of efficiency than hospitals treating a lower volume of hospitals. A dedicated algorithm is developed using callbacks, cutting off integer bilevel infeasible solutions by means of lazy constraints.

WD04

302B

Innovation in Hospital Operations

Invited Session

Healthcare Operations Management

Chair: Lesley Meng, Yale University, New Haven, CT, United States

1 - Mitigating Service Incompletion via Sunk Costs: a Queueing Approach

Jimmy Qin, The University of Texas at Dallas, Richardson, TX, United States, Carri Chan, Jing Dong

Sunk-cost bias occurs when decisions are influenced by the time, energy, or money already invested, rather than considering the future costs necessary to achieve success. While typically viewed as a source of irrational behavior that leads to suboptimal outcomes, this paper investigates how sunk cost can be used as an operational lever to increase service completion rates in a congested service system. We run a controlled online experiment and find that the abandonment rate is significantly reduced for the group of participants who incur a larger sunk cost. To better capture the dynamics of service systems and their impact on customers' behavior, we study a queueing model with strategic customers, where customers experience a disutility of balking that is proportional to the sunk cost they incur. We characterize the equilibrium behavior of the customers, from which we further derive the optimal strategy for the service provider in terms of whether to provide real-time queue length information to customers as well as the optimal level of sunk cost to impose. Our results show that the sunk cost strategy is effective only when waiting information is provided and that using a non-zero sunk cost is optimal when the queueing system is moderately congested. Numerical experiments confirm that imposing a non-zero sunk cost can substantially increase system throughput. We offer practical guidance for service managers aiming to influence customer behavior and manage congestion effectively.

2 - From Waiting to Care: How Pre-Treatment Waits Reshape Physician Decisions

Danqi Luo, UC San Diego, San Diego, CA, United States, Luxi Shen

Emergency departments face persistent congestion and pre-treatment waiting, yet little is known about how pre-treatment waiting shapes physician behavior at first contact and downstream operations.
We analyze electronic medical record and cost data from two hospital systems: a U.S.\ teaching hospital with acuity-prioritized intake and a Chinese hospital that operates first-come-first-served within departments. Using operational timestamps, we verify intake discipline and align estimation with operational queues. We identify the causal effect of pre-treatment wait using registration-time counts of patients still ahead in the relevant queue (high-priority patients in the U.S.; same-department patients in China) and a ransomware diversion shock that increased congestion in the U.S.\ system.
Across both settings, longer waits increase first-round diagnostic ordering (1-2 additional orders per hour in the U.S.; 1-7 across Chinese departments). Waiting also increases time devoted to the initial evaluation when front-end capacity is slack, but this time response collapses under high physician workload and high bed occupancy, while ordering expansion persists. We find no consistent improvements in disposition accuracy or 30-day revisits, but clear operational costs through longer post-wait service time and higher expenditures. A randomized vignette experiment with practicing physicians reproduces the ordering response, supporting a cross-agent spillover from patient waiting to physician decisions.
Our findings suggest that more than an access metric, pre-treatment wait serves as a critical input that shifts utilization and downstream resource use. Practical levers include improving communication during the wait and making downstream diagnostic constraints more visible at order entry so ordering decisions better reflect capacity.

3 - When Nurses Override the Algorithm: Human Discretion in AI-Assisted Emergency Department Triage

Lesley Meng, Yale School of Management, Yale University, New Haven, CT, United States, Huifeng Su, Rohit Sangal, Edieal Pinker

Emergency departments increasingly deploy AI-driven clinical decision support to standardize triage, yet the human decision-maker remains the final authority. We study a decision supporting tool that provides real-time, risk-based acuity recommendations to triage nurses across two EDs within a large academic health system. We show that despite the tool’s demonstrated gains in triage efficiency, nurses override the AI recommendation frequently. Using detailed administrative and electronic health record data, we investigate the operational mechanism of this discretion and its implications for technology adoption in the healthcare workforce.

4 - Evaluating AI in Echocardiography: An Experimental Study

Shuqi Yang, Yale School of Management, Yale University, New Haven, CT, United States, Lesley Meng, Katherine Clark, Robert McNamara

We examine the operational impacts of integrating an AI tool into echocardiography measurement workflows through a controlled experiment involving sonographers. Our research explores effects on measurement consistency, process efficiency, and clinician review workload, with implications for technology adoption, workforce planning, and quality management in healthcare operations.

WD05

302C

Data and Models in Public Health Policy

Invited Session

Public Health Policy

Chair: Joel Goh, NUS Business School, Singapore, Singapore

Co-Chair: Yvonne Huijun Zhu, Pennsylvania State University, University Park, PA, United States

1 - Cost-Effectiveness of a Network of Lethal Ovitraps for Dengue Control

Yvonne Huijun Zhu, Smeal College of Business, Pennsylvania State University, University Park, PA, United States, Shuzhen Sim, Joel Aik, Joel Goh

We conduct a system-level analysis of the cost-effectiveness of lethal ovitraps for dengue fever prevention, assessing both epidemiological outcomes and operational costs. The health benefits are captured using a multiple-infection epidemiological model and quantified as reductions in disability-adjusted life years (DALYs). Labor costs are estimated by modeling the maintenance workload for the ovitrap network as a Vehicle Routing Problem (VRP). Initially applied in Singapore, our analytical framework is further extended through a simulation study evaluating the prospective cost-effectiveness of implementing a city-wide ovitrap network in Bangkok, Thailand.

2 - Improving Policy Design and Epidemic Response Using Integrated Models of Economic Choice and Disease Dynamics with Behavioral Feedback

Hongru Du, University of Virginia, Charlottesville, VA, United States, Matthew Zahn, Sara Loo, Tijs Alleman, Shaun Truelove, Bryan Patenaude, Lauren Gardner, Nicholas Papageorge, Alison Hill

Human behavior plays a crucial role in infectious disease transmission, yet traditional models often overlook or oversimplify this factor, limiting predictions of disease spread and the associated socioeconomic impacts. Here we introduce a feedback-informed epidemiological model that integrates human behavior with disease dynamics in a credible, tractable, and extendable manner. From economics, we incorporate a dynamic decision-making model where individuals assess the trade-off between disease risks and economic consequences, and then link this to a risk-stratified compartmental model of disease spread taken from epidemiology. In the unified framework, heterogeneous individuals make choices based on current and future payoffs, influencing their risk of infection and shaping population-level disease dynamics. As an example, we model disease-decision feedback during the early months of the COVID-19 pandemic. Comparing the impacts of stylized policy options representing mandatory, incentivized/compensated, and voluntary work abstention, we find that accounting for disease-behavior feedback has a significant impact on the relative health and economic impacts of policies. Including two crucial dimensions of heterogeneity—health and economic vulnerability—the results highlight how inequities between risk groups can be exacerbated or alleviated by disease control measures. Importantly, we show that a policy of more stringent workplace testing can potentially slow virus spread and, surprisingly, increase labor supply since individuals otherwise inclined to remain at home to avoid infection perceive a safer workplace. This flexible and extendable modeling framework offers a powerful tool for understanding the interplay between human behavior and disease spread.

3 - When AI Listens First: Crisis Triage and Emotional Support in Mental Health Helplines

Yanhan (Savannah) Tang, SMU Cox School of Business, Dallas, TX, United States, Xiaoquan Gao

Mental health helplines play a critical role in providing timely support to individuals experiencing acute distress, yet they face severe capacity constraints and high call abandonment. Generative artificial intelligence (AI), particularly large language models (LLMs), has emerged as a potential first point of contact to support triage and engagement. In this setting, AI serves not only as a screening tool but also as a source of emotional support. We study when such AI systems improve helpline operations and how they should be designed and tuned to enhance outcomes. We develop a stylized queueing model of a mental health helpline in which callers first interact with a generative AI system and may be escalated to human counselors. The model captures the bifunctional role of AI in risk identification and emotional support, while accounting for waiting and abandonment. Our analysis shows that AI deployment can have nonmonotonic effects on outcomes: improvements in triage accuracy or conversational support do not necessarily improve system performance once congestion and abandonment are considered. Our findings offer insights into conditions under which it is beneficial to introduce AI in mental health helplines and identify concrete tuning targets for LLM design that improve system performance while balancing safety, access, and speed.

4 - Heartreference - a Cardiology-Specific Chatbot

Geneviève Larose-de Billy, University of Ottawa, Ottawa, ON, Canada, Samanvay Vajpayee, Christopher Sun

The high complexity of cardiovascular procedures and care, combined with constraints in human resources and time, creates a pressing need for technological solutions that provide readily accessible and highly reliable personalized information. In response, we propose HeartReference, a cardiology-specific retrieval-augmented generation (RAG) conversational agent. This work addresses key challenges in question-answer dataset

construction, training for multi-chunk retrieval, as well as safety and evaluation of healthcare-oriented LLM responses.

WD06

303

Emerging Applications of Operations Research in Healthcare

Invited Session

Medical Decision Making

Chair: M. Soheil Hemmati, University of Oklahoma, Norman, OK, United States

1 - Optimal Allocation of Organ Preservation Technologies to Improve Access

Ting-Tsen Lin, University of Minnesota, Minneapolis, MN, United States, Saumya Sinha

This project investigates the impact of technological advances on access to organ transplantation. Specifically, we consider novel technologies that preserve organ function for extended periods, enabling organs to be transported over longer distances, improving the quality of transplanted organs, and making certain marginal organs viable for transplantation. We study how such technologies can be efficiently implemented to maximize benefits for transplant recipients and, more broadly, enhance the overall transplant system. The system is modeled as a steady-state framework, formulated as an optimization problem. We present both analytical and numerical results.

2 - From Black to Grey: Improving Access to Antimalarial Drugs in the Presence of Deceptive Counterfeits

Jiatao Ding, Nanyang Technological University, Singapore, Singapore, Michael Freeman, Sasa Zorc

In malaria-endemic countries, the limited availability of affordable antimalarial medication has contributed to the widespread distribution of counterfeit drugs. This paper examines such markets to determine how philanthropic donors can best allocate limited funds to subsidize the purchase or sales of antimalarial drugs via private-sector distribution channels. We develop a game-theoretic model of the antimalarial supply chain, wherein the retailer strategically sources legitimate drugs from a donor-certified supplier, potentially counterfeit drugs from an uncertified supplier, or both. In contrast with the extant literature, we demonstrate that in the presence of counterfeits, exclusive reliance on purchase subsidies may no longer be optimal. Specifically, under donor budget constraints, offering sales subsidies that incorporate both legitimate and counterfeit drugs may be preferable, or in some cases, abstaining from subsidy provision completely. Additionally, we evaluate the implications of three pertinent non-subsidy strategies deployed to combat counterfeit drugs: imposing penalties for sourcing counterfeits, eliminating subsidies on counterfeit drugs with traceability technology, and implementing price controls. Our study concludes with extensive numerical analysis calibrated to malaria data from Mozambique. Overall, this paper offers strategic guidance for improving outcomes in the presence of counterfeit drugs. Our results highlight the need for governments and donors to carefully consider market-specific factors, such as retailers' pricing power and donors' budget constraints, when designing subsidy schemes and access policies for life-saving medicines. These insights could potentially be extrapolated to address similar challenges in other endemic disease contexts, offering a broader framework for enhancing public health in resource-constrained environments.

3 - Optimal Data-Driven Routing of the Mobile Lung Cancer Screening Unit of Oklahoma Based on Uncertain Screening Uptake and Incidence

Kiana Amani, University of Oklahoma, Norman, OK, United States, Iakovos Toumazis, M. Soheil Hemmati

Oklahoma ranks 8th in the United States for age-adjusted lung cancer incidence (63.3 per 100,000) and 5th for mortality (45.3 per 100,000), placing it among the most adversely affected states. Low-dose computed tomography (LDCT) is the standard screening modality for high-risk populations; however, screening uptake remains persistently low nationwide and is particularly poor in Oklahoma (approximately 10% versus 16% nationally). Mobile LDCT screening has been introduced to improve access in rural and underserved areas lacking fixed facilities. Despite this, current programs are largely reactive, operating on an “on-request” or ad-

hoc basis rather than being strategically allocated to regions with the greatest expected benefit. In this work, we develop a stochastic routing optimization model for a mobile lung cancer screening unit under uncertainty in both participation and screening outcomes. The model integrates county-level incidence data, observed screening uptake rates, and geographic characteristics. It is applied to counties with limited access to LDCT services to identify routes that maximize the net benefit of screening. The objective function captures quality-adjusted life years (QALYs) gained from early detection, adjusted for disutility from false positives, and incorporates screening, transportation, and operational costs using a willingness-to-pay threshold per QALY. The model generates optimal routing policies across representative scenarios and reports key outcomes, including total individuals screened and both screening-attributed and net QALYs gained.

4 - Identifying Minimum Performance Standards for Ovarian Cancer Screening to be Effective for the General Population Using Simulation Models

Sayaka Ishizawa, The University of Texas MD Anderson Cancer Center, Houston, TX, United States, Larissa Meyer, Iakovos Toumazis

No ovarian cancer prevention program currently exists. Past randomized screening trials failed to show significant mortality reduction. Hence, strong preliminary evidence on the effectiveness of a future screening program will be required to conduct future screening trials. The objective of this study was to identify the minimum screening accuracy levels necessary for a single or combination of screening interventions to be effective.

We simulated ovarian cancer-related life events for 1 million women from the general population starting at age 10 until death leveraging a validated natural history model of ovarian cancer. We evaluated numerous annual ovarian cancer screening strategies differing screening start age (range 45-55); stop age (range 65-75); sensitivity for early-stage disease (range 0.25-1.0); sensitivity for advanced-stage disease (range 0.5-1.0); and specificity (range 0.85-1.0). Each scenario was evaluated in terms of the expected number of true positive findings, false-positive findings, and false-negative findings, stage shift, and ovarian cancer mortality reduction.

Assuming perfect screening sensitivity and perfect adherence, annual ovarian cancer screening may yield approximately 30% mortality reduction if screening is available for the general population for 30 years. Model-derived mortality reduction from scenarios like past screening trials was similar to the reported rates (11% in UKCTOCS vs. 11-13% in simulation). As specificity of screening improved, the number of false-positives, diagnostic surgical procedures, and surgical complications decreased.

This study provides important insights to decision makers regarding the minimum required accuracy levels, and balance between benefits and harms a future ovarian cancer screening program must provide to be effective.

5 - Early Detection of Preeclampsia Through Multilevel Neural Networks

Talayeh Razzaghi, University of Oklahoma, Norman, OK, United States, Rachel Bennett

Early detection of preeclampsia is essential for improving maternal and neonatal outcomes, yet predictive modeling is challenged by large-scale clinical data, class imbalance, and bias across patient subpopulations. We propose a multilevel deep neural network (MLDNN) framework for scalable and equitable prediction of preeclampsia risk. The approach uses graph-based representations and a coarsening–refinement strategy to reduce dataset size and computational burden while preserving key data structures. To address disparities across demographic groups, we introduce a fairness-aware extension, FC-MLDNN, which integrates fair coarsening into the multilevel learning process. We evaluate the proposed methods on large-scale clinical datasets and compare them with standard deep neural networks and baseline methods. Our results show that our proposed methods achieves competitive or superior predictive performance with significantly reduced training time. These findings demonstrate the potential of multilevel learning frameworks for efficient and equitable risk prediction in maternal healthcare.

WD07

U.S. Organ Allocation Policy

Invited Session

Public Health Policy

Chair: Meghan Meredith, NYU Langone Health, New York, NY, United States

1 - Dynamic Nonuse Risk Estimation for Timely Interventions in Transplantation

Osman Ozaltin, North Carolina State University, Raleigh, NC, United States, Umur Kose, Sait Tunc, Matthew Ellis

Purpose: Kidney nonuse remains high in the United States despite increasing demand for transplantation. Organ Procurement Organizations (OPOs) often rely on expedited, out-of-sequence offers to place hard-to-allocate kidneys, but these approaches are typically reactive, inconsistently applied, and lack transparency. We propose a systematic, dynamic prediction framework to proactively guide allocation interventions and improve fairness and timeliness in decision-making during the match run.

Methods: Using Organ Procurement and Transplantation Network (OPTN) data from 2016–2021, we model kidney allocation as a dynamic time-to-event process with competing outcomes: placement, nonuse, and continuation. We develop a two-stage survival framework consisting of a baseline model aligned with match-run initiation and a post-clamp model aligned with organ clamping, when cold ischemia time (CIT) begins. Models incorporate donor characteristics, spatial features (e.g., donor hospital location, transportation access), and dynamic covariates such as accumulated CIT and provisional offer responses. We estimate cause-specific and all-cause competing risk models, including neural survival models, and evaluate performance using time-dependent concordance indices. Model interpretability is assessed with SHapley Additive exPlanations (SHAP).

Results: Dynamic survival models demonstrated strong predictive performance throughout the match run. Post-clamp models achieved concordance indices above 0.80 for nonuse prediction across observation windows up to 72 hours, while baseline models showed declining performance as CIT increased. Incorporating temporal and spatial information improved identification of high-risk kidneys compared to static approaches. SHAP analyses highlighted accumulated CIT, proximity to major airports, regional transplant center density, and offer rejection patterns as key drivers of nonuse risk.

2 - Getting the Right Kidney to the Right Patient Through Simulation Optimization

Sommer Gentry, New York University Grossman School of Medicine, New York University, New York, NY, United States, Bayzhan Mukatay, Ruiqi Xue, Allan Massie

Deceased donor kidneys vary widely in the expected survival they would offer to particular candidates who are waiting for a transplant, yet the appropriateness of each kidney for the candidate to whom it is offered is mostly ignored in the kidney allocation system (KAS). Only the 20% of kidneys with the longest predicted graft survival are preferentially directed to recipients with the longest expected post-transplant survival. We hypothesized that graft survival could be improved by allocating all kidneys to longevity-matched recipients, rather than just the top 20% of kidneys. We used simulation optimization to design an asymmetric longevity matching scale for use in a weighted composite allocation score for allocating kidneys. Using simulation to compare potential kidney allocation policies, we found that longevity matching for all kidneys would have a range of positive effects: it would narrow age differences between donors and recipients, increase transplants for candidates with better survival probabilities, and slightly increase overall 1-year and 10-year graft survival.

3 - All Together Now: Simultaneous Offers to Decrease Nonuse of Hard-to-Place Kidneys

Meghan Meredith, NYU Langone Health, New York, NY, United States, Bonnie Lonze, Dorry Segev, Nicholas Wood, Allan Massie, Sommer Gentry

Almost a third of deceased donor kidneys are recovered and not used, while thousands of kidney transplant candidates die on the waitlist every year. To decrease nonuse, the kidney allocation system needs to adapt to efficiently place hard-to-place kidneys. We propose offering hard-to-place kidneys

simultaneously to centers within distance thresholds, rather than the current sequential allocation policy. Under our proposed simultaneous offers policy, a hard-to-place kidney is offered to all candidates at all centers within a distance threshold, and the highest-priority acceptor after 1 or 2 hours receives the kidney. We simulate nonuse, the burden of evaluating offers, and transplant rates under the current policy and our simultaneous offers policy using the Scientific Registry of Transplant Recipients' (SRTR) 2024 Organ Allocation Simulator (OASim). We found that simultaneous offers reduced overall nonuse from 21.6% to 12.5% and reduced nonuse of hard-to-place kidneys from 56.0% to 20.9%, compared with the current policy. We also found that simultaneous offers decreased the burden of evaluating kidney offers, and did not decrease the transplant rate among any subgroup of candidates. Simultaneous offers within distance thresholds for hard-to-place kidneys accelerate kidney allocation, reduce nonuse, reduce burden, and maintain equity.

4 - Targeted Mortality Estimation in Liver Transplantation

Maura Hegarty, Massachusetts Institute of Technology, Cambridge, MA, United States, Dimitris Bertsimas, Georgios Margonis

Analyses of non-randomized, real world data often suffer from informative censoring. A classical setting exists in healthcare where more acute patients are prescribed treatments and less acute patients go untreated. This is systematically implemented through liver transplant policy in the United States; the sickest patients have priority on the liver transplant waitlist leaving the less sick patients without access to transplant. Thus, training a counterfactual model to predict patients' survival without transplant is prone to optimistically overestimating the probability of survival. Further, this short-coming is not revealed with the global C-index. To address this challenge, we train models on data selected with a sex- and risk-stratified covariate matching between treated and un-treated patients. We compared the model trained on a selected data set with a standard approach on the full-training cohort and an inverse-probability of treatment weighting approach. We test our model on a standard benchmark, a stratified benchmark, and a benchmark developed to more accurately represent the patient population for which we wish to predict counterfactuals for. We find that training strategy significantly impacts baseline survival while global C-index changes are modest. We show that the targeted model discriminates the best in higher strata of MELD in a head-to-head comparison with a U-statistic than other approaches. A MELD-only comparison suggests that the target model adds discrimination ability to MELD in high-acuity patients. Through this approach, we show that targeted training cohort construction improves counterfactual survival modeling for policy-relevant transplant populations.

WD08

305A

Health Across the Globe

Invited Session

Global Health and Humanitarian Operations

Chair: Xueze Song, The University of Alabama, TUSCALOOSA, United States

1 - Sparse Adaptive Tests for Emerging Wildlife Diseases

Xueze Song, The University of Alabama, TUSCALOOSA, AL, United States, Jue Wang

Wildlife disease outbreaks require rapid detection with limited surveillance capacity. Yet, the outbreak could occur across large landscape, which complicates the effective resource allocation. We propose an adaptive sampling framework that leverages sparsity in observation statistics to achieve efficient computation and identify optimal or near-optimal solutions.

2 - The Sooner, The Better? Optimal Vaccination Policy with Limited Vaccine Supply

Qi (George) Chen, London Business School, London, United Kingdom, Miao Bai, Cuihong Li

We study the optimal single-dose vaccination policy during an infectious disease outbreak, considering both limited vaccine supply and imperfect efficacy, which provides partial immunity to each vaccinated

individual. The inclusion of imperfect efficacy introduces an additional compartment to the celebrated Susceptible-Infectious-Recovered (SIR) model, giving rise to an infinite horizon nonlinear optimal control problem. We derive a closed-form expression for the optimal vaccination policy under infinite administrative capacity and establish theoretical structures of the optimal policy under finite capacity. Our results suggest that delaying the start of the vaccination process may be optimal, especially when the vaccine is less effective, the vaccine supply is more limited, and the disease is more infectious. We identify key terms as an integral part of the analysis that characterize the individual-level benefit of vaccination and, importantly, provide insights into the intuition behind vaccination delay. Building on these theoretical findings, our numerical study verifies these insights based on sensitivity results. We demonstrate the significant benefit of delay in reducing the total number of infections compared to policies without delay. Furthermore, we extend our analysis to alternative model dynamics and objective functions, confirming that our key findings remain robust and our methodology applicable across different settings. Our study contributes to the methodological framework to analyze optimal vaccination policies for infectious disease control, and uncovers an important policy structure -- vaccination delay -- that has been overlooked in the literature but has important implications for practice.

3 - R&D Strategy and the Design of Clinical Trial Eligibility Criteria

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, Zhaowei She, Liang Xu

This presentation examines how pharmaceutical firms' R&D strategies shape clinical trial design, with a particular focus on how eligibility criteria change from Phase 2 to Phase 3 trials. The central premise is that trial design involves a fundamental tradeoff: narrower inclusion and exclusion criteria can produce a more homogeneous patient population and improve the likelihood of detecting treatment effects, but they may also reduce the generalizability of evidence to the broader population that will ultimately use the treatment. Using the drug–disease pair as the unit of analysis, this study investigates whether firms adopt different strategies in modifying trial restrictiveness, and whether financial constraints lead smaller firms to make less effective design choices. Methodologically, the project leverages LLM to compare, summarize, and rate the stringency of eligibility criteria across trials on multiple dimensions, generating structured measures of restrictiveness that can be validated against observable criteria such as age limits and complemented by non-generative NLP methods and clinician evaluation. The study aims to provide evidence on how organizational strategy and resource constraints influence evidence generation in clinical development.

WD09

305B

Improving Healthcare Delivery Using OR and Data

Contributed Session

Medical Decision Making

Chair: Sundara Natarajan Panchanatham, HKU Business School, Hong Kong, Hong Kong

1 - Cost-Effectiveness of a Coordinated Drone and EMS System for Automated External Defibrillators Delivery

Marziehsadat Rezaei, University of Louisville, Louisville, KY, United States, Carmine Sorgente, Domenico Serra, Monica Gentili, Evan Kuhl

Despite evidence demonstrating the potential of drones for rapid automated external defibrillator (AED) delivery and their association with improved survival in out-of-hospital cardiac arrest (OHCA), there is limited evidence on whether integrating drones into emergency medical services (EMS) improves survival outcomes or is cost-effective. This study aims to determine the cost-effectiveness of coordinated EMS response and drone delivery of AEDs for OHCA. Using real-world data from Jefferson County, Kentucky, and optimization and predictive models, we generated 210 candidate drone-only and coordinated EMS–drone network configurations and estimated survival for each. We linked survival and response outcomes to cost estimates from the literature. This approach allowed us to compare the cost and survival across configurations and identify the best trade-off between survival gains and investment value. Across all evaluated configurations, coordinated EMS–drone networks optimized for survivability achieved the

greatest survival gains and favorable trade-offs between cost and survivability. The current EMS baseline had a mean response time of 7.68 min. In coordinated designs, mean response times were 8.9%–75.9% shorter than the EMS baseline. The expected number of survivors to hospital discharge under the EMS baseline was 126 (8.65%), which increased to 146 (10.04%)–357 (24.49%) across coordinated networks. The best-performing coordinated design with 95 platforms cost $7,403 per additional survivor. Survival gains plateaued beyond 95 platforms, indicating diminishing returns with further expansion. Coordinated drone and EMS systems for AED delivery were found to be cost-effective relative to Drone-only and standard EMS response systems.

2 - Optimal EMS Disposition Under Ed Congestion Safety

Apurv Shukla, University of Michigan, Ann Arbor, MI, United States, Yawo Kobara

We develop a dynamic model of emergency medical services (EMS) disposition in which a planner chooses between transporting patients to the emergency department (TX) and treating-and-releasing them in the field (TR), subject to two interacting queues: ED congestion and ambulance availability. Transport increases future congestion and depletes ambulance capacity, while field release avoids these externalities but carries clinical risk. We characterize the optimal policy as a monotone switching curve in the (census, availability) state space: transport is optimal when ED congestion is low relative to ambulance supply, and the threshold shifts upward as congestion rises. We prove comparative statics in the safety multiplier, congestion weight, and blackout penalty, and identify a regime shift when clinical risk is itself congestion-sensitive.

We calibrate the model to Ontario's emergency department system using CIHI NACRS administrative data (April 2024–March 2025). Ontario's 6.62 million annual ED visits generate an estimated $2.83 billion (CAD) in economic burden, of which 15.8 million waiting-hours — valued at $178.84 per hour — are responsive to arrival-rate changes. Counterfactual analysis shows that diverting 10–20% of the province's 1.18 million emergency transports to field disposition would save $50–202 million annually, depending on the queueing response. Comparative statics reveal that accurate estimation of clinical risk — and its dependence on system congestion — is the most important input for calibrating implementable disposition protocols. These findings support congestion-responsive EMS policies that condition transport decisions on real-time ED census and ambulance availability.

3 - Predicting Hospital Patient Footfall and Waiting TIMES: a Machine Learning Approach for Enhanced Operational Efficiency

Meera Simon C, Masters' Union, Gurgaon, Thrissur, India, Shovan Chowdhury

Improving patient experience requires enhancing healthcare system efficiency, as patient satisfaction is directly linked to effective management of patient footfall. Waiting times critically influence both operational efficiency and perceptions of care quality, with longer waits reducing satisfaction and impacting medical service delivery. Accurate forecasting of patient footfall and waiting times is essential for optimizing hospital operations. This study addresses these challenges through predictive modeling. Statistical time series models (ARIMA, SARIMA) and machine learning algorithms (Random Forest, Support Vector Regression, XGBoost, Neural Networks) forecast daily patient footfall using historical and time-based features. Statistical models accurately predict daily patient footfall in hospital settings with historical and time-based features. The Random Forest model delivers the most reliable and generalizable performance across departments. For waiting time prediction in specific departments, Random Forest, Support Vector Regression, and Neural Networks achieve high accuracy. Key features and their influences are identified via interpretable techniques, including partial dependence plots and SHAP values. Random Forest emerges as the top performer among machine learning models for both footfall and waiting time predictions, offering robust generalization. These insights enable proactive resource allocation, such as staffing adjustments and bed management, reducing bottlenecks. Implementing these models supports the Value Agenda by minimizing waiting times, elevating patient satisfaction, and boosting operational efficiency. Hospitals can integrate forecasts into real-time dashboards for dynamic decision-making, ultimately delivering higher-value care at lower costs. This approach fosters data-driven healthcare transformations.

4 - Telemedicine Rollout and Resource Reallocation in Health Systems: a K-Modal Approach to Heterogeneous Care Consumption

Jie Hu, University of Pennsylvania, Philadelphia, PA, United States, Yifan Jiang, Yiwen Lu, Runze Li, Yong Chen

Many operational settings generate outcomes with multiple recurring peaks rather than a single typical level. In healthcare, for example, downstream charges often cluster around several distinct resource-intensity regimes, reflecting different care pathways. Similar patterns arise in service pricing, travel demand, and transaction amounts. In these settings, decision makers often care less about the conditional mean than about the few outcome levels that are most likely to occur and how covariates shift probability mass across them.

We propose a K-modal regression framework for multi-peak prediction and decision-oriented analytics. Instead of returning a single prediction, the model estimates the top-K conditional modes of the outcome distribution together with their relative likelihoods. This produces an interpretable menu of likely outcome levels for each decision context and reveals how observable characteristics are associated with different latent behavioral or operational regimes.

The framework supports applications such as resource allocation, price menu design, demand segmentation, and capacity planning by identifying the outcome levels that are most operationally relevant. In our motivating healthcare application, the approach is designed to capture how telemedicine adoption can reallocate patients across low-, moderate-, and high-intensity care pathways, even when average effects appear modest. More broadly, the framework offers a practical way to study multimodal outcomes in service and operations environments where understanding the most likely regimes is more actionable than predicting a single expected value.

5 - Cost and Operational Projections of Adding at-Home Hpv Self-Sampling in a Federally Qualified Health Center: a Workflow-Informed Discrete-Event Simulation

Ryan Suk, Emory University, Atlanta, GA, United States, Maryam Kheirandish, Hyeiyeon Im, Mackenzie Czerner, Jessica Calderón-Mora

Introduction: At-home HPV self-sampling may improve screening reach, but its clinic-level cost and operational effects remain uncertain. We estimated the projected cost and workflow consequences of adding at-home HPV self-sampling to cervical cancer screening in a single federally qualified health center (FQHC).

Methods: We developed a 1-year, workflow-informed episode-level stochastic discrete-event simulation comparing usual in-clinic screening with a hybrid strategy adding at-home self-sampling. Inputs came from an FQHC for staffing time, wages, and kit costs, supplemented by published estimates for return, reminder response, specimen loss, invalidity, and turnaround. We evaluated scenarios in which at-home self-sampling either expanded completed screening beyond baseline (growth) or substituted for a share of baseline in-clinic screening while total initiated screening volume remained constant (switching). Outcomes included completed screens, gross annual cost, and staffing demand.

Results: Under status quo, the model projected 757 completed in-clinic screens annually and a gross annual cost of $154,662. In a 20% growth scenario, the hybrid strategy increased completed screens to 909, requiring about 206 initiated at-home episodes and increasing gross annual cost by $24,440 ($22,444 nonlabor; $1,996 labor); medical-assistant demand increased by 141 hours/year. In a 20% switching scenario, shifting clinic volume to at-home testing reduced gross annual cost by $12,997 but decreased completed screens by 40; provider demand fell by 47 hours/year. Findings were similar in a conservative sensitivity analysis restricted to directly measured local inputs.

Discussion: Workflow-based simulation may help safety-net clinics estimate near-term staffing and gross cost requirements before rollout, supporting operational and financing preparedness.

6 - Drop in the Ocean: Inventory Management for Public Cord Blood Banks

Sundara Natarajan Panchanatham, HKU Business School, Hong Kong, Hong Kong, Henri Groenevelt, Sameer Hasija

Treating many blood-related diseases requires transplantation of genetically compatible hematopoietic stem cells (HSCs) obtained either from live bone marrow (BM) donors or from umbilical cord blood (CB) units. BM registries maintain information on potential donors, whereas CB banks store collected CB units. This study examines the joint capacity management problem faced by these two complementary institutions. Because there are more than 10 million genetic variants, inventory is limited relative to demand heterogeneity, and replenishment is random, the compositions of BM registries and CB banks are interdependent, stochastic, and dynamically evolving over time. BM and CB also differ in supply processes, costs, matching criteria, and clinical implications, creating trade-offs that make exclusive reliance on either source undesirable. We develop a simulation-based optimization method to estimate time-varying matching probabilities and embed the resulting regression parameters in a mathematical model that closely reflects operational reality. We compare the results with those from a simplified analytical model and show the value of explicitly capturing system dynamics. In addition, we develop a simulation-based optimization framework to design cost-effective CB inventory policies, including screening rules for incoming units, since some units do not contain sufficient stem cells. The framework identifies admission policies that retain only units that improve the overall value of the CB bank while informing joint capacity decisions for BM registries and CB banks.

WD10

306A

Data-Enabled Management of Chronic Conditions

Contributed Session

Data Analytics

Chair: Tajma Thongrakard, Northeastern University, Boston, MA, United States

1 - The Impact of Population-Level Uptake of Glp-1 Agonist Medications on The Future Supply and Demand for Kidney Transplants in The U.S

Corwin Rhyan, University of Michigan, Ann Arbor, MI, United States, David Hutton, Amirhossein Moosavi, Neehar Parikh, Mariel Lavieri, Yuwei Zhuo

GLP-1 receptor agonist medications have increasingly been used to treat diabetes and obesity in the U.S, with an estimated 31 million patients having used the medications through 2025. Clinical trials have found these drugs lead to significant weight loss, reductions in cardiac events, and lower risks for developing Type II diabetes. Given that the pathway from obesity to diabetes to diabetic kidney disease is a primary driver of the need for kidney transplant, we investigate the impact of increasing use of GLP-1 agonists on the need for kidney transplants in the U.S. through changes in the prevalence of obesity and diabetes. We extend these analyses to also consider the effects on the available supply of kidneys, as obesity/diabetes can be a barriers to kidney donation. In combination, we forecast under what scenarios GLP-1 use would be expected to reduce the current shortage of kidneys for transplant.

We conduct these analyses by developing a compartmental state-transition Markov model for U.S. adults at risk for developing obesity and diabetes. We construct population profiles of those at risk from the National Health and Nutrition Examination Survey (NHANES) and estimate the progression of obesity and diabetes through end-stage kidney disease and transplant, using evidence from academic literature and the Michigan Model for Diabetes. Our Markov model is calibrated using historical NHANES population and transplant wait list data to ensure fit with actual observed kidney transplant need, and we

then forecast future kidney transplant trends under different scenarios of U.S. population GLP-1 use and effectiveness.

2 - Early Prediction of High-Cost Healthcare Utilization Using Explainable Machine Learning: Evidence from Meps Panels 26 and 27 (2021&Ndash;2023)

Tajma Thongrakard, Northeastern University, Boston, MA, United States, Ajit Appari

Healthcare expenditures are highly concentrated among a small proportion of individuals, making early identification of high-cost patients critical for patient-centered care management and effective. This study applies supervised machine learning to predict future healthcare expenditure levels (one year in advance) using longitudinal data for Medical Expenditure Panel Survey (MEPS) Panels 26-27 (years: 2021–2023).

The analytic cohort comprises 10,252 non-cancer adults (age>=18yrs). Future (year-2) total expenditure is categorized into six ordered tiers (Zero, Very Low, Low, Moderate, High, Very High) using inflation-adjusted (expenditure adjusted to 2023 constant dollars using Medical Care Consumer Price Index), percentile-based thresholds. Predictors include patients' current (year-1) health status, healthcare utilization and expenditure components (e.g., Rx, Inpatient, ED, Outpatient), and demographics and socioeconomic characteristics.

We evaluated six model variants: Logistic Regression, Improved Logistic Regression, Random Forest, Improved Random Forest, XGBoost, and Improved XGBoost. All models showed acceptable discrimination ability across expenditure levels (macro ROC-AUC: 76.7% to 77.7%), with Improved Random Forest having strongest balanced performance (macro F1 = 0.406). SHAP-based explainability analysis identifies prior health expenditure and utilization components as dominant predictors across all models.

Therapeutic category analysis reveals Cardiology patients, most-prevalent cluster (~49%), have relatively lower cooccurrence (18-41%) across other specialties. Whereas Nephrology patients, least-prevalent (~1%), exhibit higher cooccurrence (24-100%) indicating higher care coordination needs. Median expenditure for Nephrology patients was 9-fold higher compared to patients without kidney disease.

Our machine learning modeling support the feasibility of early risk stratification for targeted care management and resource planning using data from contemporary nationwide healthcare surveys.

WD11

306B

Impact of Financial and Insurance Systems on Healthcare

Contributed Session

Finance, Insurance, and Payment Models

Chair: Fardin Ganjkhanloo, Johns hopkins University, Washington, DC, United States

1 - Who is Paying for Performance? &Ndash; Financial Impact of Performance-Based Payment to Healthcare System Stakeholders

Bo Shi, Morehead State University, Morehead, KY, United States, Wen Chen

Effective in 2013, the Centers for Medicare & Medicaid Services (CMS) started to apply value-based payment to Medicare inpatient services offered by more than 3,000 CMS certified hospitals. Healthcare service quality became an important factor in determining providers' reimbursement. A genetic question is the subsequent financial impact on all stakeholders in the system. To shed some light on this critical question, we develop a three-party game theory model. Our model captures main features of the third-party payment system and allows stakeholders optimize their objective functions in the three-party game. Unlike traditional riskless reimbursement setup, our model introduces the downside risk to hospital payment. And health service quality is examined as a risk factor affecting the probability that the hospital

gets full payment. Equilibrium solutions suggest that policyholders of commercial insurers are going to experience premium increase, which may share the cost on quality improvement.

2 - How Health Systems Interpret and Operationalize Payment Reform Signals: Evidence from Bpci and Cjr in Orthopedic Clinical Practice

Marcus Sharper, University of South Carolina, Columbia, SC, United States

Bundled payment programs like the Bundled Payments for Care Improvement (BPCI) and Comprehensive Care for Joint Replacement (CJR) represent major Medicare reforms intended to incentivize value-based orthopedic care. Understanding how health systems interpret payment signals and how they operationalize them across the care continuum is essential for informing future payment policy. This study synthesizes evidence on organizational responses to BPCI and CJR, including spillover effects on non-participating settings and the sustainability of resulting care redesign strategies. We conducted a comprehensive review of 30 studies published between 2014 and 2025 examining implementation approaches, episode management strategies, and downstream impacts. Health systems consistently interpreted bundled payment incentives as directives to reduce 90-day episode costs by targeting controllable levers such as site-of-care decisions, patient selection, and post-acute care utilization. Common operational changes included preoperative risk optimization, enhanced care coordination, standardized discharge pathways emphasizing home recovery and physician gain-sharing arrangements. Spillover effects were observed across Medicare Advantage and commercially insured populations as systems adopted uniform protocols independent of payer. While early cost reductions and quality improvements were documented, savings diminished over time due to baseline spending variation, strategic participation decisions and shifts in surgical site of service. Overall, bundled payment programs effectively stimulated operational changes that improved value, but long-term sustainability depends on program design, target-setting methodology and organizational capacity for continuous care redesign. These findings highlight the broader influence of payment reforms on system-wide practice patterns and underscore the need for future models to account for heterogeneity in baseline spending and implementation capability.

3 - The Lingering Affordability Crisis: Why Medicare Part D Coverage Still Falls Short for Many Patients

Rebekah Angove, Patient Advocate Foundation, Hampton, VA, United States, Kathleen Gallagher

Despite Medicare Part D coverage, prescription affordability remains a major challenge for beneficiaries with chronic conditions. High out-of-pocket costs strain household budgets and compromise access to essential care. The Inflation Reduction Act (IRA) introduced landmark reforms to address financial imbalances and improve medication affordability across the system.

We conducted a longitudinal, self-administered survey of patients served in 2024 by Patient Advocate Foundation. The web-based survey was administered over a three-week period in January 2025 and December 2025.

Participants (n=2,168; n=1,166) were older adults living on fixed incomes (84% Social Security/Disability) as their primary source of income. Approximately half (53%; 47%) reported annual household income >$35,000.

Cost-related nonadherence emerged as a central concern, with (60%; 54%) indicating they would be likely to skip doses, reduce medication usage, or discontinue treatment entirely if costs increased. A significant proportion of patients already reported altering their treatment due to affordability concerns, including those who did not fill a prescription (36%; 28%), delayed filling prescriptions (40%; 32%) skipped doses (31%; 23%) or stopped/substituted medications (45%; 41%). Many (70%; 68%) requested generic instead of brand name drugs.

Our findings suggest the IRA is meaningfully narrowing the affordability gap for some patients, yet a disconnect persists between benefits and the economic realities of older adults. Many Medicare beneficiaries still face residual cost barriers across the continuum of care that impact how they seek and

receive care. These realities require continued attention to understand the social, economic, and policy-related changes necessary to continue to close the gap.

4 - The Hidden Cost of Decision Fragmentation in U.S. Healthcare&Rsquo;S Value-Based Care Transition

Gulce Cuhaci Tasdelen, Istanbul Bilgi University, Stamford, CT, United States

U.S. healthcare spending reached $5.3 trillion in 2024, 18% of GDP (CMS 2026), with approximately 25% — $760–$935 billion — classified as waste (Shrank et al. 2019). Administrative complexity, the largest waste category at $266 billion, remains the only category with no identified evidence-based interventions. We argue a substantial portion stems from the "decision governance gap": healthcare organizations' systematic challenge to integrate decision-making across competing operational, financial, clinical, and regulatory objectives during value-based care (VBC) transitions. Six decision-fragmentation channels were identified through structured review of industry reports (CMS, AMA, McKinsey, CAQH), application of paradox theory (Smith and Lewis 2011) to distinguish structural tensions from solvable inefficiencies, and multi-stakeholder friction mapping across payer, provider, and member functions. Using bottom-up micro-costing, we construct representative cases for a mid-sized health insurer (~500,000 members) and provider group (~20,000 attributed members). Case parameterization applies paradox navigation principles: each operational decision is modeled for both its intended effect and its contradictory downstream consequences across stakeholders and time horizons. Six channels are quantified: prior authorization friction, where 87% of physicians report PA increases total utilization (AMA 2024); cost-quality decision misalignment; payer-provider incentive fragmentation; care journey fragmentation; quality measurement paradoxes; and regulatory decision overload. The payer case yields $92–$164 PMPY waste; the provider case $600–$1,100 per attributed member. National extrapolation with segment-adjusted multipliers and overlap adjustment produces a primary estimate of $80–$130 billion annually. We propose three decision governance interventions — cross-functional decision integration, multi-horizon impact assessment, and payer-provider decision transparency — as a distinct policy category.

5 - Understanding Hospital Responses to Bundled Payments: a Structural Approach for Prospective Policy Evaluation

Fardin Ganjkhanloo, Johns Hopkins University, Washington, DC, United States, Kathryn Linehan, Matthew Zahn, Melinda Buntin

Bundled payment programs shift financial risk to providers by replacing fee-for-service reimbursement with a single payment covering an episode of care, creating incentives to reduce spending while maintaining quality. Yet how hospitals navigate this tradeoff in practice remains poorly understood. We develop a structural microsimulation model of hospital discharge and length-of-stay decisions under bundled payment that makes the tradeoff between cost reduction and care quality explicit. Hospitals choose a discharge destination and length of stay for each patient to maximize a utility function that trades off financial profit against readmission risk, subject to historical discharge norms and patient heterogeneity. The model nests fee-for-service and bundled payment incentives through a blended profit function incorporating target prices, stop-loss and stop-gain limits, and post-acute care costs. We estimate the model using Medicare claims data covering approximately 430,000 inpatient lower-extremity joint replacement episodes across roughly 2,500 hospitals, calibrating and evaluating the model against the Comprehensive Care for Joint Replacement (CJR) demonstration program. Because the model is structural, it supports counterfactual policy simulations, including mandatory bundled payment models and different pricing methods, that trace how hospitals would reallocate patients across post-acute settings and adjust lengths of stay under alternative payment designs. The microsimulation model provides a tool for evaluating prospective bundled payment reforms and understanding the mechanisms through which payment incentives shape post-acute care utilization.

WD12

306C

Advances in Experimental Design with Applications to Clinical Trials

Invited Session

MSOM: Healthcare

Chair: Arielle Anderer, Cornell University, Johnson School, Ithaca, NY, United States

1 - Rethinking Clinical Trial Site Recruitment: a Dynamic Replanning Approach

Bogdana Jelić, University of Oxford, Oxford, United Kingdom, Arielle Anderer, Kleio Kipourou, Agni Orfanoudaki, Holly Wiberg

Large-scale, multi-site clinical trials are a cornerstone of modern drug development, enabling the recruitment of sufficiently large and diverse patient populations needed to rigorously evaluate the safety and efficacy of new treatments. Existing research has developed statistical approaches to model patient recruitment and inform site selection. Recent work incorporates machine learning for improved prediction, though these methods remain largely static and do not fully support adaptive, real-time decision-making. But when should recruitment plans be revised? What signals indicate underperformance, and which sites should be selected to most effectively meet trial objectives?

To address this gap, we focus on optimizing site footprint management for clinical trials, a key yet underexplored determinant of trial efficiency, cost, and timelines. We propose a data-driven, adaptive framework that formulates site footprint decisions as a sequential resource allocation problem under uncertainty, enabling more responsive management of site portfolios as trials progress. Our approach integrates predictive modeling of site-level recruitment with dynamic–optimization techniques, including partially observable Markov decision processes to support real-time decisions on site activation and closure. The framework jointly considers initial site selection and ongoing trial management, enabling continuous adjustment based on accumulating recruitment data.

Empirical validation will leverage extensive historical clinical trial data from GSK to evaluate improvements in recruitment efficiency, cost-effectiveness, and probability of on-time completion. By bridging machine learning and operations research, this work aims to deliver practical decision-support tools and theoretical insights. Ultimately, our methodology seeks to enhance clinical trial execution and accelerate the generation of high-quality evidence in drug development.

2 - Da be Active: Piloting Dynamic Adaptation

John Silberholz, University of Maryland Smith School of Business, Silver Spring, MD, United States, Hamsa Bastani

Dynamic adaptation is a promising strategy for scaling the impact of behavioral interventions by personalizing interventions based on individuals' needs. Key challenges include designing pilots that create sufficient dynamic variation in intervention strategies, identifying promising sources of heterogeneous treatment effects with limited and noisy data, and evaluating the resulting policies online. We report progress on these fronts on a physical activity intervention motivated by the BE ACTIVE randomized controlled trial, which demonstrated significant and durable step count gains in a high-risk patient population when using gamification or providing financial incentives. Our pilot, which we call the Dynamically Adaptive (DA) BE ACTIVE Pilot, allocates all participants to gamification to encourage step count increases. Participants who fail to meet their step count goals in a given week are randomized to one of three interventions in the following week: study staff outreach, financial incentive, or no additional intervention. Allocation probabilities are dynamically adapted based on accruing study data.

3 - Combinatorial Bandits with Recovery and Habituation (COBRAH): Optimizing Community Health Interventions Under Resource Constraints

Katherine Adams, University of Texas at San Antonio, San Antonio, TX, United States, Justin Boutilier, Qinyang He, Yonatan Mintz

Decision makers in developing economies often struggle to manage behavior-driven health programs with limited resources and uncertain participant responses. Community-based chronic disease interventions, such as diabetes programs, are especially challenging due to high attrition, heterogeneous participant

behavior, and response dynamics that evolve over time – declining from stigma or cognitive burden, or improving as health benefits accumulate. Program managers must allocate scarce community health workers while balancing exploration of participant progress against exploitation of current knowledge. We model this problem as a nonstationary combinatorial multi-armed bandit and introduce COBRAH, the first framework to incorporate endogenous nonstationarity into the combinatorial bandit setting. In COBRAH, rewards depend on latent participant states that drift in response to past resource allocations, creating a coupling between combinatorial actions and state-dependent reward evolution. This structure breaks standard concentration arguments and complicates regret analysis, as estimation errors propagate across jointly selected participant subsets. Our key theoretical contribution extends existing concentration bounds to this setting by characterizing how habituation and recovery dynamics interact with combinatorial action sets, yielding the first finite-time regret guarantees for combinatorial bandits with endogenous state transitions. We validate COBRAH through a real-world diabetes intervention case study, where our algorithms achieved up to a 54% increase in enrollment compared to baseline approaches. Beyond performance gains, our framework personalizes treatment by identifying optimal windows of participant receptivity and provides verifiable benchmarks to support investment decisions and long-term scalability of community health programs in resource-constrained settings.

4 - Optimal Enrollment Timing in Early-Stage Patient-Centered Clinical Trials

Amin Khademi, Clemson University, Clemson, SC, United States, Hanwen Liu, Qi Luo

Early-stage patient-centered clinical trials aim to establish the toxicity profile of investigational drugs and determine safe dose ranges for future trials using cohort-specific consent. Patients considering participation face intricate benefit-risk tradeoffs, particularly regarding the timing of their enrollment: Early enrollment corresponds to low doses with potentially low efficacy and toxicity with high uncertainty around them, while late enrollment corresponds to high doses with potentially high efficacy and toxicity with reduced uncertainty, but also an increased risk of disease progression and trial termination. This paper develops a Bayesian model where a patient determines the optimal enrollment time based on health status and evolving beliefs about dose toxicity and efficacy. The patient periodically learns about the benefits and risks of doses and enrolls in the trial at the optimal time to maximize the expected benefits. We show that the optimal policy is of a control-limit type, where the threshold is determined by when the benefit of taking the current dose exceeds the benefit of waiting for future doses. We also establish sufficient conditions for the existence of a control-limit enrollment time policy based on the patient's health status and beliefs. Our study offers valuable insights into the nuanced benefit-risk tradeoffs patients face in enrolling early-stage clinical trials, especially the dynamic interplay between their deteriorating health status and evolving beliefs about the drug's efficacy-toxicity profile.

WD13

307

AI-Driven Innovations for Care Coordination

Invited Session

eBusiness

Chair: TJ Guo, Kelley School of Business, Indiana University, Bloomington, IN, United States

1 - Grounding Safety in Uncertainty: a Constrained Reinforcement Learning Framework for Personalized Chronic Disease Management

TJ Guo, Indiana University, Bloomington, IN, United States, Indranil Bardhan, Wen Wen, Yicheng Song

Healthcare systems increasingly recognize the need to weigh patient experience alongside clinical outcomes in care delivery. Accommodating patient preferences in chronic disease management, however, introduces a critical risk: systematically deprioritizing clinically necessary interventions can accumulate irreversible harm over decades. Safe reinforcement learning (RL) offers a natural framework for sequential clinical decision-making with formal safety guarantees, but existing safety formulations are incompatible with long-horizon settings in which adverse outcomes such as mortality are near-certain

features of any sufficiently long trajectory rather than avoidable events, causing safety constraints to either trivially bind or uniformly fail regardless of policy behavior. We propose UNC-SAC, a constrained RL framework that operationalizes safety by constraining cumulative uncertainty in adverse event likelihood, estimated via a Beta–Bernoulli conjugate model. We examine the proposed approach in the context of optimizing preventive care recommendations for type 2 diabetes, under a reward function that balances the long-term clinical benefits of preventive care engagement against its inconvenience to patients. Using an established diabetes progression simulator, we evaluate UNC-SAC across a range of reward configurations that vary the emphasis placed on minimizing inconvenience, benchmarking it against standard-of-care policies and extant safe RL frameworks. Under the most adversarial configuration, where the reward function heavily emphasizes convenience and all competing safety formulations fail to prevent policy collapse, UNC-SAC alone produces a stable policy that accommodates convenience-seeking behavior while maintaining a clinically meaningful bound on mortality risk. These findings establish that grounding safety in outcome uncertainty enables principled, clinically accountable personalization in long-horizon clinical decision-making.

2 - Prescription Without Pressure: Voluntary use, Interstate Spillovers, and Interoperability in Pdmp Effectiveness

Xiru Pan, University of Miami, Miami, FL, United States, Nan Clement, Catherine Tucker, Niam Yaraghi

The U.S. opioid crisis has led states to enact `Prescription Drug Monitoring Programs' (PDMPs), which use information systems to improve prescribing safety and measurement of controlled substance prescriptions. However, their efficacy depends on both uptake rates and availability of appropriate data. This study exploits a natural experiment created by stricter PDMP mandates in California, where data is not shared with Oregon, and Washington, where data is shared with Oregon, to examine physician and patient behavior in Oregon, where PDMP use is discretionary. Using comprehensive physician-level data from the Oregon Health Authority, we show that mandates in neighboring states drive substantial patient inflows into Oregon. In response, Oregon physicians significantly increase voluntary PDMP engagement, especially in regions with access to interoperable, cross-state data. Increased PDMP use is associated with safer prescribing patterns, including reduced opioid prescribing and less doctor shopping. These findings demonstrate that the efficacy of PDMPs, especially when their use is not required, depends on interoperability.

3 - Digital Value by Design: a Configurational Analysis of Digitally Enabled Mental Health Services

William Olivera, Florida State University, Tallahassee, FL, United States, Aaron Baird, Yusen Xia, Lars Mathiassen

In response to the COVID-19 pandemic, telehealth availability at MH facilities increased by 77% in a single year, and by 2023, mental health accounted for 58% of all telehealth visits nationwide. Yet this rapid expansion has not uniformly translated into improved outcomes. The same digital technology that extends a rural psychiatrist's reach may create redundancy for a hospital-based provider already operating in a well-resourced setting. Furthermore, providers differ in training, scope, service mix, and digital literacy, and the communities they serve differ in need, resources, and infrastructure. Therefore, in this research, we ask: *how different configurations of provider capabilities, digital technologies, and community context co-create value in mental health services?*

To address this research question, we construct a multi-source panel (2017–2023) linking CMS data to facility-level foot traffic data from Advan Research, transaction patterns from SafeGraph Spend, and county-level indicators from County Health Rankings, CDC PLACES, and AHRF. Methodologically, we employ heterogeneous treatment effect estimation via causal forests, complemented by policy learning to derive interpretable configuration profiles.

Preliminary findings reveal that providers with distinct service portfolios co-create value through different mechanisms when their digital adoption aligns with community needs. Facility-level analyses incorporating foot traffic and spending patterns further reveal whether digital adoption draws patients

into underutilized facilities, alleviates strain on those at capacity, or expands geographic reach by reducing travel burden. In sum, our research demonstrates that the value of digital MH services emerges from the configuration alignment of what providers offer, what technologies enable, and what communities need.

4 - How Do Artificial Intelligence-Enabled Decision Aids Impact Health Outcomes for Patient-Centric Decision Making? Evidence from Knee Osteoarthritis Patients

Nidhish Nerur, University of Texas at Austin, Austin, TX, United States, Indranil Bardhan, Wen Wen, Kevin Bozic, Karl Koenig, Prakash Jayakumar

While early studies show positive impact of artificial intelligence-enabled decision aids (AI-DAs) on health and experiential outcomes, critical questions remain: How do AI-DAs influence long-term patient health outcomes through their impact on the quality of shared decision-making (SDM)? Do patients' baseline depression risk affect the relationship between the use of AI-DAs, SDM, and their long-term health outcomes? With personalized profiles and preferences, AI-DAs empower patients, particularly those with depressive symptoms who may require additional emotional support from physicians, to voice their concerns during SDM. Drawing on the Ottawa Decision Support Framework, cognitive load theory, and the compassionate human-AI collaboration literature, we posit that AI-DAs improve health outcomes by reducing decision regret, and this mechanism is strongest among patients with baseline depressive symptoms. To test our hypotheses, we conduct an empirical study using data from a randomized clinical trial (RCT) with knee osteoarthritis patients. The intervention group received an AI-DA tool that provides personalized risk-benefit reports and elicits patient preferences to guide SDM, whereas the control group did not receive this tool. Our analysis reveals that the use of AI-DAs increased patients' long-term health outcomes by improving SDM quality through their decision regret after 3 months. Patients with depression symptoms experienced a pronounced reduction in decision regret following the use of AI-DAs, and this is the key mechanism through which AI-DAs improve their long-term health outcomes. Our research highlights the value of AI-DAs in enhancing physicians' ability to provide compassionate care, especially during high-acuity, preference-sensitive SDM consultations.

Thursday, July 30, 2026, 8:00 AM - 9:15 AM

RA01

301A

Advancing Healthcare Delivery through AI and Optimization

Invited Session

AI in Healthcare

Chair: Hamsa Bastani, Wharton School, Philadelphia, PA, United States

Co-Chair: Angel Tsai-Hsuan Chung, The Wharton School, UPenn, Philadelphia, PA, United States

1 - Improving Access to Essential Medicines via Decision-Aware Machine Learning

Tsai-Hsuan (Angel) Chung, The Wharton School, UPenn, Philadelphia, PA, United States, Jatu Abdulai, Patrick Bayoh, Lawrence Sandi, Francis Smart, Hamsa Bastani, Osbert Bastani

A critical challenge in healthcare systems in low- and middle-income countries (LMICs) is the efficient and equitable allocation of scarce resources, particularly essential medicines. This problem is complicated by limited high-quality data, which restricts the applicability of traditional data-driven techniques. We propose a novel decision-aware machine learning framework for essential medicines allocation, which additionally leverages multi-task learning to ensure sample efficiency and catalytic priors to ensure equitable allocation. In collaboration with the Sierra Leone national government, we performed a staggered, nationwide deployment of our system as a decision support tool. Our econometric evaluation finds an estimated 19% increase in consumption of allocated products in treated districts, demonstrating its efficacy at improving access to essential medicines. Our tool was subsequently scaled nationwide,

covering an estimated 2 million women and children under five. Our work demonstrates how machine learning methods can improve efficiency at very low cost in resource-constrained global health settings.

2 - Online Reinforcement Learning for Delivering Activity Suggestions and Weekly Surveys in the Adapt Study

Xueqing Liu, Harvard University, Boston, MA, United States, Susan Murphy

For the upcoming ADAPT clinical trial, we develop a reinforcement learning (RL) algorithm to deliver twice-daily activity suggestions and weekly check-in surveys. The activity suggestions are intended to improve users' core affective experience (CAE) toward physical activity. At each morning and afternoon decision point, the RL algorithm determines whether to send an activity suggestion. The reward is delayed and observed through a weekly check-in survey administered on Sunday evening. This setting presents three key challenges: (1) rewards are sparse and delayed; (2) the weekly check-in survey serves a dual role by collecting both rewards and state information, while potentially reducing user engagement because of survey burden; and (3) participants may not complete the survey. To address these challenges, we formalize the problem as a periodic partially observable Markov decision process (POMDP) by incorporating a causal directed acyclic graph, where each week defines a period and CAE is treated as a latent state. We solve this problem through an equivalent periodic belief MDP. We consider two versions of the weekly check-in survey: a full survey, which reveals the true CAE, and a short survey, which provides an emission related to CAE that can be used to update the belief state. When participants do not complete any survey, we do not update the belief state under a specific missing-at-random assumption. We evaluate the proposed approach through simulations conducted on a testbed constructed from data collected in a previous ADAPT study.

3 - Optimizing Regional Systems of Risk-Appropriate Maternal Care

Abel Sapirstein, Georgia Institute of Technology, Atlanta, GA, United States, Meghan Meredith, Lauren Steimle

Maternal mortality in the United States exceeds that of all other high-income countries, Over 40% of high-risk pregnant women receiving care at facilities ill-equipped for their needs. Regionalized systems of care have been proposed as a remedy, yet no formal guidelines exist for such systems. We develop a bi-objective optimization problem to assign levels of care to facilities, balancing risk-appropriate care and travel time while ensuring financial viability. We capture bypass and referral behavior with a novel portfolio-based IP and solve via branch-and-price. In Iowa, strategic consolidation simultaneously could reduce travel time, improve risk-appropriate care, and promote financially viable operations.

4 - Generative AI in Outpatient Care: Evidence from Somaliland

Jimmy Qin, The University of Texas at Dallas, Richardson, TX, United States, Hamsa Bastani, Tsai-Hsuan (Angel) Chung

Healthcare delivery in underdeveloped countries is often constrained by limited clinical capacity, uneven provider quality, and substantial disparities in patient access and treatment outcomes. These challenges are exacerbated by resource shortages and heterogeneity in provider expertise, leading to significant variation in care processes and efficiency across facilities and clinicians. In this work, we study the impact of introducing an AI-assisted clinical documentation tool on healthcare delivery in such a resource-constrained setting. Leveraging a staggered rollout of the technology across departments and providers, we implement a difference-in-differences research design to causally estimate its effects on physician behavior and service outcomes.

RA02

301B

AI for Public Health and Healthcare Operations

Invited Session

AI in Healthcare

Chair: Michael Lingzhi Li, Harvard Business School, Boston, MA, United States

1 - Building AI Deployment Adaptability In-Vitro: A Specification-Robust Optimization Approach

Yu Ma, University of Wisconsin Madison, Madison, WI, United States

Machine learning models deployed in regulated settings, including clinical decision support, in particular, are certified against an initial specification comprising performance, fairness, operational, and safety constraints. In practice, this specification evolves throughout the deployment lifecycle: subgroup fairness thresholds tighten as equity audits accumulate, new patient subgroups require validation, operational targets adjust in response to alert-fatigue feedback, and clinical evidence updates feature-influence requirements. Each evolution can trigger a recertification cycle that, for a deployed clinical model, commonly requires 6–18 months and substantial validation effort. We propose a framework for specification-robust training that anticipates this evolution before deployment, embedding adaptability into the model in-vitro rather than relying on costly post-deployment retraining. Future specifications are modeled as a structured neighborhood around the certified specification, parameterized by stakeholder-elicited threshold-shift intervals and constraint-addition probabilities over clinically motivated constraint families. The trained model minimizes a weighted combination of primary loss and expected violation across sampled future specifications, with the weight calibrated to the institutional cost of retraining. We characterize the tradeoff between primary performance and specification robustness, establish comparative-statics results on how the optimal model responds to retraining cost and specification volatility, and demonstrate on a clinical decision support case study that specification-robust training meaningfully reduces expected retraining frequency at modest performance cost. The framework provides ML developers and hospital administrators an explicit, operationally grounded handle on the tradeoff between deployment performance and recertification burden.

2 - Managing Public Health Surveillance Operations with Endogenous Population Growth

Nan Liu, Boston College, Chestnut Hill, MA, United States, Hui Lin, Shan Wang, Rebecca Schnall

Public health surveillance is "the continuous, systematic collection, analysis and interpretation of health-related data" to guide public health actions (WHO 2012). In addition to traditional recruitment approaches, such as facility-based methods and community outreach, modern surveillance efforts increasingly rely on social media and online platforms. A central operational challenge in these efforts is recruiting a sufficient number of participants within a specified time frame to ensure unbiased estimation of population health status.

In this talk, we introduce a dynamic programming model to support decision-making in participant recruitment and service for public health surveillance. Using data from a large-scale study designed to assess HIV risk among American women, we demonstrate that our approach can substantially improve recruitment efficiency and reduce associated costs.

3 - Sustaining Human Vigilance in Ai-Assisted Drug Safety Review Through Controlled Error Injection

Michael Lingzhi Li, Harvard Business School, Boston, MA, United States, Kacey Fang

Human-in-the-loop workflows are increasingly deployed in high-stakes domains such as healthcare, finance, and content moderation, where large language models generate preliminary outputs that are subsequently reviewed by human experts. While this paradigm improves efficiency, it introduces a structural risk: as AI outputs become more accurate and polished, human reviewers may exhibit declining vigilance over time, leading to missed errors and reduced oversight quality.

This talk presents a field-based experimental study examining a simple and scalable intervention to mitigate this risk: controlled error injection. The core idea is to deliberately introduce rare, realistic errors into AI-generated outputs to sustain reviewer attention without materially degrading trust. We evaluate this approach in a longitudinal crossover experiment conducted with PharmD students performing simulated pharmacovigilance tasks over a 13-week period. Participants review AI-generated summaries of

adverse drug event reports under varying levels of injected error rates, enabling within-subject comparisons across conditions.

The study is designed to test whether (i) vigilance deteriorates over time in the absence of intervention, and (ii) there exists an intermediate level of error injection that improves long-run detection accuracy while maintaining acceptable efficiency and cognitive load. The empirical design combines high-frequency behavioral data with survey-based measures of trust and workload, and leverages fixed-effects and mixed-effects models to isolate causal effects of error exposure.

4 - Interpretable Design of Coordinated Multi-Agent Mobile Health Applications

Mohammad Zhalechian, Kelley School of Business, Bloomington, IN, United States, Hossein Hejazian, Soroush Saghafian, Omar Robles

Mobile health (mHealth) technology offers a promising avenue for delivering effective health interventions at low cost, but it faces several challenges in achieving real-life behavior change: (i) forming a holistic health habit requires considering multiple, interacting aspects (e.g., diet and physical activity); (ii) users have limited attention and tolerance for excessive or poorly-timed messaging; and (iii) interpretability is crucial for ensuring the effectiveness and trust of an intervention policy. We introduce a multi-agent analytical framework designed to address these challenges. Within this framework, decisions are made by an automated ``manager'' as well as several agents. Each agent is responsible for one aspect of the user's health (e.g., sedentary behavior, diet, or physical activity). The manager coordinates the agents to improve the overall effectiveness by minimizing the risk of user disengagement while guiding the system to induce healthy behaviors. We develop a holistic and interpretable policy with a provable performance guarantee. We validate our proposed algorithm using two randomized clinical trials (RCTs) run by our clinical researchers. Our results indicate that our framework can help improve the performance of mHealth applications, resulting in stronger health benefits across a variety of users.

RA03

302A

Innovations and Complications: Deploying Predictive Analytics in Healthcare

Invited Session

Healthcare Operations Management

Chair: Zhaowei She, Singapore Management University, Singapore, Singapore

1 - *CANCELED Rethinking Black Box Modeling in Healthcare: the Power of Interpretable Approaches

Courtney Paulson, University of New Hampshire, Durham, NH, United States, Margret Bjarnadottir, Daniel Smolyak

Amid the rapid expansion of complex machine learning and AI-driven prediction tools in healthcare, there is growing evidence that simpler, interpretable models often match or even outperform “black box” approaches designed to optimize predictive performance. Healthcare is a uniquely problematic field for black box modeling, as providers and researchers often require transparency in their predictions as well as accuracy. This leads to a common misconception: Modelers must choose between accurate predictions or interpretable models. Instead, there are numerous conditions under which interpretable models such as regression-based approaches show advantages not only in transparency but also in performance, especially when data are limited, noisy, or biased, all common issues in healthcare data. While these models offer clear, critical benefits in both interpretability and reproducibility, they also offer increased trust in their results, which can be especially vital for provider-patient communication.

Using a series of real-world healthcare datasets, we compare these interpretable models to complex methods like neural networks and boosting algorithms commonly used for healthcare prediction modeling. In doing so, we also address the common assumption that complexity inherently yields better results by showing that predictive performance is often equalized if not exceeded by using simpler, more transparent models. Computational savings can also be significant, particularly when models have

multiple hyperparameters to tune. Ultimately, we conclude that embracing simplicity in healthcare modeling can lead to predictions which are equally accurate, more meaningful, and less costly.

2 - The Impact of AI-Augmented Diagnosis and Treatment on Patient Health Outcomes

Minghong Yuan, The University of Texas at Austin, Austin, TX, United States, Wen Wen, Indranil Bardhan, Charles Taylor

Artificial intelligence (AI) is increasingly used to support clinical decision-making, yet evidence of its real-world impact on patient outcomes and healthcare value remains limited. We examine the effects of AI-augmented coronary diagnostics procedures on patient outcomes. Using comprehensive all-payer encounter data from the Maryland Health Services Cost Review Commission (HSCRC) spanning Q1 2019 to Q1 2025, we find that AI-augmented diagnostic procedures significantly reduce future hospital visits as well as subsequent healthcare spending. In addition, we find that AI-augmented procedures reduce the likelihood of unnecessary invasive procedures, indicating higher treatment quality. We also find that the benefits of AI-augmented procedures are lower for elderly patients, suggesting that AI may be less effective in reducing downstream healthcare utilization among clinically complex populations. Our study contributes to the ongoing debate on the role of AI in healthcare by providing evidence on how AI-enabled diagnostic tools affect patient health outcomes and downstream care utilization. While much of the existing discourse centers on whether AI can reliably augment or replace elements of clinical decision-making, our findings illustrate that AI's impact extends beyond predictive accuracy to measurably influencing care delivery and patient outcomes. Second, we advance the emerging literature on AI-augmented care by showing that AI embedded directly within diagnostic procedures can meaningfully reshape subsequent care utilization. Specifically, we demonstrate that AI-enabled diagnostics reduce the use of unnecessary invasive procedures, which is central to value-based care.

3 - When Prediction Undermines Calibration: How Excessive Risk Adjustment Drives Up Health Insurance Premiums

Zhaowei She, Singapore Management University, Singapore, Singapore, Kanix Wang, Yajing Chen

Risk adjustment is a key design feature of the Affordable Care Act (ACA) marketplace. By redistributing funds across plans based on enrollees' predicted medical spending, risk adjustment is intended to mitigate adverse selection and thus stabilize premiums. Existing designs often focus on improving the risk adjustment model's discriminatory ability to separate high-cost from low-cost enrollees. We show, however, that this approach can introduce a competing distortion, i.e., miscalibration, whereby plans are systematically underpaid for low-risk enrollees and overpaid for high-risk enrollees. In this paper, we first develop a theoretical framework that formalizes the trade-off between discrimination and calibration in risk adjustment design. This framework shows how gains in predictive fit can amplify payment distortions across enrollee subgroups, thereby affecting plan pricing incentives. We then provide empirical evidence from the ACA Marketplaces using an instrumented difference-in-differences design that exploits cross-market variation induced by the 2017 expansion of risk adjusters. The results reveal a "more is less" paradox: more predictive risk adjustment models can increase the dispersion of risk transfers, which raises market-average premiums. Specifically, we estimate that a one-standard-deviation increase in risk-transfer dispersion raises average premiums by approximately 1%. To address this trade-off, we propose a calibration-aware two-stage risk adjustment method that balances predictive accuracy and calibration. Finally, we conduct simulation analyses to demonstrate that this approach can reduce payment distortions while preserving predictive performance, thereby stabilizing premiums and supporting the long-term viability of the ACA marketplace.

RA04

302B

AI, Resource Allocation, and Behavioral Responses in Healthcare Systems

Invited Session

Healthcare Operations Management

Chair: Yue Hu, Stanford University, Stanford, CA, United States

1 - Generative AI Impacts Mental Health Care Market

Yiwen Jin, Haskayne School of Business, University of Calgary, Calgary, AB, Canada, Jingyuan Hu, Jiayi Liu

Generative AI systems are increasingly used as informal mental health companions, but whether they displace professional care or broaden access remains unclear. We study the market response to ChatGPT's release using de-identified transaction records, supplemented with website traffic, smartphone mobility, and national health survey data. We find that following the launch, unique users on digital mental health platforms rose by 24%, while spending per user fell by 12%. Both effects strengthen monotonically across successive ChatGPT generations and scale with AI adoption intensity. Further heterogeneity analyses show that substitution is confined to conversation-based services while clinically integrated platforms and in-person visits are unaffected; new users skew younger and higher-income, entering at lower spending levels than incumbents. These findings reveal that generative AI appears not simply to displace mental health care, but to create a parallel, largely unregulated entry channel into mental health support.

2 - A Spatial Queueing Framework for Resource Allocation with Strategic Customer Behavior

Yuwei Zhou, Indiana University, Bloomington, IN, United States, Baris Ata, John Friedewald, Yue Hu, Ali Randa

This paper develops a modeling framework for assessing the impact of policy changes to the deceased-donor kidney allocation system analytically. To do so, we formulate two models as building blocks: a fluid model of the transplant registery and a model of organ availability and geographic sharing. Both models incorporate the dynamic structural model of transplant candidates' accept/reject decisions for organ offers developed in Ata et al. (2020). They also incorporate important features of the kidney allocation system such as patient and donor heterogeneity, the degree of tissue matching, changes in the health status of transplant candidates as they wait, organ quality, cold-ischemia time as well as the geographical sharing structure. Combining the two models, we develop an equilibrium framework and prove that there exists a unique equilibrium. The equilibrium characterization leads to efficient computations for counterfactual studies regarding possible changes to the kidney allocation system that have been considered by policy makers.

3 - Learning-Based Capacity Allocation with Endogenous Channel Selection

Wenqian Xing, Stanford University, Stanford, CA, United States, Yi Zhang, Yue Hu

Many healthcare systems offer multiple care channels that differ in waiting time, continuity, and provider availability. Patients choose among these options based on their preferences and real-time congestion, leading to endogenous demand across channels. A central operational challenge is capacity sizing—how to allocate a limited pool of providers—since capacity decisions affect congestion, which in turn shapes patient choices. This challenge is compounded by the fact that patient preferences are typically unknown and must be learned from data. We study this problem using a queueing-based framework with endogenous patient choice and develop an online learning approach that jointly learns patient preferences and dynamically adjusts capacity allocation. Patients choose between standard primary care and express care based on observed congestion and individual characteristics, modeled through a flexible choice framework. We first analyze a full-information benchmark and establish stability via a Lyapunov approach, showing that stability holds if total capacity exceeds demand under mild conditions. Building on this benchmark, we develop an episodic online learning algorithm that estimates patient preferences from observed behavior while adapting capacity decisions over time, achieving sublinear regret. We validate our approach through a detailed case study using patient-flow data from a large academic hospital. Specifically, we calibrate patient preferences and quantify how patients trade off waiting time, continuity of care, and other attributes, addressing unobserved counterfactual waiting times with a nonparametric method. Numerical results demonstrate sublinear regret and show that the proposed approach effectively learns patient preferences while improving capacity allocation over time.

4 - Ai, Measurement, and Task Allocation in Secure Messaging for Primary Care

Sean Tsung, Stanford University, Stanford, CA, United States, Jisun Lee, Yue Hu, Song-Hee Kim, Vijaya Parameswaran

Artificial intelligence is increasingly used to support operational decisions in healthcare, yet it remains unclear when and how AI improves these decisions. We study this question in the context of task allocation in secure messaging, a critical channel for patient-provider communication. Patients send messages describing symptoms, requesting refills, and seeking medical advice, and care teams must determine whether each message should be handled by a primary care provider or support staff. Assigning more messages to support staff may reduce physician workload and improve response times, but may also lower resolution rates and increase downstream utilization when complex messages require physician expertise. Using secure messaging data from primary care clinics, we estimate the effect of provider type on response time, issue resolution, and downstream workload. Evaluating these decisions is complicated by two factors: message content is unstructured, complicating measurement of request type, complexity, and similarity across messages; and handling decisions are non-random, as more complex messages are more likely to be routed to physicians. We combine three labeling approaches (patient-selected, machine-learning-based, and agentic AI-generated) with two matching methods (embedding-based and agentic AI-assisted), examining how these six combinations affect matched samples and empirical conclusions. Different approaches can produce meaningfully different estimates of task allocation effects, demonstrating how measurement choices can be consequential for operational analysis. We further identify where agentic AI adds the most value and when simpler approaches suffice. Our findings offer guidance for integrating AI into healthcare operational workflows that rely on unstructured patient communications.

RA05

302C

From Data to Policy: Transforming Population Health through AI and Machine Learning

Invited Session

Public Health Policy

Chair: Jae Park, University of Central Florida, Orlando, FL, United States

1 - A Unified Framework for Structure-Aware Clustering and Heterogeneous Causal Graph Learning

Xiang Zhong, Univ. of Florida, Gainesville, FL, United States, Honglin Du, Muxuan Liang

In complex multivariate systems, interactions among variables are defined by dependency structures, often encoded as directed acyclic graphs (DAGs). However, dependency structures can vary across subjects, and ignoring this structural heterogeneity introduces bias and obscures subpopulation-specific dependencies. To address this, we propose Directed Acyclic Graph–based Dependency Clustering via Alternating Direction Method of Multipliers (DAG-DC-ADMM), a unified framework built upon Structural Equation Modeling (SEM) that jointly learns cluster assignments and cluster-specific dependency structures. We encode acyclicity via a smooth constraint and integrate a groupwise truncated Lasso fusion penalty (gTLP) to cluster subjects based on their structural similarity. This yields a nonconvex optimization problem that incorporates sparsity, acyclicity, and structural consensus constraints. We address the nonconvexity by using the augmented Lagrangian method and solve it with an adapted version of the Alternating Direction Method of Multipliers (ADMM) for difference-of-convex programs. For certain graph structures, such as upper triangular adjacency matrices, our algorithm is guaranteed to converge to a Karush-Kuhn-Tucker (KKT) point. Experiments demonstrate that our method recovers cluster-specific causal dependency structures with a high true positive rate and a low false discovery rate. This capability enables the robust discovery of heterogeneous dependencies across subjects where the subpopulation label is unknown.

2 - Do AI-Driven Proactive Full Capacity Protocols Mitigate Emergency Department Overcrowding?

Khalid Aram, Emporia State University, Emporia, KS, United States, Orhun Vural, Bunyamin Ozaydin, James Booth, Abdulaziz Ahmed

Emergency department (ED) overcrowding remains an operational challenge linked to treatment delays, poor outcomes, and inefficiencies. Full Capacity Protocols (FCPs) are commonly used to relieve crowding, but they are usually activated only after thresholds are exceeded, which limits their impact. This study evaluated whether proactive, prediction-driven FCP activation could improve ED performance compared with the reactive approach. We developed a hybrid decision-support framework that combined discrete-event simulation (DES) with machine learning. Operational data from a large U.S. academic hospital, including 308,196 ED visits from 2019 to 2023, were used to estimate patient-flow distributions and populate a validated DES model of ED and inpatient operations. A Support Vector Regression model was then trained on simulated hourly wait-count data to predict ED crowding six hours in advance. These predictions were embedded within the simulation to trigger proactive FCP activation before overcrowding occurred. Reactive and proactive deployment strategies were compared across three crowding thresholds using 365-day simulations, a 72-hour warm-up period, and 10 replications. The simulation model showed no significant difference from observed ED length-of-stay distributions ($p = 0.4848$), supporting its validity. At the moderate threshold ($\mu+2\sigma$), proactive deployment reduced severe overcrowding events from 8.03% to 2.97% of hours above threshold, decreased mean ED length of stay from 429.76 to 425.63 minutes, and shortened average protocol duration from 646 to 520 minutes, despite more frequent activations. These findings suggest that proactive FCP deployment reduced extreme congestion events and operational volatility while maintaining comparable or improved patient flow metrics relative to the reactive FCP.

3 - When Documents Disagree: Measuring Institutional Variation in Transplant Guidance with Retrieval-Augmented Language Models

Rema Padman, Carnegie Mellon University, Pittsburgh, PA, United States, Yubo Li, Ramayya Krishnan

Patient education materials for solid-organ transplantation vary substantially across U.S. centers, yet no systematic method exists to quantify this heterogeneity at scale. We introduce a framework that grounds the same patient questions in different centers' handbooks using retrieval-augmented language models and compares the resulting answers using a five-label consistency taxonomy. Applied to 102 handbooks from 23 centers and 1,115 benchmark questions, the framework quantifies heterogeneity across four dimensions: question, topic, organ, and center. We find that 20.8% of non-absent pairwise comparisons exhibit clinically meaningful divergence, concentrated in condition monitoring and lifestyle topics. Coverage gaps are even more prominent: 96.2% of question-handbook pairs miss relevant content, with reproductive health at 95.1% absence. Center-level divergence profiles are stable and interpretable, where heterogeneity reflects systematic institutional differences, likely due to patient diversity. These findings expose an information gap in transplant patient education materials, with document-grounded medical question answering highlighting opportunities for content improvement.

4 - Improving Emergency Department Wait Time Prediction at Data-Limited Hospitals Using Multi-Site Data and Auxiliary Outcomes

Jae Park, University of Central Florida, Orlando, FL, United States, Antonio Castellanos, John M Sheele

Federal agencies (e.g., the Agency for Healthcare Research and Quality) curate data from numerous healthcare institutions across the United States. Similarly, data consortia (e.g., OneFlorida) facilitate data-sharing agreements among multiple institutions, enabling centralized data collection from affiliated sites. Despite these efforts, multi-site datasets frequently face data quality challenges stemming from variations in clinical practices and outcome reporting across institutions. For example, some hospitals collect specific health outcomes while others do not, leading to systematic missingness. Hospitals with limited data quality and availability, such as rural hospitals, may therefore be unable to develop reliable prediction models, limiting their ability to benefit from emerging technologies. In this study, we address this challenge by leveraging an auxiliary outcome that is commonly collected across sites and closely related to the primary outcome of interest.

Our simulation experiments focus on estimating emergency department (ED) wait times across (1) multiple missing-data scenarios and (2) two hospitals with differing levels of observation noise: one with high data quality and one with limited data quality. To improve predictive performance, we incorporate ED length of stay as an auxiliary outcome during model estimation. Results show that leveraging the auxiliary outcome reduces absolute bias relative to complete-case and doubly robust baselines and reduces RMSE relative to these baselines. The largest improvements occur in more challenging settings characterized by higher missingness, greater observation noise, and higher feature dimensionality. This study helps enable the application of machine learning-based clinical decision support systems in hospitals with limited data quality and availability.

RA06

303

Decision Making under Uncertainty in Chronic Disease Management

Invited Session

Medical Decision Making

Chair: You Zhou, Purdue University, West Lafayette, IN, United States

1 - Location and Capacity Planning of Substance use Treatment Facilities for Maximizing Access and Population Health Outcomes

Zixuan Feng, Penn State, State College, PA, United States, Qiushi Chen, Paul Griffin

The opioid epidemic is a critical public health crisis, with at least 5 million US adults currently living with opioid use disorder (OUD). OUD is treatable through Medications for OUD (MOUD). However, delivering this care involves complex logistics. Treatment heavily relies on recurrent patient visits, especially during initiation. While treatment resources have expanded, significant barriers remain, including low initiation and high dropout rates. Although long travel distances and extended initiation wait times significantly influence patient choices and contribute to these poor outcomes, existing resource allocation models have not explicitly modeled this key dynamic. To address this, we developed a data-driven framework to optimize the geographic allocation of treatment capacity. We formulated a discrete-time compartment model characterizing dynamic transitions between OUD, treatment, and recovery. We parameterized patient choice of treatment location using travel distance and expected initiation wait times. These choices subsequently impact their likelihood of remaining in treatment and create a feedback loop affecting wait times for other patients. We integrated this compartment model into a mixed-integer program minimizing total overdose deaths under budget and initiation wait time constraints. We applied our model to a case study in Allegheny County, Pennsylvania, using patient-level and state-level publicly available data for calibration. In simulations, our optimal allocation increased the recovery rate to 55.4%, compared to 39.8% under the status quo and 46.9% for policies solely minimizing travel distance. Our findings demonstrate that accounting for patient choice and facility congestion is essential for designing MOUD networks that effectively reduce overdose mortality.

2 - Medication-Adherence-Aware Care Model Allocation for Patients with Atrial Fibrillation via a Restless Multi-Arm Bandit Approach

You Zhou, Purdue University, West Lafayette, IN, United States, Nan Kong

Efficient allocation of limited care resources is critical for improving outcomes in patients with atrial fibrillation (AF) receiving direct oral anticoagulant (DOAC) therapy. In this paper, we study a system-level decision problem to optimally assign care models across a heterogeneous and dynamically evolving population.

We formalize the problem as a continuous-time restless multi-armed bandit (RMAB) problem, where each patient is modeled as a bandit whose state evolves according to individual risk profile, adherence behavioral trajectory and the selected care model. The state space jointly captures health status and

medication adherence, characterized by adherence levels and behavioral trajectories, which directly influence the subsequent clinical risks. We consider a multi-class, dynamic population with patient arrivals and departures, where patients within each class share common transition dynamics. The objective is to minimize the long-run average healthcare costs. Under this framework, we derive index-based policies, including the classical Whittle index policy, and adapt Verloop's priority-based allocation rules to guide resource-constrained decision-making.

This study demonstrates the feasibility of integrating advanced stochastic control methodology into healthcare resource allocation. Building on this foundation, future work will focus on rigorously validating key structural assumptions (e.g., indexability) within the context of adherence-aware care model selection. Comprehensive robustness analyses will be conducted to assess the stability of allocation policies under uncertainty in patient risk profiles, adherence dynamics, and treatment effects. We will benchmark the proposed index-based approach against alternative frameworks, including approximate dynamic programming and reinforcement learning, to evaluate long-term clinical and economic outcomes using real-world data.

3 - Optimal Sensitivity and Specificity Trade-off for Adaptive Testing

Blake Harris, University of Michigan, Ann Arbor, MI, United States, Brian Denton, Viswanath Nagarajan

Recent advancements of \textit{biomarker tests} greatly help clinicians diagnose early signs of numerous diseases, such as cancer. However, these tests can return false-positives and false-negatives leading to a misdiagnosis. In order to reduce the likelihood of these outcomes, multiple tests can be performed to improve the overall true-positive (\textit{sensitivity}) and true-negative (\textit{specificity}) rates. While all such tests can be performed to collect the most information possible, doing so can lead to high costs and certain tests, such as biopsies, can risk the health of the patient. Instead, it might be preferable to design a testing policy which only performs a subset of tests. This reduces costs while still improving overall sensitivity and specificity. There are two challenges in designing such testing policies: 1) choosing such a subset and 2) handling correlations between tests. To solve this problem, we introduce a method using dynamic programming and parametric search to build the ROC (Receiver Operating Characteristic) curve. Notably, our method constructs adaptive testing policies, where the testing sequence depends on prior outcomes. This is in contrast to prior work that only designed composite testing policies where a fixed subset of tests is used. We also show that adaptive testing can be significantly better than composite testing. In particular, there are instances where using adaptive instead of composite testing can increase the AUC (area under the curve) by almost factor two.

RA07

304

AI Agents in Clinical Workflows

Invited Session

eBusiness

Chair: Harang Ju, Johns Hopkins University, Baltimore, MD, United States

1 - Adaptive Screening Under Constraints: A Multi-Agent Reinforcement Learning Framework for Efficient Diagnostic Question Selection

Kanghyun Cho, Indiana University, Bloomington, IN, United States, Salih Tutun, Ali Tosyali, Kazim Topuz, Anol Bhattacherjee

Mental health diagnostic screening often relies on long and fixed questionnaires which can overwhelm patients and reduce their willingness to complete them. We introduce MARVEL (Multi-Agent Reinforcement for Versatile Emergent Learning), a novel framework that uses multi-agent reinforcement learning framework to adaptively select diagnostic questions. The system employs a hybrid architecture, where training

occurs centrally and execution is decentralized. This design facilitates scalable and contextually relevant question prioritization, which is further enhanced by domain-specific expert agents and a graphical representation of symptoms. In simulation using the Symptom Checklist-90-Revised (SCL-90-R), MARVEL reduced the question burden by 52% while maintaining 89.3% diagnostic accuracy. This framework offers a generalizable approach for efficient and adaptive screening in mental health and other domains.

2 - AI Confidence Heuristics and Human Trust Building: the Case of Medical Delivery

Wanqin Chen, University of Rochester , Rochester, NY, United States, xiaodong li, Ritu Agarwal, Weiguang Wang

This study investigates how AI confidence calibration affects user trust and performance in healthcare decision-making. We experimentally manipulate AI confidence levels (overconfident, underconfident, realistic) while holding accuracy constant. Drawing on confidence heuristic theory, we examine whether strategic confidence calibration can enhance human-AI collaboration effectiveness and operational productivity in clinical settings.

3 - More AI or Better-Placed Ai? Evidence from a Field Experiment on Layered Human-Ai Decision Authority for Postoperative Follow-up in Healthcare

Yichen (Willa) Wang, George Washington University, Washington, DC, United States, Sunghun Chung, Wendy Duan

Artificial intelligence (AI) in healthcare is shifting from improving prediction accuracy to embedding AI into end-to-end clinical workflows, where predictions must translate into timely follow-up actions, patient communication, and escalation decisions. Thus, the central question is not only whether AI predicts well, but how decision authority should be allocated between AI and clinicians across three sequential stages of postoperative care: Trigger (deciding follow-up is needed), Initiation (making first contact with the patient), and Escalation (determining next steps based on the patient's response).

To address this question, we partner with a refractive surgery department in Asia to conduct a field experiment among postoperative patients requiring continuous monitoring. The system integrates machine learning, retrieval-augmented generation (RAG), and patient-specific communication to support real-time follow-up. We compare five human–AI delegation structures: human-only control (T0), AI-trigger only (T1), AI-trigger plus AI-escalation (T2), AI-trigger plus AI-initiation (T3), and full automation (T4). We evaluate patient adherence and clinician workload, expecting T2, where AI handles Trigger and Escalation while clinicians retain Initiation, to outperform both full automation and fully human workflows, with effects varying by patient risk and AI aversion.

This study contributes by developing and testing a multilayer human–AI delegation framework, showing that effective automation depends on aligning authority with the relational sensitivity and error costs of each stage, not accuracy alone. It also offers a reusable experimental template and practical guidance for scaling postoperative follow-up without increasing clinician burden or undermining patient trust.

RA09

305B

Healthcare Operations Under Uncertainty: Scheduling, Emergency Care, and Patient Flow

Contributed Session

Medical Decision Making

Chair: Fu Enqi, University of St. Gallen, St. Gallen, N/A, Switzerland

1 - A Data-Driven Approach for Integrated Operating Room Scheduling and Bed Management

Nalan Gulpinar, Durham University, Durham, United Kingdom, Dimitrios Karagiannis, Vinh Doan

Operating room scheduling that accounts for downstream unit capacity poses a significant challenge for hospital management. This study integrates machine learning–based predictions with decision making process to improve scheduling decisions and resource utilization. We develop a stochastic optimization model that addresses uncertainty in initial ward bed occupancy at the start of a surgical week and evaluate its performance using real hospital data. By incorporating personalized patient length-of-stay predictions generated through machine learning, the approach reduces variability in bed usage, leading to fewer last-minute surgery cancellations and smoother ward occupancy profiles. Overall, the results demonstrate improved scheduling efficiency and more effective ward management.

2 - Distributionally Robust Balance of Unused TIME and Time-Delivery Guarantee in Tactical Block Scheduling

Saharnaz Mehrani, Florida Atlantic University, Boca Raton, FL, United States

Hospitals often assign surgeons recurring operating room (OR) time blocks for elective surgeries through a tactical block schedule that remains fixed for several months. These recurring assignments provide surgeons with predictable access to OR time and related hospital resources, but they must be determined before the detailed case mix within each block is known. A key challenge in this setting is uncertainty in each surgeon's total time demand over the recursion period. Repeated under-delivery of OR time may reduce surgeon satisfaction and, in competitive markets, can contribute to the loss of surgeons to other hospitals, whereas over-allocation leads to unused block time and inefficient use of costly OR capacity. To address this trade-off, we develop a tactical OR block scheduling model that minimizes expected unused block time while enforcing surgeon-level time-delivery requirements. To account for limited historical data and ambiguity in surgeon-specific demand distributions, we further develop a data-driven distributionally robust optimization framework based on Wasserstein ambiguity sets. We derive tractable mixed-integer linear programming reformulations for the resulting models and establish structural properties that characterize optimal solutions and support computational efficiency. Using data from a partner hospital, we show that the proposed framework provides transparent control over the trade-off between block utilization and surgeon-level time delivery, while incorporating several practical scheduling features.

3 - Queue Management for the Assignment of Anesthesia Clinician Breaks to Increase Fractions Completed Before the End of Surgical Closure and During Planned TIME Windows

Franklin Dexter, University of Iowa, Iowa City, IA, United States, Sarah Titler, Richard Epstein

At surgical suites with long workdays, anesthesia clinicians typically receive lunch breaks. We estimated the percentage relative impact of decision-making processes on 30-minute breaks completed during cases' surgical periods and during two-hour windows (e.g., 11-1).

Discrete-event simulations of breaks were performed using a retrospective cohort of a large teaching hospital's 15 years of actual dates (N=5,481 days), operating rooms (N=53), surgical times (N=460,354), and scheduled procedures. Simulations prioritizing cases with the least predicted time left used all historical data for each case from its date, calculating the probability that the log-normally distributed case finishes later than the end time of the break, conditional on the time since the case started. Simulation analyses were performed pairwise across queue disciplines using batches of one-week periods, sufficient to remove residual autocorrelation; all SE <0.02%

Giving breaks preferentially to rooms with the least (but sufficient) expected time left resulted in 16.5% fewer complete breaks vs. preference to longest ongoing cases. Single, pooled queues with preference for the longest ongoing cases resulted in overall 7.2% more breaks. Using a single queue and preference to cases ongoing the longest achieved 28.4% more breaks than assigning each clinician to serve individual rooms and prioritizing cases with the least predicted time left to the end of surgery.

For each clinician in a suite giving breaks, assign them to a room to break. While the first set of breaks is being completed, choose the next set of rooms for breaks with preference to rooms ongoing the longest.

4 - Coordinated Ambulance Routing Decisions Between Emergency and Urgent Care Under Imperfect Triage by EMS Personnel

Marco Bijvank, University of Calgary, Calgary, AB, Canada, Mahdi Shakeri

Emergency Medical Services (EMS) provide urgent pre-hospital care, but ambulance offload delays occur when Emergency Departments (EDs) are congested, which is delaying patient transfers and reducing ambulance availability. Over 40% of high-acuity ED arrivals experience delays exceeding 30 minutes, negatively affecting patient care and safety. Many patients, however, have non-severe conditions and could be treated at Urgent Care Centers (UCCs), which have lower congestion levels. EMS personnel currently triage patients to predict whether ED or UCC care is appropriate. A centralized controller then assigns patients based on predictions and facility congestion.

This study proposes a machine learning (ML)-based, error-aware dynamic allocation system to improve routing decisions. The problem is modeled as a Markov decision process, considering misclassification risks: false negatives (ED-needed patients sent to UCC) and false positives (UCC-eligible patients sent to ED). Two approaches handle misroutes: Explicit Misclassification Penalty (EMP) and Reroute on Error (RER). The ML classifier's accuracy depends on effort, and its threshold affects false-positive and false-negative rates, indirectly influencing congestion and misrouting. A joint optimization framework determines the optimal effort and threshold to minimize total system costs, including congestion and misrouting penalties.

Numerical analysis reveals: (i) higher false-positive rates make policies more conservative in sending patients to UCCs, (ii) policy is more sensitive to false positives than false negatives, (iii) RER is more aggressive than EMP, and (iv) investing in classifier accuracy is always beneficial, even with high learning costs. Overall, effective patient flow requires integrating triage accuracy and routing, not just expanding capacity.

5 - How Hospital Market Liberalization Reshapes Patient Mobility, Hospital Operation and Clinical Safety

Enqi Fu, University of St. Gallen, St. Gallen, Switzerland, Alexander Geissler

We investigate how hospital market liberalization affects patient mobility, hospital operations, and clinical safety in planned care. In 2012, Switzerland introduced a nationwide hospital reform that implemented SwissDRG payment, coordinated regional capacity planning, and expanded patients' statutory right to receive elective inpatient care outside their home canton. Using the Swiss Federal Statistical Office Medical Statistics covering more than 9.8 million planned admissions from 2010–2023, with a primary analysis of 7.2 million admissions from 2010–2019, we apply a high-dimensional two-way fixed-effects difference-in-differences design comparing patients in border-adjacent (treatment) and interior (control) residence regions (n = 647,269 and 800,475 before the reform; n = 2,565,059 and 3,204,478 after the reform).
The reform immediately and persistently increased patient mobility. Patient outflow rate increased by 3.6% relative to the baseline of 27.1%, and average driving time increased by 3.1% relative to the baseline of 27.9 minutes for patients living in border regions. Hospitals responded by improving bed throughput, average length of stay declined by approximately 2%, while ICU admission rates decreased slightly (0.1% relative but insignificantly). With controlling the effect of outflow rate and driving time, ICU usage hours and ventilation usage hours increased by 0.6% and 8.3% relative, suggesting a reallocation of resources toward more complex cases. Importantly, inpatient mortality remained unchanged, indicating that greater operational efficiency was not achieved at the cost of clinical safety. Overall, our findings suggest that hospital market liberalization can increase patient mobility and improve operational efficiency without compromising clinical safety.

RA10

306A

Topics on Mental Health and Postpartum Care

Contributed Session

Public Health

Chair: Marco Patino-Murillo, North Dakota State University, Fargo, ND, United States

1 - Improving Access to Electroconvulsive Therapy: a Discrete-Event Simulation Workflow Optimization in Behavioral Health

Liam Coen, Northwell Health, New Hyde Park, NY, United States, Sally Chin, Sohag Sanghani, Shaghayegh Norouzzadeh

Background:
Electroconvulsive therapy (ECT) is a highly effective, rapid-acting treatment for severe psychiatric conditions, often succeeding when medication fails. However, operational bottlenecks limit timely patient access. We analyzed a psychiatric hospital ECT suite serving inpatient and outpatient populations where workflow inefficiencies constrained throughput capacity.

Objective:
To reduce patient turnaround time, eliminate peak-hour constraints, and expand capacity by 5% without capital investment through workflow optimization.

Method:
Using workflow mapping, time studies, and statistical analysis as inputs for a discrete-event-simulation (DES) model, we evaluated staff utilization, treatment room capacity, and patient-flow bottlenecks. The validated model captured decision-making processes including patient routing, resource allocation, and scheduling patterns.

Results:
Baseline turnaround time from Check-In to Exit averaged 102 minutes, with peak-period (10am-12pm) utilization imbalances limiting throughput. Key findings: (1) RN handoffs between preassessment and treatment added 3 minutes per patient, (2) anesthesiologist availability constraints during peak delayed treatment, (3) concentrated scheduling created artificial constraints despite 69% room utilization (most peak appointments were controllable inpatients), and (4) RN direct patient-care utilization reached 70% at peak. Modeling demonstrated that—eliminating RN handoffs, dedicating anesthesiologist time to clinical care during peak (30% wait reduction), and redistributing inpatient appointments—would enable 5% volume growth while maintaining baseline performance.

Conclusions:
This is the first DES application to optimize throughput in outpatient interventional psychiatry. Modeling results indicate that streamlining workflows, optimizing staff allocation, and redistributing scheduling can expand patient access by 5%. These strategies extend DES applications to procedurally intensive behavioral health settings and are generalizable to other procedure-based services.

2 - Risk Factors for Postpartum Depression and Anxiety: Implications for Early Detection and Resource Allocation

Vianka Lazaro Vazquez, North Dakota State University, Fargo, ND, United States, Diana Lopez Soto, Shannon Harriger

Postpartum Depression (PPD) and anxiety (PPA) are prevalent conditions, affecting 1 in 7 mothers in the U. S. Beyond their clinical impact, perinatal mood disorders impose substantial economic burden; recent estimates suggest that untreated maternal mental health conditions cost the U.S. healthcare system and society over $14 billion annually. Despite their significant consequences, these conditions are frequently under diagnosed, which prevents timely and effective intervention. This highlights the need for data-driven methods to categorize individuals by risk level, ensuring that health policies, diagnosis procedures, and treatments are based on solid evidence.

This study aims to identify clinical and demographic predictors associated with elevated PPD and PPA risk . A retrospective chart review was conducted using data from 1,696 patients, who were assessed via the Edinburgh Postnatal Depression Scale (EPDS) at two intervals: 24 hours and 6 weeks postpartum. Following hospital protocols, scores greater than 12 were classified as positive for PPD, while scores between 9 and 12 were categorized as positive for PPA.

Preliminary findings suggest that limited partner support, delivery complications, and specific socioeconomic indicators are associated with an increased risk of PPD and symptom progression. However, the impact of these factors varies according to age, marital status, and existing support system, among others. Ultimately, this study aims to provide decision-makers with actionable insights early detection such as targeted screening protocols, efficient resource allocation, and personalized interventions.

3 - Maternal Morbidity Across Bmi Categories: Identifying Patterns of Risk Beyond Obesity for Equitable Interventions

Marco Patino-Murillo, North Dakota State University, Fargo, ND, United States, Diana Lopez-Soto, Ph.D.

During pregnancy, women are often subject to a variety of medical conditions, whether pre-existing or arising during gestation, that significantly complicate the maternal experience. Current estimations indicate that maternal morbidity costs exceed 32 billion dollars annually. Conditions such as preeclampsia, gestational diabetes, hemorrhage, and preterm birth represent the most frequent complications driving these escalating expenses. While obesity is one of the factors most strongly associated with these outcomes, with studies indicating that it raises maternity care expenses by 23 to 37 percent, gestational complications are certainly not exclusive to this demographic. A significant number of women within a normal body mass index range also suffer from severe complications.

In this context, labeling obesity as the primary cause of poor outcomes may overshadow meaningful differences among various at-risk patients. Such a narrow focus limits the ability of healthcare providers to design effective and targeted interventions for the broader population. This study seeks to identify complex patterns regarding maternal complications among pregnant women by analyzing the intersection of clinical, behavioral, and demographic factors. By evaluating how these diverse variables influence health outcomes, the research provides decision makers with a clear picture of the risks individuals face. This comprehensive approach enables the development of more precise, proactive, and equitable healthcare strategies for all expectant mothers.

RA11

306B

Secure and Scalable Deployment of AI and Digital Systems in Healthcare

Contributed Session

Healthcare Technology

Chair: Sai Teja Rayabarapu

1 - Operationalizing Cybersecurity in Healthcare Systems: a Risk-Based Framework for Secure Medical Device Integration and Patient Data Operations

Shekar Munirathnam, IEEE, Cumming, GA, United States

Healthcare systems face unprecedented operational challenges in managing the convergence of medical devices, electronic health records, and interconnected clinical systems while maintaining patient safety and data security. This research presents a comprehensive operations management framework that integrates cybersecurity risk assessment into healthcare workflow optimization, focusing on the operational lifecycle of connected medical devices and health information systems.

We propose a quantitative model that balances operational efficiency metrics (device uptime, patient throughput, clinical workflow continuity) with security posture indicators (vulnerability remediation rates, identity federation success, threat detection latency). Drawing from real-world implementation at enterprise healthcare environments, our framework addresses three critical operational domains: (1) secure provisioning and lifecycle management of IoT-enabled medical devices, (2) identity and access management optimization for multi-stakeholder healthcare ecosystems, and (3) incident response coordination that minimizes clinical service disruption.

The methodology employs fuzzing-based validation techniques to ensure medical device reliability under adverse conditions, coupled with KPI-driven security operations that align with HIPAA compliance requirements and emerging EU Cyber Resilience Act (CRA) mandates. Case studies demonstrate measurable improvements in both operational resilience and security outcomes, including 40% reduction in security-related operational downtime and 60% improvement in credential-based access efficiency across federated clinical systems.

This work contributes practical, implementable strategies for healthcare operations managers seeking to enhance system security without compromising clinical operational effectiveness.

2 - A Piecewise Affine Architecture for Regional Value Function Approximation in Approximate Dynamic Programming

Jonathan Patrick, University of Ottawa, Ottawa, ON, Canada, Shahryar Moradi, Antoine Sauré

Motivated by the shortcomings of using a single affine approximation to represent the value function for problems with complex dynamics, we propose a novel approximation approach that adaptively partitions the state space into different regions and fits a local affine approximation to each. Unlike traditional piecewise approaches that rely on predefined regions, our method dynamically learns partitioning breakpoints and refines them within an approximate dynamic programming framework that combines simulation and linear programming. The method focuses on regions where the value function exhibits significant non-linearities or a non-monotonic behavior. We demonstrate the effectiveness of the proposed approach through applications in queueing control, battery management, and multi-priority patient scheduling. Our approach provides a better approximation of the value function compared to single affine approximations, resulting in improved performance while maintaining interpretability and computational efficiency.

3 - Measuring and Governing AI Impact Under Real-World Constraints: Lessons from Non-Randomized Deployments

Aman Tripathi, Hyatt Hotels Corporation, Sugar Land, TX, United States

Organizations are increasingly deploying AI-driven interventions in production environments, yet measuring their impact remains challenging due to constraints such as non-randomized rollouts, limited control groups and high variance outcome metrics. These challenges are particularly relevant in domains such as healthcare, where controlled experimentation is often constrained.

In this talk, we draw from large-scale deployments of AI-enabled prioritization systems for high-value decisions, where traditional experimental design was not feasible due to phased rollout and region-specific implementation constraints. These challenges were consistently observed across multiple AI deployments involving high variance business metrics and non-randomized rollouts.

A key challenge was the reliance on high variance business KPIs, including revenue-based metrics, which limited statistical sensitivity despite extended test durations and observable directional trends. Baseline differences between test and control groups further complicated interpretation.

We present a practical measurement and governance framework that:

1. **Distinguishes causal impact from directional evidence.**

2. Integrates lower variance leading indicators with outcomes metrics.
3. Establishing guardrails for interpreting results under limited statistical power.

This framework enables consistent and transparent decision-making even when statistical significance is difficult to achieve. While illustrate through operational deployments, the approach generalizes to healthcare and other settings where outcome variability is high and randomized experimentation is limited.

This work is intended as a practitioner-focused talk providing actionable guidance for evaluating and governing AI systems under real world constraints.

4 - Advancing Learning Health Systems via One Digital Health and Multilingual Semantic Interoperability

Arriel Benis, Department of Biomedical Engineering, Duke University, Durham, NC, United States, Gaia Benis, Stefan Darmoni, Oscar Tamburis

Learning Health Systems (LHS) necessitate a continuous cycle of data-to-knowledge transformation, requiring high-fidelity, semantically enriched data streams to drive clinical and operational improvements. To address fragmented terminologies, we developed ODH-UT (One Digital Health-Unified Terminology), a semantic bridge between the One Digital Health (ODH) architecture and the Medical Informatics and Digital Health Multilingual Ontology (MIMO).

The ODH framework provides the systemic foundation, organized into two enabling keys (One Health and Digital Health), three perspectives (Individual Health, Population/Society, and Ecosystem), and five dimensions (Education, Environment, Human/Veterinary Healthcare, Healthcare Industry 4.0, and Citizen's Engagement), all unified within a Technology ring.

By extending the LHS data perimeter beyond Electronic Health Records to include heterogeneous population-level determinants and environmental streams, ODH-UT enables a holistic approach to system optimization. This is operationalized through MIMO, integrated within the HeTOP portal, offering over 4,000 concepts in more than 30 languages. Critically, ODH-UT aims to ensure that multidisciplinary datasets adhere to FAIR data principles, specifically enhancing Interoperability and Reusability.

This structural alignment minimizes feature ambiguity in predictive models, driving sustainability by reducing resource duplication. By facilitating international harmonization, ODH-UT strengthens global health surveillance. Ultimately, the convergence of ODH and MIMO via ODH-UT provides a scalable infrastructure for actionable knowledge discovery, fostering a resilient, FAIR-compliant ecosystem for data-driven healthcare innovation worldwide.

5 - From AI Extraction to Trusted Answers: Using Knowledge Graphs and Ontologies to Power Explainable Ai in Healthcare

RAMA KRISHNA KUMAR Lingamgunta, Cigna Evernorth Services Inc, Raleigh, NC, United States

Healthcare organizations increasingly rely on AI to generate insights and support operational decisions, yet many struggle to trust and operationalize these outputs due to fragmented data, inconsistent terminology, and limited explainability. AI systems that operate on disconnected datasets often fail to capture context, relationships, and governance constraints required for reliable decision-making at enterprise scale.

This session examines how AI-assisted ontology creation and knowledge graph construction can transform structured and unstructured healthcare data into a governed intelligence layer for explainable AI. Attendees will learn how AI can accelerate the extraction of concepts and relationships from documents, policies, and operational data, while human-in-the-loop validation ensures accuracy, contextual integrity,

and governance alignment. The session highlights how ontologies standardize meaning across systems and how knowledge graphs preserve relationships needed for traceable, auditable reasoning.

The session also demonstrates how enterprise knowledge graphs can be consumed operationally by AI agents and chatbots to deliver trusted answers across workflows such as analytics, compliance, and decision support. Emphasis is placed on translating governed knowledge into measurable outcomes, including improved efficiency, reduced manual reconciliation, and increased confidence in AI-driven decisions.

6 - Automating Healthcare Operations Without Compromising Compliance

SAI TEJA RAYABARAPU, Health Care Service Corporation, Richardson, TX, United States

Healthcare automation is often positioned as a solution to rising costs and operational inefficiencies, yet many organizations hesitate due to concerns around compliance, regulatory risk, and patient safety. This session explores how healthcare operations can be automated responsibly—without compromising regulatory integrity—by focusing on process design, governance, and evidence-based decision-making.

Grounded in research on U.S. healthcare affordability and administrative inefficiencies, this talk demonstrates how excessive manual workflows, insurance complexity, and fragmented processes contribute significantly to healthcare cost inflation. Using insights from decentralized care models and AI-driven administrative automation research, the session presents a structured approach to identifying automation opportunities that reduce cost while strengthening compliance.

Attendees will gain practical insight into how business analysts and project leaders evaluate automation readiness, design compliant workflows, and embed regulatory safeguards into AI-enabled processes. The session highlights how automation can streamline administrative operations such as intake, scheduling, documentation, and billing—reducing overhead while maintaining transparency and auditability.

Rather than focusing on automation as a purely technical initiative, this session reframes it as a compliance-aware transformation strategy. Participants will leave with a clear understanding of how healthcare organizations can adopt AI and automation in a way that improves efficiency, protects regulatory requirements, and supports long-term sustainability.

RA12

306C

Empirical Healthcare

Invited Session

MSOM: Healthcare

Chair: Xinyu Liang, INSEAD, Singapore, Singapore

1 - *CANCELED Disparities in Maternal Care and Operational Factors: Role of Specialization, Exposure, and Ownership in Eliminating the Gap

Max Yakovlev, Vanderbilt University, Nashville, TN, United States, Maria Ibanez

Problem definition: Annually, approximately 60,000 U.S. women experience *severe maternal morbidity* (SMM)—unexpected labor and delivery outcomes with serious short- or long-term health impacts. Many of these outcomes are preventable with consistent clinical decision-making. However, disparities in obstetric care and outcomes persist, and Black mothers consistently experience higher odds of labor complications. In this paper, we examine how hospital- and physician-level operational factors moderate the disparity for Black mothers relative to other mothers in the likelihood of (1) riskier delivery procedures (e.g., first-time unscheduled Cesarean Section) and (2) resulting 90-day readmission rates

following these procedures. Methodology/Results: Using cross-sectional data of all labor and delivery admissions across six U.S. states, we uncover operational factors mitigating these disparities: stronger hospital-level *C-section specialization* reduces the gap, as does greater hospital- and physician-level *exposure* to Black patients. Notably, hospital *ownership* also impacts disparities: giving birth at a non-profit hospital entails relatively lower odds of receiving a first-time unscheduled C-Section for a Black mother, whereas receiving treatment at a government-owned facility corresponds to higher odds of experiencing riskier procedures and 90-day readmissions. Finally, using text analysis techniques, we find that patients treated at hospitals using more *process-oriented* mission language face smaller disparities in first-time unscheduled C-section use compared to patients seen at hospitals with *patient-oriented* missions. Managerial implications: These findings highlight the importance of more detailed quality information disclosure, and more decision and behavioral support for clinicians. Such measures are crucial to promoting consistent clinical decision-making and improving health outcomes for all patients.

2 - Physician Networks in Accountable Care Organizations

Dimitrios Andritsos, HEC Paris, Paris, France, Claire Senot

Effectively coordinating a patient's health care services is critical for the delivery of high-quality and efficient care. Accountable Care Organizations (ACOs) are a prominent model for incentivizing more coordinated care delivery. Nevertheless, no consistent approach has emerged in terms of how different ACOs organize and coordinate care between participating providers. In aggregate, the evidence to date suggests that ACOs' success in achieving their stated objectives of lowering costs and improving quality has been modest. However, this hides the fact that there exists a wide disparity of performance across ACOs. Thus, they provide an excellent empirical setting for the study of care coordination. In this work, we exploit this variability to understand how the nature of care coordination across ACO-centered physician referral networks is associated with performance across both cost and quality dimensions. In our analysis, we put a particular emphasis on coordination between primary care providers and specialists. To characterize communication and care coordination among physicians we examine their patient-sharing relationships via network analysis.

3 - Incentive Contracts and Operational Design: Evidence from Bundled Payments

Anand Bhatia, IE Business School, Madrid, Spain, Jayashankar Swaminathan

Incentive contracts are widely used to improve performance, yet little is known about how they reshape internal operations. In healthcare, bundled payment models shift financial risk to providers, and while often linked to savings through changes in post-acute care (PAC), their effect on inpatient operational design is less understood. This study examines how hospital adoption of the Bundled Payments for Care Improvement (BPCI) program impacts inpatient operations and costs.

Using a quarterly panel dataset from California acute care hospitals, we employ a hospital-episode fixed-effects model to estimate the impact of BPCI enrollment on performance. Our analysis controls for time-invariant hospital-episode differences and secular time trends.

Consistent with prior work, we find BPCI adoption reduces cost per discharge by 3-4%. However, these savings persist after controlling for PAC discharge patterns, pointing to other operational channels. We find that that BPCI adoption significantly increases inpatient care standardization, a measure of consistency of health service delivery, by 5.3%. This increase in standardization is strongly associated with reduced episode costs. Conversely, average length of stay modestly increases, suggesting savings are not driven by faster throughput. We find no significant effect on readmission or mortality rates.

This research contributes to the operations management literature by demonstrating that financial incentives induce multi-dimensional operational adjustments. It identifies inpatient process reconfiguration and standardization, not just patient discharge decisions, as a primary mechanism for achieving cost savings in complex service systems like hospitals.

4 - The Bias Toward Action: Evidence from Asymmetric Physician Responses to Clinical Signals

Xinyu Liang, INSEAD, Singapore, Singapore, Srishti Arora, Michael Freeman

Knowing when to stop care is as important as knowing when to start it. In clinical settings, the harder decision is often not whether to act, but whether to discontinue follow-up when clinical signals suggest that no further surveillance is warranted. We study physician behavior in test ordering at a large hospital from 2022 to 2024 by deploying a local LLM to recover the clinical signal from ultrasound reports and map them to the hospital's follow-up protocol. We find evidence of action bias in physician ordering: physicians frequently continue surveillance even when the prior scan suggests stopping. Among overuse cases, over 70% lack documented clinical justification, and less than 50% yield no meaningful new findings. We also find that continuity and operational conditions further amplify unnecessary surveillance. Our study provides insights into the design of decision support and workflow systems that mitigate action bias in diagnostic surveillance.

Thursday, July 30, 2026, 9:45 AM - 11:00 AM

RB01

301A

Opening the Black Box: Interpretable Multimodal AI in Healthcare

Invited Session

AI in Healthcare

Chair: Kimberly Villalobos Carballo, New York University, New York, NY, United States

1 - Feature Selection for Machine Learning with Unstructured Data: Applications in Healthcare

Kimberly Villalobos Carballo, New York University, New York, NY, United States

Modern machine learning models trained on unstructured data—such as clinical text—have achieved strong predictive performance in healthcare tasks ranging from diagnosis to outcome prediction. However, these models remain fundamentally limited in their ability to provide faithful interpretations of their outputs, as they cannot reliably explain why a prediction was made or which information was most influential. This motivates revisiting the classical feature selection problem in the context of unstructured language data: identifying small, informative subsets of the input (e.g., sentences) that are sufficient to drive model predictions.

In this talk, we formulate feature selection as a learning problem jointly optimized with the predictive task, leading to a maximum-likelihood objective over both the predictive model and the selection function. While the resulting combinatorial problem is intractable, we introduce a greedy sequential approximation that extends forward feature selection methods from tabular settings to unstructured text, enabling feature selection while remaining computationally tractable.

Empirical results on real-world healthcare datasets demonstrate the effectiveness of the approach. For end-of-stay mortality, next-day discharge, and coronary heart disease risk prediction using clinical text, the method produces interpretable models that rely on substantially fewer sentences per patient while achieving predictive performance comparable to models trained on the full text.

2 - Interpretable Multimodal AI for Clinically Actionable Decision Support

Ignacio Fuentes, MIT Jameel Clinic, Cambridge, MA, United States

Recent advances in multimodal foundation models have created new opportunities to integrate imaging, clinical notes, structured EHR data, genomics, and wearable signals into unified predictive systems. Yet in healthcare, predictive performance alone is insufficient, models must be interpretable, reliable, and aligned with how decisions are actually made in clinical practice.

In this talk, I will present recent work from the MIT Jameel Clinic on the development and deployment of interpretable multimodal AI systems designed to support real-world clinical decision-making. Rather than treating interpretability as a post hoc diagnostic, our approach integrates transparency directly into model design and evaluation. We focus on three complementary directions: (1) architectures that preserve modality-specific interpretability while enabling cross-modal reasoning; (2) methods for uncertainty quantification and subgroup analysis to assess reliability, safety, and equity; and (3) evaluation frameworks that prioritize clinical utility, including calibration, actionability, and integration into care delivery workflows.

Through case studies in medical imaging and longitudinal risk prediction, I will highlight how these systems can move beyond retrospective prediction toward actionable decision support. I will also discuss lessons from deployment in healthcare settings, where constraints such as workflow integration, user trust, and operational considerations fundamentally shape model impact.

This perspective shifts the goal of interpretable AI from explaining predictions to enabling decisions, advancing multimodal systems that are not only accurate, but trustworthy and clinically actionable.

3 - Xhaim: Improved Performance and Explainability

Vasiliki Stoumpou, Massachusetts Institute of Technology, Cambridge, MA, United States, Periklis Petridis, George Margaritis, Dimitris Bertsimas

With the increasing interest in deploying Artificial Intelligence in medicine, we previously introduced HAIM (Holistic AI in Medicine), a framework that fuses multimodal data to solve downstream clinical tasks. However, HAIM uses data in a task-agnostic manner and lacks explainability. To address these limitations, we introduce xHAIM (Explainable HAIM), a novel framework leveraging Generative AI to enhance both prediction and explainability through four structured steps: (1) automatically identifying task-relevant patient data across modalities, (2) generating comprehensive patient summaries, (3) using these summaries for improved predictive modeling, and (4) providing clinical explanations by linking predictions to patient-specific medical knowledge. Evaluated on the HAIM-MIMIC-MM dataset, xHAIM improves average AUC from 79.9% to 91.3% across chest pathology and operative tasks. Importantly, xHAIM transforms AI from a black-box predictor into an explainable decision support system, enabling clinicians to trace predictions back to relevant patient data, bridging AI advancements with clinical utility.

4 - Use of Multi-Modality Classification/Prediction Modeling in Chronic Pain

Catherine Chong, Mayo Clinic, Phoenix, AZ, United States

Pain chronification or recovery from chronic pain remains challenging to predict. My research focuses on the development of interpretable multimodal artificial intelligence (AI) frameworks that integrate longitudinal neuroimaging and clinical data to improve classification, prediction, and mechanistic understanding of headache pain trajectories. Using large-scale, multi-site datasets, we have applied deep learning and advanced machine learning techniques to identify imaging-based biomarkers and patient subgroups associated with differential outcomes (migraine vs post-traumatic headache vs healthy controls). Structural MRI-based models enable automated headache classification and brain-age estimation, while functional imaging and graph-based approaches reveal dynamic pain-related network alterations. Longitudinal modeling further allows prediction of headache trajectories following mild traumatic brain injury, identifying individuals at risk for persistent post-traumatic headache. Importantly, our work emphasizes interpretability enabling identification of brain regions and network features that drive model predictions. These efforts demonstrate that multimodal, interpretable AI can move beyond prediction toward actionable insights, supporting precision medicine approaches for chronic pain by identifying high-risk patients, informing treatment selection, and uncovering underlying neurobiological mechanisms.

RB03

302A

Data- and AI-informed models for Health Decisions

Invited Session

Healthcare Operations Management

Chair: Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States

Co-Chair: Farzin Ahmadi, Johns Hopkins University, Baltimore, MD, United States

1 - Multimodal Time Series and Text LLMs

Kimia Ghobadi, Johns Hopkins University, Baltimore, MD, United States, Felix Parker, Nimeesha Chan, Chi Zhang

Many operations decisions rely on time-series forecasts informed by unstructured context, but conventional pipelines separate perception from prescription and offer little auditability. We introduce a two-stage framework that equips a pretrained language model to reason over numerical sequences. A residual vector-quantized variational autoencoder tokenizes time-series patches into a discrete vocabulary that interleaves with text, enabling mixed-modality prompts. We test this approach across electrocardiogram question answering, context-aware forecasting, and few-shot UCR classification. The result shows improved accuracy over baselines while producing machine-readable rationales suited to downstream optimization and healthcare operations.

2 - The Impact of AI Scribe on Primary Care Workload

Danielle Ripsman, Johns Hopkins University, Baltimore, MD, United States, Ramya Chiplunkar, Kimia Ghobadi

Primary care clinicians have historically spent nearly double the direct patient care time on electronic health record (EHR) documentation, contributing to clinician burnout and reducing the time available for urgent patient needs. Recently, LLM-based AI tools have been dispatched to primary care settings with the goal of reducing clinical burden and improving efficiency. This study aims to quantitatively investigate the impact of such AI medical scribe tools on clinics and clinician work hours across an eighteen-month period.

Using a database of over 20,000 patient encounters, we investigate 23 primary care facilities, employing over 900 providers, around 300 of whom are active scribe users. The study contains a pre-/post-analysis on the AI scribe's impact on visit volumes, documentation hours, and EHR time across the US, with varied facility and provider backgrounds. We examine uptake patterns by facilities and clinicians, providing insights into the benefits and drawbacks of these emerging technologies.

3 - Similarity-Guided Inverse Optimization for Personalized Dietary Recommendations via Food Swapping

Farzin Ahmadi, Towson University, Towson, MD, United States, Felix Parker, Kimia Ghobadi

Dietary adherence to evidence-based guidelines such as DASH remains persistently low in free-living populations despite strong clinical evidence linking diet to blood pressure reduction and chronic disease prevention. A central barrier is that existing recommendation systems either replicate observed behaviors without enforcing nutritional goals, or enforce goals without accounting for individual preferences — producing recommendations that are clinically valid but behaviorally impractical.

We propose SGIO (Similarity-Guided Inverse Optimization), a framework that integrates large language model-based food embeddings into an inverse optimization backbone to generate personalized, item-level dietary swap recommendations. For each food item in an individual's observed diet, SGIO queries a FAISS similarity index built over Qwen3 text embeddings of USDA FoodData Central item descriptions to identify high-substitutability neighbors, then encodes these as switching costs within the Modified Goal-Integrated Inverse Learning (MGIL) optimization framework. The resulting recommendations simultaneously satisfy DASH nutritional constraints, minimize behavioral disruption relative to habitual

intake, and produce human-interpretable food swap sequences across a graduated constraint-tightening pathway controlled by a single parameter r.

Holdout validation on 356 hypertensive NHANES adults and 200 MyFitnessPal users demonstrates that recommendations fall within natural intra-person dietary variation at all four tradeoff iterations ($p<0.001$ in both nutrient and food-item space), with a 3% non-amenable rate on NHANES and 0% on MyFitnessPal. The r parameter, algorithmically bounded at r=4, functions as a graduated behavior change roadmap with direct clinical interpretability. All theoretical guarantees of the MGIL backbone are preserved under the augmented item space.

4 - When Delays Propagate: Endogenous Sliding Windows for Home Care Scheduling

Jing Liu, Johns Hopkins University, Baltimore, MD, United States, Dew-Anne Langcaon, Kimia Ghobadi

In home care scheduling, caregiver delays propagate across tasks, and a late arrival cascades through the day, causing realizations to deviate from expectations. To better mitigate this inherent uncertainty propagation, we develop a Dynamic Bayesian Network to derive a per-visit appointment window (a "sliding window"), whose size is determined endogenously by the schedule rather than fixed by the planner. This framework introduces a tradeoff between reliability and client burden, namely, looser schedules raise reliability but require clients to reserve a longer block, while tighter schedules reduce idle/burden cost but amplify delay propagation and expose clients to late arrivals. We capture the trade-off between caregivers and clients in our mixed-integer optimization framework, which also includes consistency metrics, task assignment, patient preference, and caregiver availability, and is solved via a decomposition approach. Using real operational data from our partner collaborator (Vivia Cares; a home care provider in Hawaii), our models show reductions of 71 to 15 minutes on the average daily idle time, improved consistency for 98% of clients vs 79% historical, and over 50% reduction in weekly overtime, while substantially reducing delay propagation across visits.

5 - Uncertainty-Aware Risk Score Optimization with Applications in Inpatient Fall Risk Assessment

Emmett Springer, Johns Hopkins University, Baltimore, MD, United States, Fardin Ganjkhanloo, Erik Hoyer, Daniel Young, Kimia Ghobadi

A fundamental challenge in improving clinical risk assessment is the lack of ground truth risk from observational retrospective data. Preventative interventions effectively reduce rates of adverse events but obscure the relationship between risk factors and outcomes. In this work, we apply interpretable and black-box machine learning models to recalibrate an existing risk assessment model, the Johns Hopkins Fall Risk Assessment Tool (JHFRAT), to more accurately reflect latent fall risk. Our interpretable constrained score optimization (CSO) model preserves the JHFRAT structure and is tuned to consider clinical and resource-based constraints. We employ two approaches for assessing this latent fall risk from retrospective clinical data. First, we apply an intervention-based proxy label which categorizes hospital encounters according to their consistent reception of targeted fall prevention interventions, or lack thereof. Secondly, we apply a fall occurrence-based proxy label which considers all encounters with one or more fall to be high risk, and matches by targeted interventions to identify patients who did not fall as low risk. Since neither type of proxy label truly captures latent fall risk, we perform a thorough ablation analysis on cohorts derived from a hybrid of the two proxy label approaches. The ablation analysis results then form the basis for a proposed range of tuned JHFRAT score coefficients for each risk item which are robust to uncertainty in true underlying fall risk. Our data-driven, uncertainty-aware approach to risk assessment optimization provides a foundation for future research and development of trustworthy and explainable risk assessment decision support tools in healthcare.

6 - The Shape of You: Incorporating Patient Geometry and Controlling Extreme Dosage in Radiation Treatment Planning

Angela Morrison, University of Calgary, Calgary, AB, Canada, Amanda Monegat, Bita Sadeghi, Sarah Weppler, Yuriy Zinchenko

Cancer treatment has advanced significantly over recent decades, with improvements in technology enabling radiation therapies that more precisely target tumors while sparing healthy tissue. One such approach is Intensity-Modulated Radiation Therapy or Intensity-Modulated Proton Therapy, which commonly relies on inverse planning. In inverse planning, clinicians specify acceptable radiation limits for healthy tissues while ensuring a sufficiently high dose to the tumor, and a computer-generated plan is produced to meet these criteria.

In this work, we consider a key inverse-planning metric, the Dose-Volume Histogram (DVH). By interpreting the DVH as a probability distribution, we propose a framework to assess worst-case possible deviations from a reference DVH when generalized equivalent uniform dose constraints are imposed. We present several optimization models for solving the associated bounded moment problem, alongside closed-form bounds derived for four specific sequences of moments. To mitigate overly pessimistic outcomes associated with these formulations, we then introduce a mixed-integer programming approach that incorporates patient-specific anatomical and dosimetric information through a dose-deposition matrix. By embedding geometric considerations directly into the optimization model, the resulting DVH approximations achieve greater clinical relevance.

Along with the incorporation of patient geometry, we also investigate strategies for controlling high-dose regions for organs at risk, comparing mean-tail dose and Conditional Value-at-Risk as two complementary approaches. Preliminary numerical experiments demonstrate how our formulations perform and highlight the tradeoffs between tractability and conservatism. The findings contribute to a clearer understanding of the effectiveness of moment-based DVH approximations in the context of radiation therapy planning.

RB04

302B

Artificial Intelligence Across the Healthcare Ecosystem

Invited Session

AI in Healthcare

Chair: Arlen Dean, Washington University in St. Louis, St. Louis, MO, United States

1 - A Two-Stage Predictive Framework for Next-Day Discharge in Hospital Operations

Sogand Soghrati Ghasbeh, University of Michigan, Ann Arbor, MI, United States, Arlen Dean, Mark P. Van Oyen

Timely discharge planning plays an important role in improving hospital throughput, reducing avoidable delays, and supporting effective bed and capacity management. As healthcare systems increasingly adopt artificial intelligence and predictive analytics to support operational decision-making, next-day discharge prediction offers a promising opportunity to improve inpatient flow. In this work, we develop a two-stage predictive framework for next-day discharge using routinely collected electronic health record data. First, patients are stratified into subgroups based on their expected remaining length of stay to better capture heterogeneity in hospitalization trajectories. Second, subgroup-specific models are used to dynamically predict next-day discharge using updated clinical and operational features observed throughout a patient's stay. The proposed framework is designed to generate daily patient-level predictions that can support discharge planning and near-term capacity management. We evaluate the approach using retrospective inpatient data from a large hospital setting and assess performance using metrics relevant to real-world implementation. Our results highlight the value of combining patient stratification with dynamic modeling to improve visibility into near-term discharges. More broadly, this work illustrates how AI-based decision support tools can be integrated into hospital operations to strengthen discharge planning and enhance healthcare delivery efficiency.

2 - Optimizing Diuretic Dosing to in Hospitalized Heart Failure Patients: a Machine Learning Approach

Yu Ma, University of Wisconsin Madison, Madison, WI, United States

Heart failure (HF) is a leading cause of hospitalization, and optimal diuretic dosing strategies remain poorly defined. We developed a machine learning model to identify individualized diuretic dosing that maximizes alive discharge in 48 hours. We performed a retrospective study using EHR data from the University of Chicago (development) and University of Wisconsin (validation). A gradient-boosted tree classifier was trained given clinical predictors and diuretic dose (0–20 units). For each patient, the optimal dose was selected as the dose maximizing predicted discharge probability, and the average treatment effect (ATE) was calculated as the mean difference between optimal and observed dosing outcomes in the test set. The cohort included 12,641 patients (40.2% female; mean age 65.7 years) across 21,183 encounters. Model-recommended dosing improved ATE by 28.1% compared to observed dosing among patients within a 5-day stay window, with similar results in a 7-day window (ATE = 0.277). These findings suggest that machine learning-guided diuretic optimization may substantially improve early discharge rates in hospitalized HF patients.

3 - Scalable Approximate Dynamic Programming for Family Genetic Testing

Kanix Wang, University of Cincinnati, Cincinnati, OH, United States, Daniel Adelman

Hereditary genetic testing increasingly relies on multi-gene panels, but sequential family testing becomes computationally intractable as pedigree size and panel width grow. We develop an approximate dynamic programming framework for family panel testing that reformulates the sequential problem as a linear program, decomposes the value function additively by gene, and solves the resulting approximation by row generation with embedded Bayesian-network inference. This reduces the burden of exact dynamic programming while preserving a transparent accuracy–tractability tradeoff.In exact-solvable single-gene benchmarks, the proposed ADP achieves a mean policy gap of 1.85% across 33 test instances. As pedigree complexity and panel width grow, the method delivers substantial computational savings while remaining computationally usable in regimes where exact dynamic programming becomes burdensome. In large-family multi-gene benchmarks, the current best scalable specification attains an 11.72% mean policy gap, indicating that the method already makes these instances computationally tractable, while highlighting a remaining accuracy–tractability tradeoff at larger scales.

RB05

302C

OR Models in Public Health

Invited Session

Public Health Policy

Chair: Ankit Bansal, State University of New York, Binghamton, Binghamton, NY, United States

1 - Uncertainty-Aware Inventory Planning in Drone-Enabled Healthcare Logistics

Pardis Bahmani, University of Missouri - Saint Louis, St Louis, MO, United States, Shakiba Enayati

Rural healthcare systems face persistent challenges in delivering time-sensitive medical supplies due to geographic dispersion, limited infrastructure, and demand uncertainty. Emerging drone technologies offer the potential to improve responsiveness and accessibility; however, their integration into healthcare logistics requires coordinated decision-making across inventory planning and delivery operations.

This study develops a two-stage, multi-period optimization framework that integrates tactical inventory prepositioning with operational routing and scheduling in a drone-enabled, multi-modal delivery system. The model captures uncertainty in demand while incorporating time- and location-dependent energy consumption for drones, reflecting the impact of environmental conditions on operational feasibility. Inventory is prepositioned at the beginning of the planning horizon under a fixed replenishment policy, and delivery decisions adapt across time periods based on realized demand, allowing for inter-period delays and shortage penalties.

Computational experiments, informed by real-world healthcare logistics data, evaluate system performance under varying demand and operational conditions. Results show that drone-enabled delivery improves service responsiveness and reduces delayed and unmet demand for time-sensitive items while operating under energy and capacity constraints. The value of uncertainty is also examined by comparing predictable and volatile demand settings, highlighting its impact on inventory utilization and system performance.

This work provides a foundation for uncertainty-aware healthcare logistics design and motivates future extensions toward jointly optimizing replenishment policies, inventory positioning, and delivery operations in integrated, data-driven decision-support systems.

2 - A Multidimensional Framework for Strategic Peer Matching: Lessons from the West Virginia Opioid Crisis

Jean-Philippe Richard, University of Minnesota, Minneapolis, MN, United States, Bouchra Er-Rabbany, Ankit Bansal, Samuel Workman, Gordon Smith, Henry Brownstein

Local governments frequently enhance performance by adopting successful strategies from peer organizations. Traditionally, this diffusion of innovation is achieved through an organic process driven by geographic proximity, static demographic affinity, or sigmoid (S-curve) adoption patterns. However, these natural pathways often fail to account for the complex multidimensionality of policy effectiveness.
In this talk, we introduce a novel optimization framework designed to systematically identify peer groups for strategic knowledge exchange. We formulate an Integer Programming (IP) model that integrates concepts from Data Envelopment Analysis (DEA) to identify "best-practice" benchmarks. Moving beyond traditional DEA, our scheme endogenously constructs peer groups by optimizing the potential for benchmark improvement over time. The model explicitly balances socio-economic patterns across clusters while strategically allowing for a controlled fraction of non-peer outliers to maintain structural diversity.
We demonstrate the utility of this approach through a case study of county-level governance in West Virginia. Specifically, we apply the IP model to analyze budget allocations aimed at mitigating two intersecting public health and safety crises: violent crime and opioid-related mortality. By clustering counties into optimized learning cohorts, our framework provides a data-driven roadmap for organizations such as the National Association of Counties, offering a scalable tool to convert regional data into actionable, evidence-based peer networks.

3 - Integer Programming Models for Interpretable Clustering with Applications to Managing the Opioid Crisis In West Virginia

Ankit Bansal, State University of New York, Binghamton, Binghamton, NY, United States, Mahdi Saadati, Jean-Philippe Richard, Samuel Workman, Gordon Smith, Henry Brownstein

We develop novel Integer Programming (IP) formulations for interpretable clustering. Our approach follows a two-stage framework. In the first stage, we propose IP models for grid-based clustering, and in the second stage, we introduce IP formulations to merge low-density grids into coherent clusters. We show that both stages admit min-cost network flow interpretations, which provide structural insights and enable the derivation of strong formulations. Building on these insights, we design a Benders decomposition algorithm to efficiently solve large-scale, high-dimensional instances. We evaluate the proposed approach on benchmark datasets commonly used in the interpretable clustering literature and demonstrate its effectiveness relative to existing methods. In addition, we apply our framework to study the temporal impact of county-level government budget allocations on opioid-related outcomes in West Virginia, a state heavily affected by the opioid epidemic. Our results highlight the potential of interpretable optimization-based clustering to support data-driven decision making and inform policy interventions aimed at mitigating the opioid crisis in highly affected regions of the United States.

4 - Mitigating Harm from Disruption in Illicit Fentanyl Supply Networks: A Stochastic Bilevel Approach

Melika Jahan Beikloo, Clemson University, Clemson, SC, United States, Farhatul Janan, Thomas Sharkey, Mary Beth Kurz

Illicit fentanyl supply networks rapidly adapt to enforcement and supply disruption, complicating public-sector decisions about how to reduce overdose harm without unintentionally increasing risk. We develop a scenario-based stochastic bilevel interdiction model to study the tradeoff between disrupting the network and shifting flow toward riskier backup sources with higher potency or greater uncertainty. A public-sector decision maker chooses disruption and harm-reduction actions, while traffickers respond by re-routing shipments to maximize profit. Uncertainty is represented through scenario-dependent drug potency, which determines harm at consumption nodes after routing decisions are made. The leader minimizes expected harm or conditional value-at-risk (CVaR), allowing us to compare average-risk and worst-case-oriented strategies. To solve policy-relevant instances, we reformulate the bilevel model as a mixed-integer linear program using strong duality and linearization. Computational experiments examine how large-scale disruptions, such as restricting or removing a stable supplier, may shift both the amount of flow reaching users and the risk profile of that flow. We then evaluate how targeted harm-reduction activities can be paired with disruption strategies to prevent post-disruption harm from exceeding baseline levels. In particular, we estimate the required budget and the placement of harm reduction to offset unintended increase in overdose risk. We also compare delayed harm reduction responses with coordinated benchmarks that incorporate harm reduction into the optimization structure. This framework provides insights for when enforcement actions require complementary public health investments under uncertainty.

RB06

303

Operations Research for Personalized Healthcare (ORPH)

Invited Session

Medical Decision Making

Chair: Fatemeh Nosrat, University of Florida, Gainesville, FL, United States

1 - Temporally Feathered Radiation Therapy Planning

Aysenur Karagoz, Bilkent University, Bilkent, Turkey, M. Soheil Hemmati, Fatemeh Nosrat, S. Mohammad Hosseinian, Andrew Schaefer

Temporally feathered radiation therapy (TFRT) for head-and-neck cancer (HNC) radiotherapy combines variable-dose daily subplans to increase the rest time of organs-at-risk (OARs) as sought in intensity modulated radiation therapy (IMRT). While the standard TFRT recommends uniform rest time for each OAR, improved toxicity outcomes may be achieved through variable rest time for OARs by incorporating the OARs' variable radiosensitivity profiles.

We develop a decision-making framework to optimize OAR-specific rest times by maximizing overall recovery. The model combines two key components: biologically effective dose based on the linear-quadratic model, and a dynamical recovery model that captures dose-dependent repair over time. The framework allows for heterogeneous radiosensitivity across OARs, reflecting differences in tissue response to radiation. Using this approach, optimized TFRT schedules are compared with standard TFRT and IMRT in terms of end-of-treatment recovery.

The results indicate that accounting for variability in radiosensitivity can improve recovery outcomes for several OARs while maintaining comparable performance for others. When cumulative dose levels are similar, optimized TFRT provides improved recovery through more effective temporal dose modulation. When cumulative dose differs across approaches, trade-offs arise between total dose and recovery; however, the optimized approach consistently improves the worst-case OAR recovery.

Overall, incorporating organ-specific radiosensitivity into TFRT enables adaptive rest-time optimization and provides a more flexible framework for OAR sparing, with the potential to reduce treatment-related toxicity in HNC radiotherapy.

2 - Evaluating the Effectiveness of Optn Regulations in Organ Transplantation

Michael Olabode, University of Minnesota, Minneapolis, MN, United States, Saumya Sinha

The performance of transplant programs in the United States is routinely evaluated by the Organ Procurement & Transplantation Network (OPTN). In 2024, the OPTN introduced a new performance monitoring system to replace the one previously in place. Despite the OPTN's efforts to monitor the effect of these changing metrics, the latter's impact on transplant programs is yet to be fully established. In this paper, we evaluate the effectiveness of OPTN's performance monitoring system on transplant programs using a simulation-based approach. Our simulation design involves three analytic stages —we start by fitting appropriate regression models on real patient data. These models produce estimates (pre-transplant mortality rate, cumulative risk of graft failure and offer acceptance probability) that are then used to simulate patient outcomes. Finally, we apply OPTN metrics to evaluate transplant programs based on their simulated outcomes. Compared to the old metrics used by OPTN which only evaluate post-transplant outcomes, our findings reveal that the new metrics will reduce false positive flags by 45% and produce greater statistical power (sensitivity, specificity, positive and negative predictive value). Additionally, we observe that under the new metrics, programs have a lower probability of being flagged, supporting OPTN's goal of not identifying too many programs. Finally, our analysis suggests that 1-year conditional on 90-day graft survival —is most associated with program flags with very few or no programs flagged for poor waitlist outcomes. This new direction of evaluation of programs shows a significant improvement upon the previous system and is achieving its goals.

3 - Optimal Decision Support Systems for Adaptive Radiation Therapy

Fatemeh Nosrat, University of Florida, Gainesville, FL, United States

Prior work on adaptive organ-at-risk (OAR)-sparing radiation therapy has typically reported outcomes based on fixed-number or fixed-interval re-planning, which represent one-size-fits-all approaches and do not account for the variable progression of individual patients' toxicities. The purpose of this study was to determine the personalized optimal timing of re-planning in adaptive OAR-sparing radiation therapy, considering limited re-planning resources, for patients with head and neck cancer (HNC). A novel Markov decision process (MDP) model was developed to determine optimal timing of re-planning based on the patient's expected toxicity, characterized by normal tissue complication probability (NTCP), for four toxicities. The MDP parameters were derived from a dataset comprising 52 HNC patients treated between 2007 and 2013. Kernel density estimation was used to smooth the sample distributions. Optimal re-planning strategies were obtained when the permissible number of re-plans throughout the treatment was limited to 1, 2, and 3, respectively

RB07

304

Performance and Policy in Healthcare Systems

Invited Session

eBusiness

Chair: Moiz Bhai, University of Arkansas, 2801 South University Ave, Little Rock, AR, 72204, United States

1 - Screens to Smarts: Regularized Apprenticeship Learning with Attention for Inferring Behavioral Pathways Linked to Health Knowledge Gains in a Digital Intervention

Rahul Ladhania, University of Michigan, Ann Arbor, MI, United States, Shreyas Vajjhala, Rema Padman

Digital health interventions are increasingly deployed by organizations seeking to promote healthy behaviors at scale, yet managers lack evidence-based methods for identifying which program components drive measurable outcomes, limiting their ability to optimize, personalize, or justify these investments. We address this gap by analyzing gameplay telemetry from an 11-week school-based digital health intervention deployed across grades in a school in India, linking children's in-game behavioral patterns to measured gains in health knowledge. To recover interpretable behavioral strategies from high-dimensional, small-sample data, we develop an inverse reinforcement learning (IRL) framework augmented with sparsity constraints and an attention mechanism to learn latent gameplay strategies most predictive of learning outcomes, measured by survey-based health knowledge gains. We find that students who consistently engaged with health-aligned mechanics, such as exercising and strategically deploying power-ups, exhibited significantly greater knowledge gains. Empirically, incorporating attention and regularization substantially improves alignment between inferred rewards and knowledge gains compared to the baseline IRL framework, yielding transparent reward structures that highlight the most behaviorally relevant gameplay features. These findings translate directly into operational guidance: which mechanics to emphasize, which to redesign, and how to differentiate experiences across user segments. More broadly, this work demonstrates how behavioral telemetry from digital platforms can be systematically converted into a feedback loop for evidence-based intervention management, offering a generalizable framework for organizations operating at the intersection of digital engagement and health outcomes.

2 - Designing Adaptive Responses to Fluoridation Withdrawal: Policy Tradeoffs in Safety-Net Oral Health Systems

Raymond Smith, East Carolina University, Cary, NC, United States, Mark Moss

Community water fluoridation has long served as a foundational public health intervention, providing population-wide protection against dental caries. Recent decisions to reduce or eliminate fluoridation introduce a structural change whose downstream effects on oral health, service demand, and system equity are delayed and difficult to anticipate using static approaches.

This study develops a conceptual system dynamics framework to examine the withdrawal of fluoridation as a disruption to prevention within safety-net oral health systems. Focusing on East Carolina University's Community Service Learning Centers as a representative rural care delivery network, the framework integrates population oral health dynamics, service capacity constraints, and feedback mechanisms that link disease burden to care demand and prevention–treatment trade-offs. A causal loop diagram is used to make system structure explicit and support stakeholder-informed model development.

The framework is used to define and compare a set of adaptive response strategies, including treatment-oriented expansion, targeted preventive interventions, and balanced approaches. Outcomes of interest include untreated disease burden, service access delays, care mix, and indicators of system strain, with attention to impacts on high-risk populations.

Model development and empirical calibration are underway. Preliminary insights emphasize the importance of time delays, the risk of prevention crowd-out under capacity pressure, and tradeoffs between short-term service stabilization and long-term system performance. This work illustrates how system dynamics can support the design of adaptive responses and inform decision-making under preventive policy change.

3 - U.S. Hospital Performance During Crisis: Analyzing Productivity and Efficiency in the COVID-19 Era

Seungjae Shin, Mississippi State University, Meridian, MS, United States

This study comprehensively examines productivity and efficiency within the U.S. hospital industry during the COVID-19 pandemic (2020-2023). Addressing long-standing concerns regarding escalating costs and

persistent productivity challenges, the analysis leveraged panel data and econometric techniques on 22,055 hospital-year records to investigate workforce dynamics, operational efficiency, and financial performance. For labor productivity, the analysis reveals complex annual trends, including a recovery in revenue-generating efficiency but persistent challenges in profitability. Significant heterogeneity exists across ownership types, with state and local hospitals demonstrating uniquely robust financial productivity per staff, possibly linked to their higher nurse-doctor ratios and public service mission. Regarding cost determinants, the analysis highlighted predictable input influence and substantial, escalating time-specific cost pressures, primarily driven by enduring pandemic challenges. While full-time staff, hospital beds, and patient volume positively influence total costs, the nurse-doctor ratio exerts an inverse effect. For Total Factor Productivity (TFP), the Cobb-Douglas production model confirmed labor and capital's positive influence on output. TFP consistently increased over the four-year period, reflecting hospitals' adaptation and their ability to generate more output from given inputs. This research offers crucial insights for policymakers and administrators regarding workforce, cost, and financial sustainability in healthcare, particularly during periods of crisis.

4 - *CANCELED Scope-of-Practice Reform and the Geographic Reallocation of the Healthcare Workforce

Moiz Bhai, University of Arkansas, Little Rock, AR, United States

Healthcare access in the United States is shaped not only by insurance coverage, but also by the supply of providers available to deliver care. Yet policy discussions have disproportionately emphasized demand-side interventions while giving less attention to regulatory barriers that constrain labor supply. This study examines how scope-of-practice regulation influences the geographic distribution of healthcare providers, with particular attention to nurse practitioners (NPs), certified nurse midwives (CNMs), and physicians. Using county-level data from the Area Health Resource Files for 2013 and 2023, we document national changes in provider supply and evaluate how state Full Practice Authority (FPA) laws are associated with differential workforce growth across regulatory environments and local market conditions.

Our empirical strategy combines descriptive geographic analysis with long-difference county-level regressions. We examine changes in provider counts, log supply, and providers per capita, and then assess heterogeneity by baseline physician density and by state FPA status. The results indicate substantial restructuring of the healthcare workforce over the past decade. NP supply increased from 42 to 135 providers per county on average, greatly outpacing physician growth, while CNM supply declined. We also find strong geographic heterogeneity: counties in the lowest quartile of baseline physician density experienced a 0.73 increase in the NP-to-physician ratio, compared with 0.15 in the highest-density quartile. This pattern suggests that NPs disproportionately expanded in underserved markets. Stratified analyses further show that NP growth was especially pronounced in states that adopted or maintained FPA.

RB08

305A

Modeling the Impact of Social Determinants of Health on Healthcare Outcomes and Decision-Making

Invited Session

Health Opportunities for All

Chair: Karen Hicklin, Icahn School of Medicine at Mount Sinai, 1 Gustave L. Levy Place, New York, NY, 10029, United States

1 - Using a Customizable Web-Based Modeling Tool to Assess the Impact of Colorectal Cancer Screening Strategies on Addressing Disparities in Colorectal Cancer Outcomes

Ashley Stanfield, NC State University, Raleigh, NC, United States, Meghan O'Leary, Clayton Ulm, Sai Srikar Puppala, Kristen Hassmiller, Maria Mayorga

Decision-makers, including clinic leadership, state health departments, and policymakers, are often faced with how to equitably improve colorectal cancer (CRC) screening and outcomes in their local context and for specific subgroups. To address this challenge, we developed a customizable metamodeling web-based tool to support decision-making about the potential impact of different strategies to improve CRC screening on long-term health outcomes. The tool allows users to specify their patient population's size and demographic characteristics and to enter in CRC screening rates for stool-based testing and routine colonoscopy before, during, and after intervention for their local population and for specific subpopulations. We present multiple use cases about how this tool can be used to examine potential disparities in long-term CRC-related outcomes by race, age, and sex associated with different strategies. We also show how this tool can be used to set priorities for intervention strategies to improve CRC screening to address known disparities within the local population.

2 - Examining Food Insecurity and Social Need Among Cancer Patients

Jorge Cruz-Muniz, University of Michigan, Ann Arbor, MI, United States, Julie Ivy, Hayley Falk

Food insecurity (FI), defined as limited access to adequate food for a healthy lifestyle, is a critical social determinant of health (SDoH) associated with negative health outcomes, worse quality of life, and treatment nonadherence. While there are estimates of FI among U.S. households, there are no national estimates of FI among cancer patients. Across different geographic and clinical subpopulations, FI among cancer patients ranges from 17 to 55%. FI is often underrecognized in cancer care due to irregular SDoH screening practices and poor incentives for coding SDoH in the electronic health record (EHR). Data on SDoH are captured in the EHR as structured (e.g., survey responses, ICD codes) and unstructured (e.g., clinical notes). Research suggests that unstructured clinical notes often capture more comprehensive information about SDoH than structured data. In this study, our approach is two-fold: (i) using the National Health Interview Survey (NHIS), we estimated the rate of FI cancer patients through time and by type of cancer, and characterized the features of FI cancer patients via clustering, and (ii) for a retrospective cohort of breast, gynecologic, and pancreatic cancer patients from Michigan Medicine, we developed methods to identify patients with FI using structured and unstructured EHR data. We evaluated FI across cancer types and data sources, and assessed its impact on health outcomes. Our analysis of the NHIS suggests that FI differs statistically across cancer types, with gynecologic cancer patients having the highest rates of FI.

3 - The Impact of Food Insecurity and Co-Occurring Social Determinants of Health on Outcomes in The Perinatal Period

Hayley Falk, University of Michigan, Ann Arbor, MI, United States, Julie Ivy

Food insecurity is associated with adverse perinatal outcomes, but rarely occurs in isolation and frequently coexists with other social determinants of health (SDoH). These overlapping factors may create a cumulative burden that disproportionately affects maternal mental health and perinatal outcomes. However, their effects remain underexamined within healthcare decision-making frameworks. This study aims to examine the association between food insecurity, co-occurring SDoH, and health outcomes across the perinatal period. Using data from the University of Michigan, we identify individuals experiencing food insecurity through ICD codes, proxy indicators, and validated screening tools. We use food insecurity as a marker to identify co-occurring SDoH and examine its influence on outcomes during the perinatal period, with a focus on maternal morbidities (preeclampsia, gestational diabetes, anemia, hemorrhage), maternal and neonatal mortality, and neonatal outcomes such as gestational age, birth weight, and living status. By situating food insecurity within the broader context of social and clinical vulnerabilities, our work examines how SDoH burden relates to maternal psychological distress and adverse perinatal outcomes. This study aims to support targeted, evidence-based interventions and strengthen healthcare decision-making for vulnerable populations.

4 - Nutrition-Aware Optimization for Equitable Food Distribution

Nowshin Sharmile, North Carolina A&T State University, Raleigh, NC, United States, Lauren Davis

Food banks play a vital role in addressing food insecurity, yet allocation strategies rarely consider nutritional quality. This work develops an optimization-based framework that explicitly incorporates nutrition using established healthy eating guidelines. Nutritional attributes are structured into decision

variables through a tiered classification, enabling the model to balance serving capacity, accessibility, and nutrition equity across agencies. The framework captures trade-offs between efficiency and equity, particularly under constrained supply, and evaluates how different policy priorities influence distribution outcomes. By integrating nutrition into allocation decisions, the model moves beyond traditional weight-based fairness and promotes more equitable access to healthier food options. This approach provides a practical and scalable pathway for applying data science methods in nutrition and healthcare systems.

RB09

305B

Analytical Models for Public Health Policy and Substance Use Interventions

Contributed Session

Public Health

Chair: M Gabriela Sava, Bowling Green State University, Bowling Green, OH, United States

1 - A System Dynamics Model of Post-Service Opioid use Disorder Risk in U.S. Veterans

Steven Fusco, North Carolina State University, Raleigh, NC, United States, Eda Kemahlioglu-Ziya

Despite a decade of success by the Veterans Health Administration (VHA) in reducing opioid prescriptions following the 2013 Opioid Safety Initiative, overdose mortalities continued to rise among veterans. This suggests a possible shift from controlled clinical use to high-risk illicit use over time. The first 12 months post-service represent a "lethal gap" during which unaddressed latent risk and failed healthcare handoffs may accelerate transition into illicit use and untreated opioid use disorder (OUD).

We build a system dynamics model utilizing an aging chain to track veterans through stocks of clinical exposure, sporadic illicit use, and OUD. Key policy levers include the Department of Defense's Total Force Fitness (TFF) program, which reduces latent risk by addressing accumulated trauma, and a handoff protocol that integrates high-risk screening with VHA appointment scheduling, helping veterans smoothly transition to the VHA system.

Simulations reveal that the sporadic use and the OUD Not in Treatment stocks serve as a death trap, where overdose mortality is projected to spike as the illicit market shifts from fentanyl to high-potency synthetics like nitazenes by 2028. Our analysis demonstrates that handoff success is non-linear: a synergy between screening and VHA appointment scheduling is required to effectively divert high-risk populations into clinical care before they accumulate in untreated, high-risk stocks.

Integrating TFF effectiveness with standardized healthcare handoffs during the military-to-civilian transition may reduce the number of veterans entering illicit opioid pathways. This systems-based approach has the potential to save a significant number of veteran lives during the 2028 synthetic opioid transition.

2 - Impact of Physician Training on Opioids and Non-Opioids Pain Medication Prescribing Patterns in the Emergency Department

Rajni Hiroshima Gurugoda Arachchilage, Bowling Green State University, Bowling Green, OH, United States, M Gabriela Sava, Lauren Maziarz, Jerrold May, Tiffany Sayre

The opioid crisis remains a critical public health challenge in the United States, with Ohio consistently reporting among the highest drug overdose death rates. Emergency Departments (EDs) play a pivotal role in this crisis, as they are often the first point of clinical exposure to opioid medications in a clinical setting. This study examines the effectiveness of the Prescribing Alternatives Instead of Narcotics (PAIN) program, a state-funded physician training initiative, on opioid and non-opioid prescribing patterns across two EDs in Northwestern Ohio. A two-stage analytical framework is applied to (a) compare opioid and alternatives to opioids (ALTO) prescribing rates across seven pain conditions; and (b) identify patient-level prescribing patterns through association rule mining. The results show statistically

significant reductions in opioid prescribing for multiple pain conditions following physician training, accompanied by increased ALTO utilization. Association rule analysis reveals shifts toward non-opioid centered prescribing patterns. These findings suggest that structured physician training, supported by decision tools, can effectively influence prescribing behavior. The PAIN program demonstrates a scalable approach to reducing inappropriate opioid use while promoting evidence-based, multimodal pain management in ED settings.

3 - Data-Driven Analytical Model for Automatic Opioids Screening in the Emergency Department

M Gabriela Sava, Bowling Green State University, Bowling Green, OH, United States, Abraham Sanyang, Millicent Berko, Jerrold May, Lauren Maziarz, Tiffany Sayre

The opioid crisis has claimed roughly 806,000 American lives since 1999, according to the CDC. In 2023 alone, opioids were involved in nearly 80,000 of the approximately 105,000 recorded drug overdose deaths — nearly ten times the toll seen in 1999. There are tentative signs of progress: overdose deaths tied to prescription opioids dropped by nearly 12%, and the overall opioid overdose death rate fell 4% between 2022 and 2023. Yet the crisis is far from resolved, and Emergency Departments (EDs) remain a vital point of intervention for identifying individuals at risk before their situations become fatal.

It is against this backdrop that the present study introduces a data-driven framework for improving opioid use disorder (OUD) screening in ED settings — with the aim of connecting more patients to life-saving treatment. Recognizing the well-established shortcomings of manual screening methods, this study moves beyond sole reliance on clinician judgment or patient-completed questionnaires. Instead, an analytical model is trained to identify patients at elevated OUD risk using information already captured in the healthcare system's electronic health records (EHR), drawing on variables such as patient demographics, prior diagnoses, medication history, and ED visit patterns.

RB10

306A

Forecasting Methods in Healthcare Data Analytics

Contributed Session

Data Analytics

Chair: J.J. Van De Klundert, Universidad Adolfo Ibanez, Viña del Mar, Chile

1 - Predicting Hidden Transnational Organ Trafficking Links Using Supervised Link Prediction

Hasini Balasuriya, University of Louisville, Louisville, KY, United States, Cagla Acun, Monica Gentili, Naoru Koizumi

Illicit organ trade operates as a hidden transnational market driven by severe shortages in legal transplantation and the economic vulnerability of organ sellers. However, the underlying network is rarely observed in full, since available evidence is both sparse and fragmented across sources. This creates a fundamental analytical challenge: the observed network reflects detection rather than true operations, limiting situational awareness and making it harder for law enforcement to prioritize targets and design effective interventions. This study addresses this limitation by modeling organ trafficking as a partially observed country-level criminal network and applying supervised link prediction to infer likely missing connections. Using an incident-based dataset covering transnational organ trafficking from 2000 to 2022, we construct a directed, role-aware network where countries participate as sellers, buyers, brokers, or surgery locations. We develop a supervised prediction pipeline that integrates local structural similarity measures, with contextual country-level indicators. We evaluate five classification models under a temporally ordered train-test framework reflecting real-world prediction settings. Results show that tree-based ensemble methods, particularly XGBoost and Random Forest, achieve the strongest performance (ROC AUC = 0.953), substantially outperforming linear and kernel-based approaches. Feature importance analysis indicates that local network structure dominates prediction, suggesting trafficking

ties emerge through repeated interactions and embedded operational neighborhoods rather than random pairings. These findings demonstrate that supervised link prediction provides a scalable analytic approach for uncovering hidden transnational trafficking relationships and prioritizing plausible connections for further investigation and policy response.

2 - Forecasting Long-Term Surgical Readmissions and Costs For Capacity and Payment Planning

Raymond Martin, Louisiana State University, Baton Rouge, LA, United States, Bruno Rego

Accurate long-term demand forecasts are essential for hospital capacity planning, multi-year budgeting, and the design of risk-bearing payment contracts, yet current models focus mainly on binary outcomes at a single time (e.g., 30-day readmission) and provide limited guidance for long-horizon operational decisions. Using the Nationwide Readmissions Database, we developed models to forecast patient-level, multi-year readmission incidence and resource use following five high-volume surgical procedures frequently targeted in bundled payment programs: cholecystectomy (gall bladder removal), hip replacement, knee replacement, coronary artery bypass graft, and carotid endarterectomy (arterial plaque removal). We estimated time-dependent readmission risk using a regularized Weibull proportional hazards model with index-admission demographic, clinical, and hospital predictors, and modeled readmission cost and length of stay with multivariate regression. Cumulative readmission for median-risk patients is estimated to reach 50% by 10 years for cholecystectomy and 2.6-5.4 years for the other procedures, with broad heterogeneity (10th-90th percentile spreads of 38-58 percentage points). Median readmission costs and length of stay were similar across procedures ($47,800-$55,400; 3-4 days), but variability was large (interquartile ranges $61,500-$93,500). Cost predictability was modest yet consistent (Spearman's correlation 0.34-0.44). These results show that long-term readmissions create substantial and forecastable workload that varies markedly by patient mix. These models can support operational decisions by translating current case mix into multi-year demand projections, enabling improved resource planning, risk-adjusted bundled payment pricing, and targeted strategies to mitigate high-variance utilization. For health systems and payers, this framework offers a scalable tool to apply statistical prediction for improved capacity management and financial risk control.

3 - Forecasting Drug use Under Market Shifts: a Hybrid Prophet-LSTM Model

Qiaoyuan Li, UC Berkeley, Berkeley, CA, United States

Drug utilization forecasts are widely used to support high-stakes decisions in pharmaceutical inventory planning, payer cost budgeting, pharmacy reimbursement strategy, and insurance policies. However, traditional time-series methods often struggle to achieve high accuracy in complex healthcare environments. Especially during market-shifting events, traditional models cannot capture such changes from past monotonic trends, which often leads to suboptimal downstream decisions.

To tackle this problem, this study proposes an innovative hybrid Prophet + LSTM model that combines the flexibility of the Prophet model with LSTM to boost accuracy. The Prophet component captures complex seasonality and trends, while the LSTM component subsequently learns nonlinear residuals and complex patterns. The external regressor detects the shock events and incorporates market shifts into forecasted values.

Using Medicaid drug utilization data to forecast across major therapeutic classes, the hybrid model demonstrates substantial performance gains. Compared with ARIMA, Holt-Winters, and the standalone Prophet model, the hybrid model consistently reduced MAPE from 11.76% to 4.61% and reduced cumulative MAE by 1.1 million prescriptions across seven Medicaid drug classes during the 2023–2025 evaluation period. More importantly, this model correctly identified utilization inflection points following the Medicaid Unwinding trend shifts.

This novel framework provides an enhanced forecasting architecture for decision-making in complex, evolving healthcare environments. Improved forecasts support reduced market frictions, more robust pharmaceutical inventory and production planning, and more credible healthcare policy evaluation.

Overall, the framework helps decision makers leverage AI-driven insights and dynamically respond to market and policy changes.

4 - Privacy-Preserving AI Testing at Scale Using Ai-Generated Synthetic Data to Validate Healthcare Systems Without Exposing Phi

RAMA KRISHNA KUMAR Lingamgunta, Cigna Evernorth Services Inc, Raleigh, NC, United States

As healthcare organizations scale the use of artificial intelligence across clinical, operational, and analytics platforms, the need for realistic test data in non-production environments has become a major barrier to innovation. Development and validation efforts often rely on real patient data, creating significant privacy, security, and compliance risks that slow AI adoption and increase organizational exposure.

This session explores how AI-driven synthetic data generation can be used to safely test healthcare systems at scale without exposing protected health information. The discussion covers both structured data, including relational and FHIR-based resources, and unstructured content such as clinical notes, documents, and attachments. Emphasis is placed on preserving referential integrity, ensuring that relationships across patients, encounters, observations, claims, and documentation artifacts remain realistic and usable for end-to-end system validation.

Through practical implementation patterns, the session demonstrates how synthetic data can support AI model testing, interoperability validation, and workforce training while aligning with enterprise privacy and security requirements. Attendees will gain actionable insight into how synthetic data enables faster innovation, reduced risk, and more reliable AI deployment across healthcare environments.

5 - Improving Survival Prediction: Higher C Index Scores and Tighter Upper Bounds for Proportional Hazards Models

J.J. Van De Klundert, Universidad Adolfo Ibanez, Viña del Mar, Chile

Survival prediction has gained much interest as an area in which machine learning models can resolve perceived poor performance of classical models such as the Cox proportional hazards model. This is especially valuable when using predictive analytics in allocation policies, such as in organ allocation. The C Index measures the discrimination performance of prediction models, i.e., the fraction of pairs of population members for which a model predicts correctly who lives longest. For applications such as survival prediction after kidney transplantation, traditional and machine models alike deliver C Index score that are much closer to the score of 0.5 obtained by random predictions than to the perfect score of 1.

We present theoretical and empirical upperbounds for C Index scores that prediction models can be expected to provide. We provide theoretical upperbounds of ln 2 and 0.75 in case the proportional hazard ratios are uniformly, resp. exponentially distributed. We present comparable empirical results for graft survival after kidney transplantation in the US. These empirical results are extended to subpopulation based C Indices which help to understand whether differences in prediction performance are explained by the upperbounds or by estimation biases.

Additionally, we formulate the problem of estimating proportional hazards models as an integer linear programming problem. We show this ordering problem to be strongly NP hard and present computational results for the C Index attainable by proportional hazards models on UNOS data for graft survival after kidney transplantation.

This research is partially funded by Fondecyt Regular grant 1230361

RB11

306B

Health System Performance and Policy Across Contexts

Contributed Session

Infectious Diseases

Chair: Sanjay Mehrotra, NORTHWESTERN UNIVERSITY, C246, Tech Institute, Evanston, IL, 60208, United States

1 - Epidemiological Modeling Using Network Analysis on Cellular Geolocation Data and Autotuning Hyperparameters in Ensemble Machine Learning Algorithms

Carlos Pinheiro, SAS Institute Inc, Cary, NC, United States

Cellular geolocation identifies mobile device positions by analyzing network connections through triangulation and signal strength. This generates high-resolution temporal-spatial data that serves as a proxy for population-level urban mobility. By applying network analysis on geolocated cellular data, it is possible to extract critical metrics to describe urban mobility behavior, including overall routing, general paths and locations connectivity. These network centralities reveal how populations flow through metropolitan regions over time.
When integrated with epidemiological data, these mobility patterns become powerful predictors for viral transmission. In this framework, positive disease cases serve as dependent variables for supervised machine learning models. Because viral outbreaks move quickly, the temporal observation window is narrow, requiring models that can adapt to shifting human behaviors in real-time.
To maintain predictive accuracy and strong generalization, two advanced computational strategies are employed:
Hyperparameter Autotuning: As viral spread accelerates, mobility behaviors shift (e.g., due to lockdowns or social distancing). Autotuning continuously recalibrates model parameters, ensuring the algorithms capture the evolving correlation between urban movement and infection rates.
Ensemble Learning: By aggregating the posterior probabilities of multiple machine learning algorithms, the system enhances generalization. This ensemble approach mitigates the risk of overfitting specific data snapshots, ensuring robust estimations even as new mobility patterns and infection clusters emerge.
Ultimately, this data-driven approach transforms raw network signals into a sophisticated predictive tool. By modeling the relationship between movement and contagion, it allows public health officials to visualize the dynamic path of a virus across geographical regions with high temporal precision.

2 - Retrospective and Sequential Calibration of Seir Models with Delayed Observations

BumJoon Cho, NCSU, Raleigh, NC, United States

This work studies calibration and inference in deterministic SEIR epidemic models under partial and delayed observation. We consider a setting in which core epidemiological parameters—such as transmission, latency, recovery rates, and initial conditions—are assumed known, and the focus is on reconstructing time-varying transmission dynamics from observed data streams, including reported cases, hospitalizations, and deaths. These signals are linked to latent infections through fixed delay structures and severity parameters, enabling a multi-signal calibration framework.

We propose a retrospective calibration approach that leverages the full availability of historical data to infer piecewise transmission dynamics over past time horizons. Using a structured sampling strategy inspired by Latin hypercube designs, candidate transmission values are generated at multiple time points and evaluated jointly to identify combinations that best reproduce observed signals. This retrospective fitting establishes a consistent "grounded" trajectory once all delayed observations have been realized.

Building on this, we introduce a sequential forward calibration procedure in which candidate trajectories are propagated over future time periods using a branching structure, with pruning based on goodness-of-fit. This produces an ensemble of plausible transmission paths that naturally captures uncertainty arising from delayed observations. The resulting ensemble can be interpreted as a cone of uncertainty around recent estimates.

Finally, we demonstrate how deviations between observed data and the ensemble of calibrated trajectories can be used for anomaly detection, signaling potential structural changes in disease dynamics. This framework provides a unified approach for retrospective inference, short-term forecasting, and uncertainty quantification in epidemic simulation models.

3 - Medical Tourism: Boon OR Bane for the Host Nation

Sundara Natarajan Panchanatham, HKU Business School, Hong Kong, Hong Kong, Omkar D Palsule Desai, Saurabh Bansal

Medical tourism has a growing role in global healthcare, where foreign demand can impact host-country's healthcare capacity, equity, and welfare. Using analytical modelling, our study explores capacity competition, congestion effects, and regulatory levers to inform policies that balance efficiency, access, and investment for sustainable cross-border healthcare systems.

4 - Decision-Dependent Utility Robust Service Facility Location with Application to Achieving Herd Immunity and Decision-Dependent Utility

Sanjay Mehrotra, NORTHWESTERN UNIVERSITY, Evanston, IL, United States, Shibshankar Dey, Esma Gel

We studly a novel decision-making framework for the facility location problem where the location decisions maximize an unknown concave utility, which belongs to an ambiguity set. The resulting model is computationally challenging because the expressions appearing in the model are not closed-form. We develop a surrogate model that yields an optimization model with a more tractable mixed-binary bi-convex structure. We develop a novel linearize–branch–and–union (LBU) decomposition algorithm for solving the resulting problem. Enhancements for the bi-convex setting involve a six-way disjunction strategy that generates valid separations to cut bi-convex infeasible solutions. A guided cut-generation technique is introduced to accelerate convergence. Computational performance is further improved by decomposing the cut-generating linear program at the expense of looser capacity constraints.

Computational experiments focusing on a vaccination center location problem that incorporates the immunization effect of vaccination through the herd effect using the Lambert function demonstrate the superior performance of our proposed algorithm and provide valuable managerial insights. On smaller instances generated from zip codes of the five major counties of the state of Illinois, our approach achieves optimality gaps below 1\%, compared to $>45\%$ from commercial solvers with a 12-Hour time limit. On state-level instances, our framework with a customized adaptive-large-neighborhood-search meta-heuristic reaches about a 20\% gap, while commercial solvers run out of memory.

Solutions provide valuable insights: the use of concave utility yields more balanced geographical coverage. When combined with herd immunity, it further promotes allocations that minimize aggregate susceptibility to infection, supporting resilient and equitable service network design.

RB12

306C

AI and Data-Driven Healthcare Operations

Invited Session

MSOM: Healthcare

Chair: Yue Hu, Stanford University, Stanford, CA, United States

1 - Incorporating AI into Risk Prediction and Triage Support Using Secure Messaging

Sean Tsung, Stanford University, Stanford, CA, United States, Jiayi Wang, Mohsen Bayati, Yue Hu

Secure messaging enables patients to communicate asynchronously with their care teams through electronic health record portals. As message volumes have grown, so has the challenge of identifying which

messages signal imminent clinical deterioration. We develop and validate AI models that predict emergency department (ED) admission within seven days of a secure message, aiming to support triage workflows and enable timely escalation of time-sensitive requests. Using secure messaging data from primary care clinics spanning January 2023 through February 2024, we define the prediction target as a binary indicator at the patient-day level of ED visit within seven days and evaluate models on a temporally held-out test set. Baseline features include demographics, comorbidity measures, and one-year utilization counts. We compare three feature generation strategies: baseline structured features alone; baseline features augmented with representational text features including bag-of-words and embedding representations reduced via singular value decomposition; and baseline features augmented with clinically interpretable variables extracted using a multi-agent large language model system adapted for secure messaging. This agentic system reads a sample of patient messages, proposes candidate clinical features relevant to the prediction task, validates that they can be reliably extracted across the message corpus, and iteratively refines extraction logic through self-correcting validation loops. Our results demonstrate the relative value of structured, interpretable feature extraction from secure messages compared to representational approaches for short-term ED risk prediction. This work offers a practical framework for clinics to prioritize message responses and reduce avoidable downstream utilization.

2 - Tiered Service Architecture for Remote Patient Monitoring

Isha Thapa, Stanford University, Stanford, CA, United States, Siddharth Chandak, Nicholas Bambos, David Scheinker

tract— We develop a remote patient monitoring (RPM) service architecture, which has two tiers of monitoring: ordinary and intensive. The patient's health state improves or worsens in each time period according to certain probabilities, which depend on the monitoring tier. The patient incurs a "loss of quality of life" cost or an "invasiveness" cost, which is higher under intensive monitoring than under ordinary. On the other hand, their health improves faster under intensive monitoring than under ordinary. In each period, the service decides which monitoring tier to use, based on the health of the patient. We investigate the optimal policy for making that choice by formulating the problem using dynamic programming. We first provide analytic conditions for selecting ordinary vs intensive monitoring in the asymptotic regime where the number of health states is large. In the general case, we investigate the optimal policy numerically. We observe a threshold behavior, that is, when the patient's health drops below a certain threshold the service switches them to intensive monitoring, while ordinary monitoring is used during adequately good health states of the patient. The modeling and analysis provides a general framework for managing RPM services for various health conditions with medically/clinically defined system parameters

3 - Scalable Stewardship of An LLM-Assisted Clinical Benchmark with Physician Oversight

Junze Ye, Stanford Graduate School of Business, Stanford, CA, United States, Daniel Tawfik, Alex Goodell, Nikhil Kotha, Mark Buyyounouski, Mohsen Bayati

Reference labels for machine-learning benchmarks are increasingly synthesized with LLM assistance, but the reliability of such labels remains underexamined. We audit MedCalc-Bench, a clinical benchmark for automated medical score computation whose original labels were partly derived using GPT-4, and develop a scalable physician-in-the-loop stewardship pipeline to reassess them. Across the test set, at least 27% of cases were likely mislabeled or incomputable. We further release physician-informed recomputed train and test labels and show that correcting label errors materially changes downstream evaluation. On our recomputed test labels, today's frontier LLMs achieve a mean accuracy in the 85-90% range, whereas using the original labels would have consistently *underestimated* model performance by nearly 25%. These findings suggest that LLM-assisted benchmarks can propagate systematic errors into both model assessment and post-training unless they are actively stewarded. Benchmark development in high-stakes domains should therefore pair scalable automation with expert adjudication.

Thursday, July 30, 2026, 11:00 AM - 12:00 PM

Thursday Plenary

Ballroom B

Operationalizing AI in Health and Science: Governance, Alignment, and Execution

Plenary Session

Plenary

1 - Operationalizing AI in Health and Science: Governance, Alignment, and Execution

Jamie Pina, RTI International, Raleigh, NC, United States

Many organizations are investing heavily in analytics and AI, yet still struggle to translate those efforts into sustained performance improvement. Drawing on his experience leading enterprise data and AI strategy at RTI International, Dr. Pina will share practical lessons on how emerging technologies can be positioned to drive meaningful organizational results. The session will focus on governance, leadership alignment, and disciplined execution approaches that help turn technical capability into measurable gains in cost, quality, access, and overall system performance.