All of the artworks in this report were created by Microsoft Bing Image Creator, powered by DALL·E 2 or DALL·E 3.

The cover image was created by providing a prompt to create an abstract image representing impact, "accessible, flow, artificial intelligence, assistant, machine learning, health care, healthcare, hospital, river, and operations, in the style of Roy Lichtenstein"
About DIHI

The Duke Institute for Health Innovation (DIHI) catalyzes transformative innovations in health and healthcare through implementation of high-impact innovations, leadership development, and cultivation of a community of entrepreneurship. We bring innovative solutions to the most pressing challenges in health and healthcare through multidisciplinary teamwork across Duke University and Duke Health and by fostering collaborations with national and international thought leaders.
Contents

Letter from the Directors ................................................. 07
Impact Summary .............................................................. 08
RFA Impact Write-ups ...................................................... 12
Predicting Behavioral Emergencies in the Hospital ............. 12
eProvider using the EHR .................................................. 18
Prescription for Repair ...................................................... 24
Implementation and Evaluation of the Duke Health Patient Initiated Note about Goals (PING) ......................... 30
Improved Perioperative Care Coordination via Enhanced Decision Support Tools ........................................ 36

AI Life Cycle Best Practices ........................................... 55
Introduction ........................................................................ 56
Building Common Governance and Best Practices for AI/ML Tools in Healthcare .......................................... 57
Fostering an Ecosystem to Advance Governance Standards for Responsible AI Adoption in Healthcare ......... 58
Leading the Effort via Health AI Partnership (HAIP) .......... 60
Equity Analysis for a Pediatric Sepsis Model to Inform AI Governance Guides ............................................... 63
Scaling AI/ML Tools and Governance Practices in Lower-Cost Settings ..................................................... 64
The Algorithm Journey Map .............................................. 65

Our Team ........................................................................... 80
Publications ...................................................................... 82
List of Collaborators ....................................................... 84
DIHI Perspectives

The Truth About Large Language Models and Their Potential to Revolutionize Healthcare .......... 10
Harnessing the Power of ChatGPT in My Day-to-day Work as a Healthcare Innovator .......... 16
Introducing Bed Watch ................................................................. 21
Identifying Candidates for PrEP to Prevent Incident HIV in Louisiana and Mississippi ............. 28
Duke Opioid Sedation Assessment (DOSA) ........................................ 33
The Vitality of Real-Time Surgical Outcomes Data .................................................. 40
DIHI AutoML Platform .......................................................................... 45
Automation of SEP-1 Quality Reporting ...................................................... 50
Monorail: The Quest for Universal EHR Monitoring & Analytics ............................................... 68
Cohort Builder ....................................................................................... 74
Entity Resolution ..................................................................................... 79

DIHI Experiences

FREYA GULAMALI .................................................................................. 11
JIAYU YAO .......................................................................................... 15
TIMOTHY OCHOA ................................................................................. 17
GAURAV SUDERSHMUKH ............................................................... 27
SHEMS SALEH ...................................................................................... 35
KAIVALYA DESHPANDE ................................................................. 49

Diffusion & Scaling

Advancing Surgical Risk Prediction: Unveiling the Power of Python ........................................... 43
Sepsis Watch and Adult Deterioration ....................................................................................... 46
Improving Outcomes for Patients Who Develop Sepsis at Scale ............................................... 52
Maternal Early Warning System (MEWS) ............................................................................. 70
Thomas Owens, MD
Executive Director, DIHI
Executive VP and COO
Duke University Health System

Suresh Balu, MBA, MS
Director for DIHI
Associate Dean, Innovation and Partnership, School of Medicine
Letter from the Directors

The rising cost of healthcare has become a pressing concern, especially in the wake of the operational challenges faced by health systems here and worldwide due to the impact of the COVID-19 pandemic. Enhancing operational efficiency and optimizing patient care is not only essential for effective cost control, but by allocating resources more effectively, such measures also improve health equity, enhance patient care, reduce the risk of errors, help address capacity challenges, and ultimately lead to greater resilience and preparedness in the face of unexpected challenges.

Achieving operational efficiency in healthcare operations often involves the integration of multiple methodologies and best practices. Some of the key methods used to improve efficiency include utilizing the principles of lean management, data-driven decision-making, increased focus on collaboration and adaptability, and, more recently, the equitable and safe deployment of AI and machine learning algorithms and predictive analytics that utilize vast amounts of data leading to the generation of more targeted interventions and workflow management.

This year, our innovation efforts focused on automation to enhance healthcare operational efficiency. This report highlights some projects our leadership selected to align with the above themes through the annual DIHI call for proposals, a competitive process for developing, implementing, and scaling interventions across the clinical enterprise. Outside the DIHI RFA process, our team led efforts to establish the Health AI Partnership in collaboration with local, regional, and national peer institutions, community partners, professional societies, and regulatory bodies. Funded by the Gordon Betty Moore and McGovern Foundations, the partnership seeks to empower healthcare organizations to use AI safely, effectively, and equitably while being a trusted resource for healthcare professionals vested in leveraging innovation and technology for a human-centric approach to providing care. And we have continued to engage with learners – from local schools in Durham to undergraduate and professional students at Duke, medical fellows, and residents. We also describe their contributions to advancing innovation in this edition of the DIHI Impact Report.

In the new academic year, we look forward to continuing to support health system operations by incorporating digital technologies to improve efficiency, increase access to the highest-quality care, reduce workload for our frontline clinical care teams, and support the overall growth of our organization. This is possible, given our incredible and highly motivated faculty, staff, and students and strong track record of deploying innovative technologies and processes to meet the needs of our patients and clinical enterprise.

Thomas Owens, MD
Suresh Balu, MBA, MS
Innovation Portfolio Impact Summary

Innovation Portfolio Impact: >$245K per DIHI staff member per year

eProvider in Duke Family Medicine
Burnout among MDs and PAs in June ‘23 vs June ‘22 dropped by 24%, according to surveys. 3,419 eProvider visits brought $480K in revenue. Provider hours managing patient advice request messages per clinical day went down from 1.91 to 1.50 hours per day. The after-work hours decreased from 2.72 to 2.13 hours. 40 Duke Family Medicine providers gained 49 minutes daily due to optimizing Patient Advice Requests: 32.6 hours saved per day. At a conservative average of $30/hour and 260 work days, this saved $255K in a year.

HealthGuard™
DIHI sent 1,955 page/e-mail notifications to clinical care teams for goal-concordant-care initiation since 2021 go-live and they identified 991 hospitalized patients needing advanced care planning (ACP). In the last 6 months, ACP use increased by 53%. $1.0M - $64.2M is the range of reduced care delivery costs to DUHS (estimated) as a result of patient goals of care established and documented in the EHR discussions.

<table>
<thead>
<tr>
<th>Tool</th>
<th>A tool that...</th>
<th>Type</th>
<th>Calculation Per Day</th>
<th>Calculation in the last 12 months</th>
<th>All Time Calculations</th>
<th>Influence Scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acute Kidney Injury</td>
<td>Post-surgical acute kidney injury prediction tool</td>
<td>Model</td>
<td>1,705</td>
<td>491,446</td>
<td>530,521</td>
<td>4,812 Unique Patients</td>
</tr>
<tr>
<td>Predicting BE</td>
<td>Inpatient Behavioral Response Team alerting tool</td>
<td>Model</td>
<td>196</td>
<td>196</td>
<td>196</td>
<td>1 Unique Patients</td>
</tr>
<tr>
<td>Cardiogenic Shock</td>
<td>Seven phenotypes to identify cardiac decompensation and shock</td>
<td>Model</td>
<td>26,057</td>
<td>8,630,724</td>
<td>17,772,015</td>
<td>131,751 Unique Encounters</td>
</tr>
<tr>
<td>RT Decomp Mortality</td>
<td>Real-time prediction of decompensation or need for ICU-level care</td>
<td>Model</td>
<td>31,164</td>
<td>126,611</td>
<td>602,556</td>
<td>3,882 Unique Patients</td>
</tr>
<tr>
<td>DOSA</td>
<td>Duke Opioid Scoring Algorithm</td>
<td>Model</td>
<td>18,024</td>
<td>6,430,013</td>
<td>6,747,798</td>
<td>47,438 Unique Patients</td>
</tr>
<tr>
<td>OR Efficiency</td>
<td>OR asset scheduling model</td>
<td>Model</td>
<td>381</td>
<td>105,762</td>
<td>276,040</td>
<td>127,340 Unique Patients</td>
</tr>
<tr>
<td>ED Triage</td>
<td>Emergency Department patient flow prediction model</td>
<td>Model</td>
<td>2,200</td>
<td>1,170,986</td>
<td>3,581,817</td>
<td>210,270 Unique Patients</td>
</tr>
<tr>
<td>GOPOP</td>
<td>Nomogram for post surgery opioid pill prescription</td>
<td>Cohort Identification</td>
<td>14</td>
<td>3,581</td>
<td>7,367</td>
<td>6,746 Unique Patients</td>
</tr>
<tr>
<td>HealthGuard</td>
<td>Inpatient, 30-day, 60-day mortality prediction</td>
<td>Model</td>
<td>3,598</td>
<td>2,421,392</td>
<td>9,185,765</td>
<td>127,907 Unique Patients</td>
</tr>
<tr>
<td>IRAE</td>
<td>Readmission prediction model for patients on immune therapy</td>
<td>Model</td>
<td>50</td>
<td>4,969</td>
<td>5,532</td>
<td>1,278 Unique Patients</td>
</tr>
<tr>
<td>MEWS-Sepsis</td>
<td>Real-time prediction of Sepsis in L&amp;D patients</td>
<td>Model</td>
<td>526,725</td>
<td>9,904,912</td>
<td>15,178,826</td>
<td>1,200 Unique Patients</td>
</tr>
<tr>
<td>MEWS-Hemorrhage</td>
<td>Real-time prediction of 1 m, 2 rs hemorrhage in L&amp;D patients</td>
<td>Model</td>
<td>526,725</td>
<td>10,514,598</td>
<td>17,335,942</td>
<td>1,243 Unique Patients</td>
</tr>
<tr>
<td>NAFLD</td>
<td>Cohort identification of potential patients with NAFLD</td>
<td>Cohort Identification</td>
<td>520</td>
<td>15,280</td>
<td>28,359</td>
<td>20,977 Unique Patients</td>
</tr>
<tr>
<td>Neuroscience Mortality</td>
<td>Inpatient, 30-day, 60-day mortality among neurosciences patients</td>
<td>Model</td>
<td>1,608</td>
<td>351,315</td>
<td>383,458</td>
<td>1,374 Unique Patients</td>
</tr>
<tr>
<td>Sepsis Watch v1</td>
<td>Sepsis phenotype and risk of sepsis prediction across DUHS hospitals</td>
<td>Model</td>
<td>22,582</td>
<td>8,164,680</td>
<td>34,796,892</td>
<td>622,127 Unique Encounters</td>
</tr>
<tr>
<td>Sepsis Watch v2</td>
<td>Sepsis phenotype and risk of sepsis prediction across DUHS hospitals</td>
<td>Model</td>
<td>230,961</td>
<td>65,038,215</td>
<td>87,483,415</td>
<td>516,132 Unique Encounters</td>
</tr>
</tbody>
</table>

1,392,510  113,374,670  193,916,499
Predicting Hospital Admissions and ER Visits from Immune-related Adverse Events
We have deployed and are preparing to scale a solution to identify immunotherapy patients at high risk of admission or Emergency Department visits. We aim to reduce planned ED and hospital visits and improved outcomes by 121 fewer readmissions per year. The average cost of an IRAE readmission is $18,819. We will save $2.1M annually.

Sepsis Watch
Providers used Sepsis Watch to identify 3,020 patients who met the CMS “sepsis time zero” in the ED. For DUH, DRAH, and DRH, respectively: Sep-1 Bundle Compliance has increased by 120%, 130%, and 49%; Observed/Expected length-of-stay for sepsis patients has decreased 1%, 9%; Duke Sepsis patient O/E Mortality has decreased by 50%.

Gynecology Postoperative Opioid Predictive (GOPOP) Calculator
Duke Health reduced prescription of opioid pills by 4,000 without any adverse effect on patient care and experience for ~800 Oncology Division cases / year. Compared to a historical cohort, GO-POP reduced opioid prescribing from 15 to 10 pills per patient.

The Health AI Partnership and AI adoption Framework
We launched the Health AI Partnership, a nationwide collaborative community, for AI governance and best practices in healthcare. We lead the initiative, which brings together healthcare experts and influencers to create guidelines for developing, assessing, and integrating AI solutions. The collaboration has produced an AI adoption framework based on eight key decision points, comprising thirty-one topic guides. These guidelines help healthcare system leaders ensure the safe, effective, equitable and responsible use of AI by healthcare organizations.

Prescription for Repair
Over this one-year pilot, we hired a program coordinator, enrolled 30 gunshot survivors or their families, trained 28 community-based facilitators, and conducted 56 listening sessions. Forty-five sessions were transcribed and made accessible to research teams for analysis.

Perioperative Care Coordination
The immediate outcome is an uptick in the volume of surgeries at Ambulatory Surgery Centers, which directly enhances their operational efficiency. Simultaneously, this would liberate more block time at Duke North, catering to surgeries requiring an inpatient framework.

Predicting Behavioral Emergencies in the Hospital
Our algorithm identified 5,261 encounters (4,759 patients) who met the behavioral emergency outcome definition during the encounter. A May-November 2022 comparison of the behavioral emergency outcome (n=203) with behavioral emergency clinical note documentation (n=506) and dual physician adjudication (n=31) yielded a precision of 90% and sensitivity of 36% for the behavioral emergency outcome.

Patient Initiated Note about Goals (PING)
Over 15 months, 678 unique patients responded to the PING via direct-to-patient MyChart message or the Duke Primary Care Population Health Nurse workflow. Through the project and increased nurse partnerships in the year, direct-to-patient communication more than doubled. 85% of patients surveyed believed they were more able to care for themselves.

Pragmatic Framework to Ensure Equity of AI Solutions in Healthcare
The Health Equity Across the AI Lifecycle (HEAAL) framework was developed with input from a community of AI experts in healthcare that the Health AI Partnership spearheaded. Our framework offers a systematic approach to evaluate how AI affects health equity. It consists of 37 procedures for assessing existing AI solutions and 34 for new ones. This framework emphasizes accountability, fairness, and transparency, enabling healthcare institutions to manage risks and allocate resources for equitable implementation of AI in healthcare.

Pediatric Sepsis and Decompensation Prevention
Within the cohort of all pediatric hospitalizations from DIHI and Pediatric Emergency Medicine, we created and implemented paging notification systems for real-time digital phenotypes for pediatric sepsis and for febrile pediatric patients with high risk conditions. Among the cohort of all pediatric hospitalizations from November 2016 through April 2023 (28,399 total hospitalizations), 3,265 (11.5%) met the pediatric sepsis phenotype and 777 (2.7%) met the high-risk condition + fever phenotype. The real-time sepsis phenotype identifies 90% of patients who eventually become septic, with a positive predictive value of 1-in-4. The high-risk condition + fever phenotype will help improve timely administration of antibiotics from 48% to 75% (goal).
In October 2022, OpenAI unveiled ChatGPT, showcasing for the first time the capabilities of Large Language Models (LLMs) on a grand scale. Within weeks, the enthusiasm for LLMs grew exponentially, with industries worldwide, including healthcare, eager to exploit their potential. Major Electronic Health Record (EHR) companies and health systems were among those captivated by its promise.

So, where do we stand now? As of this article’s publication, less than a year has passed since LLMs were thrust into public awareness. The momentum behind LLMs is only accelerating, as evident from enterprise activity and emerging trends in the LLM sphere. Tech giants like OpenAI, Microsoft, Google, and Meta are spearheading developments within the enterprise space and the open-source community. Meta, in particular, has led and leveraged the open-source community, releasing powerful open models such as Llama and Llama2 for public use. These open-source models have ignited a race to refine and adapt them to rival ChatGPT and other proprietary systems. Furthermore, the influx of venture capital into LLM-based startups suggests that humanity can capture real value if people (or machines) deploy these models effectively.

However, despite the flurry of activity, LLMs seem predominantly in the demo and pilot phases, especially within healthcare. Though the application surface is vast, the industry realizes that instantaneous benefits might be rarer than anticipated. LLMs represent an entirely novel technological paradigm. Hence, the infrastructure to employ, interact with, and ensure the reliability and safety of LLM-dependent applications is still an active development area.

At the Duke Institute for Health Innovation (DIHI), our data science team has delved deep into LLMs and the complexities of developing solutions in this realm. Peering behind the veil of newly-coined concepts such as Retrieval Augmented Generation (RAG), Low-Rank Adaptation (LoRA), and Parameter Efficient Fine-Tuning (PEFT), we’ve discovered ample reasons for optimism. However, harnessing these technologies for enterprise use is more complex than many realize and requires specific skills and expertise. Specialized hardware, such as Graphics Processing Units (GPUs, essential for scaling LLMs), presents another challenge, partly explaining NVIDIA’s meteoric stock price.

Michael Gao, MS
Our verdict? LLMs, integrated with existing software practices, hold the potential to revolutionize vast sectors of healthcare. The promise and excitement surrounding them are immense. Yet, realizing their full potential demands meticulous research, development, and a dedicated, multidisciplinary team to ensure that solutions are efficient and secure and uphold quality standards. The mere existence of this technology does not imply that Duke will emerge as a leader in its use or that the technology will resolve all of our challenges unless there is a genuine commitment in time, energy, and resources toward the development and roll out of LLM-centric solutions.

At DIHI, our mission has always been to catalyze transformative innovation in health and healthcare, and we are excited to contribute to developing LLM solutions as another tool in our toolbox. Stay tuned!

The Duke Institute for Health Innovation (DIHI) gave me the unique opportunity to experience multiple sides of healthcare—business, technology, clinical, and policy. This past year, I learned about qualitative methods through the Health AI Partnership (HAIP) and quantitative methods with Pythion (page 43), a model that predicts post-operative surgical complications. As a member of the HAIP team, I had the opportunity to listen to interviews from operational, clinical, technical, and administrative health systems leaders across the US and help craft best practices for how health systems can integrate AI into health care safely and effectively. This tremendous opportunity enabled me to more deeply understand the numerous factors that affect model development—defining the problem, change management, etc. so that I am more prepared to work on the technical side. Working on Pythion has given me incredible insight into the numerous deliberations that precede model development and implementation. I learned how to do due diligence in developing something that will affect real people.

Entering my junior year, I am incredibly grateful to be a part of the DIHI team. Not only is DIHI a special place of inspiration, learning, and innovation, but it is also a wonderful community. Everyone across the team has unique experiences and varying backgrounds. Working with this team has contributed so much to my growth during my time in college, and I greatly look forward to working with this team over the years to come.
Predicting Behavioral Emergencies in the Hospital

Problem
Nearly half of all hospitalized patients in the US have comorbid psychiatric disorders.¹ Behavioral health crises are associated with most patient-perpetrated assaults and physical threats.² However, the current practice for patients experiencing a behavioral emergency is reactive rather than proactive, focusing more on containing the patient and suppressing violence rather than improving treatment and outcomes. Containment-focused responses to behavioral emergencies use chemical sedation and physical restraints, which contribute to longer lengths of stay and increased rates of injury and nosocomial infections for patients.³ There is a call to health systems to progress the standard of care for behavioral emergencies by forming trained behavioral response teams (BRTs).⁴

Solution
To address this gap, Duke University’s Departments of Medicine and Psychiatry and the Duke Institute for Health Innovation (DIHI) formed a transdisciplinary team to develop a machine-learning model for real-time detection of behavioral emergencies using electronic health record (EHR) data. Our goal is to support an established BRT’s proactive monitoring of and intervention in patients’ psychiatric destabilizations to improve patient care and safety for clinicians.

We used data from 310,873 inpatient encounters for 179,416 unique adult patients at three Duke University Health System (DUHS) hospitals from 1/1/2017 to 12/31/2021. We excluded patients under 18 years old at the time of admission. A behavioral emergency outcome was defined as the occurrence of any of three interventions during the encounter while the patient was on an intermediate or step-down unit (excluding emergency and perioperative units): a violent restraint order placed, a medical hold order placed, or a non-violent restraint order placed plus the ordering or administration of antipsychotic medication. We compared the outcome to a limited data set of ground truth clinical note documentation of a behavioral emergency or consultation or dual physician adjudication. We designed the machine learning model to predict the first occurrence of a behavioral crisis during hospitalization. Model inputs included historical event data and 31 temporal data inputs related to patient assessment scores, medication ordering and administration, and toxicology screening labs. We trained the model using a light gradient boosting machine (LGBM), with 70% of the data used for training and 15% each for validation and testing. We evaluated the model performance based on sensitivity, specificity, positive predictive value (PPV), and area under the curve (AUC) of the receiver operating characteristic (ROC) curve.
Outcome
Of the 310,873 encounters for 179,416 patients, 5,261 encounters (1.69%) for 4,759 patients (2.65%) met the behavioral emergency outcome definition during the encounter. A May-November 2022 comparison of the behavioral emergency outcome (n=203) with behavioral emergency clinical note documentation (n=506) and dual physician adjudication (n=31) yielded a precision of 90% and sensitivity of 36% for the behavioral emergency outcome. (See model results in Table 1). The clinical workflow applies the real-time behavioral emergency outcome, the predictive model, and secure paging technology to alert the BRT nurse, prompting a chart review and bedside assessment as needed to identify and intervene in imminent behavioral emergencies. We set up the behavioral emergency outcome label to run in real-time, with a bed view visualization tool and event paging system in place. Our next steps are to complete temporal validation of the model and integrate it into the BRT team workflow (Figure 1). We aim to evaluate the solution’s potential to decrease the use of violent restraints, antipsychotic medications, physician hold, suicide precautions, length of stay, and incidence of patient violence.

Next Steps
We are piloting the behavioral emergency prediction solution in the Duke University Hospital (DUH) inpatient units to evaluate its performance. The clinical workflow utilizes push notifications and snoozing logic to alert the BRT team when a patient is at high risk of meeting or has met the real-time behavioral emergency phenotype. After the pilot period, we will evaluate the impact of the solution on behavioral emergency event rate, use of restraints and antipsychotic meds, patient length of stay, and employee injury rates.

Team
Cara O’Brien, MD
Will Ratliff, MBA
Dustin Tart, BSN
Catie Dunn, BSN
Kristen Shirey, MD
Kelley Kester, DNP
Lisa Pickett, MD
Linda Tang, BS
Shems Saleh, MSc
Mark Sendak, MD, MPP
Michael Gao, MS
Suresh Balu, MS, MBA

Project in Brief
Behavioral emergencies by patients during admission represent a growing problem for health systems. At Duke, there is no good way to identify patients at risk of such emergencies to mitigate that risk. To help our clinical colleagues who respond to these emergencies, including our behavioral response teams (BRTs), our team developed a machine-learning solution that alerts clinicians of high-risk patients at the time of their inpatient admission and 24 hours after admission. We have implemented the model in real-time and are evaluating its performance using metrics such as behavioral emergency event rate, use of restraints and antipsychotic meds, patient length of stay, and employee injury rates.
**Academic Output**

Our work was submitted to, presented at, and selected as the winner of the GSA Applied AI Healthcare Challenge, which took place in May 2023 (virtual). GSA awarded our team a $25K cash prize, which we are using to support the project.

Our work contributed to an abstract, poster, and spotlight presentation at the Machine Learning for Healthcare Conference, which was held in August 2023 at Columbia University in NY.


**References**


<table>
<thead>
<tr>
<th>Model</th>
<th>Description</th>
<th>AUC</th>
<th>AUPRC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Admission</td>
<td>At the time of inpatient admission, what is the patient's likelihood of having a behavioral emergency during the encounter?</td>
<td>0.846</td>
<td>0.107</td>
</tr>
</tbody>
</table>

Table 1. Model Performance
I joined the Duke Institute for Health Innovation (DIHI) as an intern in the summer of 2022. Although the internship is short, my experience with DIHI has significantly reformulated my way of approaching healthcare problems with Machine Learning (ML) techniques.

Before joining DIHI, I have been working in the field of ML for healthcare as a Ph.D. student for many years. I have been reading many works that utilize ML frameworks to facilitate decision-making for healthcare workers. Although these works demonstrate that ML has the potential to bring benefits to the healthcare domain, I found myself doubting the applicability of these approaches in real life. Additionally, I started wondering if the problems researchers are interested in are clinically meaningful and if solving these problems will make a real impact in the healthcare domain. I decided to join DIHI to find my answer.

I joined the "Development of a Machine Learning Model for Early Detection of Pediatric Sepsis" project. Although my main focus was model development, I was involved in the entire workflow, from data preprocessing to model deployment. I was fortunate to see how the DIHI team makes great efforts to bring various stakeholders, such as clinicians, nurses, and patients, into the process from end to end. I realized that quantitative metrics should be one of many tools to evaluate project success. A good project should tackle clinically meaningful problems, capture realistic factors closely related to the downstream task, involve numerical tools and end-users in careful assessment, and smoothly conform to the healthcare workers' daily workflow. This experience with DIHI has significantly changed my view of good research work, which greatly impacted my future career as an ML researcher.

Another valuable thing about the DIHI team is that with the focus on building end-to-end frameworks to facilitate healthcare workflows, the team brings in people from diverse backgrounds and forms a collaborative culture. I gained a lot of insights and valuable skills from different team members, such as how to properly extract and store data, communicate with clinicians efficiently, and write reusable codes. I am grateful for what I have learned over the summer and will carry those valuable lessons to my future work.
In today’s rapidly-evolving healthcare landscape, artificial intelligence (AI) is now an accessible differentiator for driving innovation and improving patient outcomes. One remarkable tool at the forefront of this revolution is ChatGPT, an AI language model with diverse applications in healthcare. As a project manager and data scientist working on innovation in the healthcare provider space, I have found ChatGPT immensely helpful in my daily work. Specifically, it has augmented my speed and accuracy in Python coding, research synthesis, and communication for project implementation:

1. **Expediting Python coding** for retrospective analysis: Retrospective data analysis forms the bedrock of our clinical decision support tool development and quality improvement initiatives at Duke Institute for Health Innovation (DIHI). However, curating and analyzing vast volumes of complex EHR data can be challenging, especially under the time constraints of a 12-month project timeline. Enter ChatGPT. This intelligent language model is a valuable assistant, helping me create efficient Python functions and resolving errors so that I no longer need to schedule time with a DIHI data engineering colleague or sift through online coding forums to find answers. It also generates associated steps to test the code it produces, which I can opt to include in my code or learn from to help improve my coding “best practices” knowledge base.

2. **Synthesizing research** on clinical conditions and outcome labels: I need help to keep up with the ever-expanding body of healthcare research across the half-dozen projects I’m managing at any point in time. ChatGPT offers a novel solution by synthesizing research papers on specific clinical conditions and outcome labels (clinical, operational, etc.). I can utilize ChatGPT to quickly summarize and extract valuable insights from a particular article or an extensive corpus of literature. This utility accelerates the process of evidence-based decision-making. It helps me stay updated on the latest advancements to consider as we refine our approach to a project.

3. **Iterating communication and education materials** for innovation project implementation: The successful implementation of healthcare innovation projects requires effective communication with leadership and education sessions with front-line users. ChatGPT is helpful in this project phase by allowing me to draft and iterate on associated written materials quickly. ChatGPT serves as an AI-driven content creator and proofreader, from project summary one-pagers to stepwise training documents. Its natural language processing capabilities help me rapidly create understandable, accurate content that aligns with the target audience, thereby fostering increased adoption of innovation initiatives.
ChatGPT has saved me approximately 30 work hours since I began using it this past spring. However, it is essential to acknowledge that while ChatGPT is a powerful tool, its implementation in healthcare requires careful consideration. Data privacy and security are paramount, and we must always protect sensitive patient information. Ensuring transparency and equity in AI-generated outputs is critical to building trust among stakeholders and ensuring the responsible use of AI in healthcare. ChatGPT has proven to be a game-changer in healthcare innovation and data science, at least in my experience. It has empowered me to achieve greater efficiency and precision in my work. As the technology evolves, infusing AI systems like ChatGPT into healthcare innovation will lead to transformative advancements in our industry and hopefully shape a brighter and healthier future for us all.

A third-year scholarship at the Duke Institute for Health Innovation (DIHI) allowed me to work outside traditional healthcare research and focus on innovative machine learning (ML) projects. The summer of 2022 kicked off my second year with the DIHI team, and at that point, I felt more confident and comfortable in the skills I had been developing. In my first year, I spent time with this team learning how to code in Python, while this year, I could apply much of what I had learned. My primary project was the “Development of an ML Model for Prediction of Textbook Outcomes.” However, I could also work on and support other projects, which allowed me to explore ML in different clinical projects.

This year I continued to utilize the skills I learned last year to work on data quality, cleaning, and exploration on various projects. I expanded my skills this year by working closely with experienced data scientists to learn how to develop a predictive ML model for my primary project. I was fortunate to have presented our work internally at Duke Research events and at a national surgical conference. After two years of working with and learning from this incredible DIHI team, I’ve developed programming skills and learned how to apply innovative thinking in healthcare settings. My time with DIHI has been an incredibly unique and transformative experience. My research and career interests now lie heavily in ML and predictive modeling to help solve problems in medicine and health care. The mentorship provided by the entire DIHI team has left me confident and prepared to continue exploring the world of ML in medicine.
eProvider Using the Electronic Health Record

Problem
The Electronic Health Record (EHR) has created substantial changes in the practice workflow of the primary care provider, both for direct and indirect patient care. Over the last decade, the expansion of patient health portals has been rapidly expanding. These portals are embedded in the EHR, allowing patients to interact with their healthcare team and securely view their health information. There is pressure directly from the health system and indirectly from patients to respond nearly immediately to patient messages.

Marketing this continuous access to clinicians has led to unintended consequences. The marketing of digitalized health sets unrealistic patient expectations for 24-hour per day, seven days per week access to their primary care provider. Consequently, patients send many inappropriate patient portal messages to providers and have inappropriate expectations for their management. For example, although the portal is not to be used for urgent matters, patients continue to use it for urgent health concerns. These messages can range widely in complexity from the very straightforward that can be managed by other staff (requiring no provider input) to the complex responses requiring a scheduled clinical encounter.

High expectations and complex needs increase the messages sent and perpetuate a burnout cycle. Providers can only complete some patient treatment, patient encounter notes, other EHR duties, and digital communications during scheduled business hours.

Figure 1 – eProvider workflow for patient
Hence, they spend their evening and weekend off-hours responding to patient portal messages to keep up with health system pressures. Off-hours responses reinforce patients’ expectations of continuous access to clinicians. Additionally, message volume for providers has increased over time, often resulting in increased indirect unreimbursed patient care work. This work is often done during the provider’s “free time” and can lead to increased burnout.

We needed more standardization in managing patient portal messages. Our solution was to create a new “eProvider” clinical role: a liaison for reimbursed digital clinical care and organizing indirect duplicate work within the Electronic Health Record (EHR). The eProvider (eP) would address patient portal messages using a novel workflow without additional hiring. The Duke Institute for Health Innovation (DIHI) funded staff and provider time to create this clinical innovation.

**Team**
Maribeth Kuntz, PA-C  
Megan Arguello, RN  
Brian Halstater, MD  
John Ragsdale, MD  
Anthony Viera, MD  
Will Knechtle, MBA, MPH

**Project in Brief**
Patients have growing demands for digital healthcare experiences and continuous electronic access to clinicians. These demands overwhelm clinicians and raised a need for more standardized approach to the management of patient messages in the EHR. Our solution was to create a new ‘eProvider’ clinical role: a liaison for reimbursed digital clinical care and organizing indirect duplicate work within the EHR. Post-implementation survey results showed that patients and providers valued the protected time for digital work and wished to sustain this new role.
The Process
We recruited ePs from among our faculty, developed a schedule distinguishing their digital from in-person time, and incorporated this among the department’s electronic scheduling templates. We trained nurses and allocated a portion of their time toward determining whether patient concerns are appropriate for a video visit with a designated provider. We worked with patient access employees to train them to convert the nurses’ designations for eP need into eP virtual visits. We centralized faculty and staff personnel and schedules so they could spend daylight hours responding to patients’ digital needs. Finally, we kicked off eP care in July 2022 and have sustained it.

Impact
We asked patients for verbal feedback via the message portal system, and we surveyed physicians about EHR time and burnout. Patients reported satisfaction with getting needed meds promptly via the patient portal. Providers were pleased. Surveys captured that provider hours managing patient advice request messages per clinical day for patients went down from 1.91 to 1.50 hours per day. We asked providers how many additional daily hours they would need to feel like clinical care and patient-digital-response work were complete. This count decreased from 2.72 to 2.13 hours from baseline to pilot end. Providers across the department, not just ePs, felt they gained 49 minutes daily due to optimizing Patient Advice Requests.

Academic output
Presentation at a national conference, STFM Annual Spring Conference 2023

Two pending publications in Family Practice Management Journal, late 2023: “Utilizing a Nurse Triage Model to Address Patient Concerns via Electronic Medical Record” and “The eProvider: A Novel Approach to a Modern Age-Old Problem.”

Next Steps
We are involved in MyChart Task Force and other work groups across Duke Health, looking at the best ways to implement in different fields and specialties of medicine.

Figure 2 – Volume of patient messages over time (June 2019 – January 2022)
STAFF PERSPECTIVES

Introducing Bed Watch

Matt Gardner BS, Will Knechtle, MBA, MPH

In healthcare, being ahead of the curve means being in the right place at the right time with the right interventions. At the Duke Institute for Health Innovation (DIHI), we aim to practice human-centered design by proactively engaging care providers from problem identification to solution implementation. Since the dawn of Sepsis Watch, we have striven to identify sets of observable characteristics with technology near the same time of their observation by a human eye (e.g., on page 40, read about finding readmissions when they arrive rather than months later, when they are billed). Most recently, we have developed methods for identifying the current place of the patient. The user interface we have pioneered to help interventions occur in the right place and at the right time is “Bed Watch.”

Today, “Bed Watch” describes Tableau dashboards that display a box for each hospital bed in a health system, color-coded according to match criterion, and can be highlighted to see the identity and additional information about the patient in the bed. “Bed Watch” was first designed to present the Patient Flow Coordinator team with Emergency Department (ED) capacity, flow, and admission likelihood. A few months later, it helped small groups identify the available beds for COVID-19 patients and locations of patients with COVID-19, flu, Respiratory Syncytial Virus (RSV), or using equipment such as Extracorporeal Membrane Oxygenation (ECMO) or Mechanical Ventilators.

The new value “Bed Watch” immediately provided inspired our 20th Impact Volume (aka “Issue”)’s cover design. Since then, Bed Watch has helped identify patients with outlier Length of Stay (LOS), ready for discharge, sepsis, high risk for mortality, ready for hospital care at home, and more.

Admittedly, the “Bed Watch” visual is not a new concept. DIHI’s Sepsis Watch’s red and black cards idea inspired the colored bed rectangles. An Epic user can identify a patient bed location after several clicks or can view patient location in tiles or rows with General Electric (GE) Healthcare’s “Hub.” However, to our knowledge, identifying patients’ places, care timing, or care procedures can be done by few other products at such a low cost. “Bed Watch” can be implemented quickly with only a Tableau license and access to an Electronic Health Record (EHR) and Monorail (page 68).

Bed Watch has become an effective real-time tool when paired with a clean pipeline of EHR data, subsequent clever data engineering, and proactive care providers. In 2023, we improved the Bed Watch architecture to manage a growing volume of patient risk model results and phenotypes (observed conditions).
Bed Watch Architecture Overview

Before 2023, the need for timely access to a wide breadth of clinical data and model results, such as active infections, anticipated discharge destination, and mortality risk - presented a significant technical challenge. Extracting a consolidated view of this information for admitted patients required complicated Structured Query Language (SQL) joins and aggregations that spanned many tables and databases. Complex user demands, materials, and code caused the initial implementation of the Bed Watch dashboard to be resource-intensive and response times to be unacceptable – minutes in some cases.

The DIHI data engineering team needed an organized and optimized data mart to provide quick snapshots of any admitted patient’s clinical status. Previously, DIHI housed the Bed Watch dashboard source data in the DIHI Pipeline databases (Active and Archive) and several model results databases. The goal was to update this data mart every five minutes – so the Bed Watch dashboard could deliver “near” real-time insight.

Extraction the source data and updating the data mart every five minutes required a parallel processing approach that an engineer could scale up as they identify new data subject areas for extraction and inclusion in the Bed Watch dashboard.

The following narrative and diagram depict the current architecture we implemented that ensure the Bed Watch data mart is updated every five minutes. The Bed Watch dashboard now consistently presents refreshed results in seconds.

Bed Watch Data Mart ETL
(Extract, Transform, Load) Data Flow

- Every five minutes “near” real-time ADT data is extracted to determine admitted encounters/patients and their current bed: (location → department → room → bed).

Figure 1
For each data subject area of interest, such as active infections, the list of encounters is split into equal chunks and a message for each chunk is placed into a work queue.

Workers process the queued message. For example, they extract the active infections for the chunk of encounters and update the Bed Watch data mart. These workers run in parallel and can be scaled up – the current implementation has 16 workers.

This parallel worker paradigm allows simple, less resource-intensive SQL queries to extract data in parallel – and to load results into the Bed Watch data mart in parallel. The parallel extracting and loading minimized the window required to refresh the data mart every five minutes.

Conclusion
We foresee Bed Watch becoming a standard tool Duke Health employees use to ensure they meet daily access, quality, and cost goals. Any bedside care provider can view analytics at the patient level at their current time and place while, in one view, maintaining the perspective of the floor, service line, and hospital. A manager can view their whole area or unit for a quality assessment but drill down to the beds driving the failure or success. Any employee could view benchmarks set from the past and see how the last five minutes of work today will compare to them. The innovative “Bed Watch” dashboard makes it possible to understand the current state better and implement improvements on the spot.
In recent years, gun violence has increased across the United States (US), particularly in communities of color. In 2022 and Durham alone, over 400 victims of gun violence were treated within the Duke Hospital system, with a large proportion of these victims being young, black men. Across the US, programs have been growing to address the rising epidemic of gun violence, including collaborations between health systems, government agencies, and community groups. Often missing from these programs is attention to the needs and experiences of survivors of gun violence.

From August 2022 to December 2023, we operated a program called Prescriptions for Repair (P4R). Through a series of structured listening sessions using a restorative justice framework, trained community-based facilitators helped survivors of gun violence tell their stories through a non-judgmental process. In doing so, gunshot survivors defined their “prescription for repair” to continue healing. They also provided a “prescription for repair” for the community to address the harm from gun violence. This program was a collaboration between the City of Durham, Duke Medical School, North Carolina Central University (NCCU), and Restorative Justice Durham.

Over this one-year pilot, we hired a program coordinator, enrolled 30 gunshot survivors or their families, trained 28 community-based facilitators, and conducted 56 listening sessions. We coded all transcripts from 18 consecutive participants, which included 45 listening sessions and ran over 70 hours. We coded over 400 separate statements and summarized these statements within a restorative justice framework. The extended length of our listening sessions and depth of engagement between facilitators and participants far exceeds existing studies to probe into the complexities of the harm resulting from gun violence. Some of the major lessons learned from our program are summarized below:

**The inherent value of listening to gun violence survivors**

One of the most important lessons we learned from P4R is the inherent value of listening to survivors of gun violence using a non-judgmental process. Almost all program participants expressed gratitude for the program, and several respondents were unable to socialize, work, or even leave their homes following the original trauma until participating in P4R. These experiences attest to the high burden of mental and physical sequelae following gun violence.
Although our program facilitators were not practicing therapists, this program attests to the therapeutic nature of a structured listening program for people impacted by gun violence. Social support services for survivors of violent injury across the US remain limited. Most public mental health resources go to treating severe mental illness, while the survivors of violence and their loved ones often lack mental health and social support.

Restorative justice practices date back to early Abrahamic religions, and contemporary restorative justice programs are often aligned with faith-based groups to support programs directed toward gun violence. With most Durham residents identifying religion as important to their lives and a significant portion attending religious services, partnerships incorporating faith-based groups, restorative justice-based programs, government, and academic experts can optimally leverage the shared expertise of all partners.

**Race impacts the experiences of gun violence survivors**

Almost all participants in the P4R program commented on the central role of race in their personal experiences of gun violence, how they view community responses (or lack thereof), and the challenges Black communities continue to face in addressing the harm from gun violence. These voices mirror expert opinion, documenting high violent offending rates among Black males in urban contexts. Respondents expressed the importance of structural disadvantages and residential segregation in Black urban communities as drivers of gun violence. Respondents also spoke of the role of gang activities and “street culture” in perpetuating the cycle of gun violence. Finally, respondents voiced both victimization and a sense of inevitability in breaking the cycle of gun violence. In addition to how race impacts individuals following gun violence, several respondents cited the need for broad healing of the greater Black community.

---

**Team**

Henry Rice, MD  
Lorraine Graves, PhD, MSW, LCSW  
Marcia Owens  
Christopher Solomon, MSW  
Azmen Johnson  
Will Ratliff, MBA  
Will Knechtle, MBA, MPH

---

**Project in Brief**

Survivors of gun violence have long-term healing and community needs after facing the societal challenges of gun violence. Other programs have not focused as much on learning from the experiences of the victims. Prescriptions for Repair (P4R) is a program that supports structured listening sessions using a restorative justice framework for trained community-based facilitators to help survivors of gun violence tell their stories through a non-judgmental narrative process. Our experience has demonstrated that low-cost, trained community facilitators can help build resilient health systems.
The integration of gun violence into daily life among many respondents reflects what is recognized as “community trauma,” which is characterized by a breakdown of social networks, relationships, and positive social norms across the community. Many P4R respondents cited the importance of historical violence in driving gun violence, such as slavery, economic inequities, and predatory housing practices. Recognizing how views of race are deeply integrated with gun violence within communities can support implementing effective programs to address the community’s harm from gun violence.

**The focus should be on children and youth**

When asked to identify community-level strategies to address the harm and reduce the rates of gun violence, many respondents emphasized the need for increased resources for family support and youth development. These responses mirror those of experts who suggest that when strong family units or community organizational infrastructure are lacking, violence and trauma have a more profound impact on individuals and communities.

Other suggestions to improve youth support, as voiced by respondents, also align with prior research, including reclaiming public space to reflect community culture and access to positive family role models and other examples of healthy behaviors and relationships. Participants frequently voiced that improved economic opportunities for young adults are critical to breaking the cycle of gun violence and healing from community trauma. Other suggestions from respondents include expanding resources to increase the number of young people who complete high school, attend college, and job training for non-college-bound youth.

**Sustainability requires new thinking among public, private, and community partners.**

As P4R is a pilot program, we recognize the need to implement sustainable programs to address the harm of gun violence within existing social and public health networks. We are discussing with local government leaders to develop a government-based Office of Survivor Care (OSC) for gun violence victims, survivors, and their families. The OSC would represent a transformative community response to gun violence:

1. This center would represent a historic opportunity to provide survivor-centric resources to augment existing law enforcement systems.
2. The OSC would serve as a community health model that begins with listening to the experiences of those most affected by gun violence to strategically develop programs to address the personal and community harm of gun violence.
3. This platform would support partnerships between public, academic, and private groups that have traditionally not worked together to address public health challenges.

For the P4R program, we learned many practical lessons through operating a program with partners from a historically Black university (NCCU), a research university (Duke), the City of Durham, and community partners.

**Academic Output**

Grant in development: Duke Endowment grant ($620K)

Research in action:

1. Six students from Duke and NCCU overseeing multiple subprojects to analyze gun survivor experiences qualitatively
2. Standard monitoring and evaluation of program outputs

**Affiliations**

Department of Surgery, Duke University School of Medicine; Duke Center for Global Surgery and Health Equity; Restorative Justice Durham; Department of Social Work, North Carolina Central University; Duke Institute for Healthcare Innovation; Department of Community Safety, City of Durham
I have worked with the Duke Institute for Health Innovation (DIHI) as an undergraduate research intern for the past two years. In 2023, during my senior year at Duke, I worked on a project to predict 30-day readmissions for post-operative patients in the Duke Health system. This research allowed me to complete a senior thesis in the Duke Statistics department and ultimately graduate with distinction. DIHI gave me the unique opportunity to take the lead on the project and tackle the myriad challenges that came our way while receiving continual support and mentorship from my DIHI advisor.

Working with DIHI was one of the most impactful parts of my undergraduate career. I gained valuable research experience, saw how data science projects go from ideation to implementation, and enjoyed afternoons in the DIHI office. I’m grateful that DIHI welcomed me into the team, and I look forward to working with the group again!
Identifying Candidates for PrEP to Prevent Incident HIV in Louisiana and Mississippi

Mark Sendak MD, MPP

Part of the Duke Institute for Health Innovation’s (DIHI’s) mission is to “cultivate a community of innovation and entrepreneurship across Duke University and the health enterprise.” One of the ways this mission manifests is durable changes in how clinical faculty and staff develop, test, and scale their ideas. A hub of innovation at Duke has centered on infectious diseases. DIHI helped launch the Sepsis Watch program in 2016, a project for pre-exposure prophylaxis (PrEP) for Human Immunodeficiency Virus (HIV) in 2017, and in 2019 DIHI and infectious disease collaborators were awarded an internal Duke Center for AIDS Research (CFAR) grant to develop a machine learning model to predict incident HIV. Over two years, our transdisciplinary team successfully developed and validated an incident HIV model, replicating prior work conducted at Kaiser Permanente. The Duke HIV model marked the first time anyone built an incident HIV machine learning model on a Southern U.S. cohort, which represents the geographic region most impacted by the current HIV epidemic.

Building upon this initial work, in June 2022, the NIH awarded our transdisciplinary team an End the Epidemic R01 grant from the National Institute of Health (NIH) to adapt the incident HIV model for use in Louisiana and to conduct a prospective, multi-site clinical trial. This grant, led by co-Principal Investigators Meredith Clement and Lance Okeke, will be the first time DIHI will refine, validate, and implement a machine learning model entirely for external organizational contexts. In addition to the incredible opportunity to learn how to conduct highly technical projects with multiple external sites, this project also represents a unique opportunity to impact health inequities. Black persons are disproportionately affected by HIV, making up 42% of new infections and only 13% of the US population. And more than half of new infections occur in the Southern US. If we can accurately identify patients at risk of incident HIV and help those patients initiate PrEP, there’s an incredible opportunity to reduce HIV transmission.

While we’re still in the early stages of the project, we’ve already encountered several challenges that differentiate HIV modeling from other acute or chronic conditions. First, the incidence of HIV is very low, meaning model development requires large cohorts of patients. The Kaiser Permanente model used a cohort of 3.75 million patients with 784 incident cases to build a model. We built the Duke model on a cohort of 1 million patients and 162 incident cases. The low incidence of HIV also means that models with strong sensitivity have meager positive predictive value. For example, the Duke model had a PPV of 0.38% at a sensitivity of 48.89% and a PPV of 2.77% at a sensitivity of 26.67%. Many high-risk patients would need to be initiated on PrEP to prevent a single incident case of HIV.
The second major challenge with HIV modeling is the significant number of undetected cases. Of the 1.1 million people with HIV in the US, an estimated 13% are undiagnosed. This means that any model trained on existing data will almost certainly perpetuate bias, especially for disadvantaged subgroups more likely to have undiagnosed HIV. Data scientists cannot precisely measure model performance on historical data because a portion of patients identified as high risk who appear not to have incident HIV (i.e., false positives) would, in fact, be true positives if all positive cases were known. This high rate of undiagnosed cases also means that implementation of an incident HIV model must also broaden access to diagnostic testing. Before any patient identified by the model is initiated on PrEP, HIV testing must be completed to confirm negative status.

The third major challenge with HIV modeling is the different risk factors and disease dynamics between men and women. For example, the Kaiser Permanente model built in Northern California did not identify a single positive HIV case in females. The Duke model could identify positive HIV cases in females, but accuracy was much lower than in the general population. This gender-specific disparity will need to be accounted for in the model features included in the model and optimizations to improve performance among females.

Lastly, there is a significant stigma associated with HIV. After the Kaiser Permanents model was published, the New York Times ran a story titled “Would You Want a Computer to Judge Your Risk of HIV Infection?” Modeling the risk of a sexually transmitted disease like HIV feels invasive and intrusive in ways that inherently differ from modeling the risk of sepsis, chronic kidney disease, or pulmonary embolism. The clinical workflow for model integration will need to be carefully designed to ensure that patients feel comfortable discussing HIV risk, and clinicians will need to be trained to have these conversations.

Our team looks forward to working closely with clinical collaborators over the coming years to navigate these challenges. We are thrilled to be able to work with Franciscan Missionaries of Our Lady Health System (FMOLHS) and LCMC Health System to design, develop, and implement machine learning models and clinical workflows to identify patients at high risk of HIV for PrEP. Through this collaboration, we hope to directly impact the Southern HIV epidemic and build new capabilities to navigate bias in data and address health inequities in diverse geographic settings.

References
Implementation and Evaluation of the Duke Health Patient Initiated Note about Goals (PING)

Problem
Only 25% of Duke Health patients near the end of life have had documented conversations with healthcare providers about their goals of care. Consequently, plans of care, including plans for end-of-life care, are being developed and implemented without recorded input from patients and their families. A large body of evidence has shown that goals of care conversations are essential in promoting care concordant with patients’ preferences. Communication improves patient and family satisfaction, reduces healthcare utilization, and lowers care costs. In response, Duke Health began a broad mission to promote goal-concordant care.

Solution
Duke Palliative Care, the Duke Institute for Health Innovation (DIHI), and Duke Population Health Sciences formed a team to oversee the implementation and measure the impact of direct-to-patient goal concordant care support. Whereas previous work has focused on clinician-initiated conversations, our project’s overarching aim was to implement, optimize, and evaluate the Patient Initiated Note about Goals (PING), encouraging patients to share their goals with their care team via MyChart proactively.

The PING was developed in late 2021-early 2022 with clinicians and patient stakeholder groups, including Duke Patient and Family Advisory Council, the Duke Patient Education Office, and underrepresented minorities in the community. It includes a brief three-minute survey on patient goals using multiple-choice and free-text options and is written at a fifth-grade reading level. Categories focused on important goals for patients, such as: “not being a burden on others” and “living as long as possible,” and answers included “not very important,” “somewhat important,” and “very important.”

The project began with a go-live: Duke Palliative Care, Duke Population Health Sciences, and Duke Health Technology Solutions implemented the PING in the spring of 2022. A “silent BPA” in Maestro Care began identifying candidates to receive the PING based on their comorbidities and recent clinical events (e.g., a recent hospitalization). This instantiates a MyChart message to the patient, encouraging them to complete the PING. Our project team began analyzing response data. After a review of preliminary data, however, we observed that the direct-to-patient messaging was not having the scale of impact we had hoped, with minimal responses from direct-via-MyChart outreach. So, starting in January 2023, we partnered with the Duke Primary Care (DPC) Population Health Nurse (PHN) Team, led by Beth Soule and Sarah Tucker, to incorporate the PING into their workflow for assessing patients during their annual wellness check visits. Figure 1 shows the workflow used by PHN Team.
Outcomes
We analyzed results from the PING responses since the 2022 go-live (04/01/2022 – 07/15/2023), wherein 678 unique patients responded to the PING via direct-to-patient MyChart message or the DPC PHN workflow. From the period directly after going live (i.e., the direct-to-patient only route via MyChart, 04/01/2022-01/30/23), 279 patients (27.9 per month) responded to the PING.

After the partnership with the DPC PHN nurses began, 400 patients responded to the PING (71.4 per month, 02/01/2023-07/18/2023), more than doubling the direct-to-patient-only route. Our 678 PING respondents were, on average, 61.5 (SD 17.2) years old, predominantly female (67.6%), White (70.6%), and Non-Hispanic (92.8%). It was “very important” for respondents to “not be a burden on others” (85.5%) and “take care of myself than depending on others” (84.4%), and “not important at all” to “avoid travel to and from clinic appointments” (40.9%).

Next Steps
Our next steps are to improve our reach and understanding of Duke patients’ goals of care so that we can best provide concordant care. We will prioritize two objectives: (1) scale our inclusion of the PING across additional wellness visits and other appropriate ambulatory/outpatient settings and (2) refine our understanding of direct-to-patient outreach via MyChart and other platforms. Additionally, we plan to measure the impact of a completed PING on subsequent clinical outcome metrics: hospitalization and Emergency Department (ED) visit rates, palliative care utilization, and Intensive Care Unit (ICU) utilization. We hope to show the positive impact of goal-concordant care the alignment of healthcare utilization with our patients’ goals.

Team
David Casarett, MD (PI)
Will Ratliff, MBA
Beth Soule, MSN
Sarah Tucker, BSN
Sara Gonzales
Jessica Ma, MD
Salam Ibrahim, MS
Devon Check, PhD
David Flores
Suresh Balu, MS, MBA

Goals of care conversations are often initiated and documented by clinicians. Prompts to patients can encourage documentation of patient goals and values. A patient-initiated survey on goals in the electronic health record is feasible and can guide goal-concordant care.
### Population Health Nurse (PHN) pre-work for Annual Wellness Visit (AWV)

<table>
<thead>
<tr>
<th>MONDAY</th>
<th>TUESDAY</th>
<th>WEDNESDAY</th>
<th>THURSDAY</th>
<th>FRIDAY</th>
</tr>
</thead>
<tbody>
<tr>
<td>Send MyChart message and PING questionnaire 1 week prior to visit</td>
<td>Pre-work normal</td>
<td>Send MyChart message and PING questionnaire 1 week prior to visit</td>
<td>Pre-work normal</td>
<td>Send MyChart message and PING questionnaire 1 week prior to visit</td>
</tr>
</tbody>
</table>

### PHN workflow with patient during AWV

1. Review any previous PING answers in ACP activity
2. Review and complete PING with patient, determine next steps (e.g., ACP visit)
3. If patient agrees, add to list for follow up call from PING research team
Patients with Opioid-Induced Respiratory Depression (OIRD) benefit from increased monitoring while taking prescribed opioids. A multidisciplinary team developed the Duke Opioid Sedation Assessment (DOSA) model – led by Duke’s Safe Opioid Prescribing Committee.

In the summer of 2022, Duke Health leadership tasked the Duke Institute for Health Innovation (DIHI) engineering team with transitioning the trained model into a real-time operational setting for silent evaluation.

For training purposes, complex Structured Query Language (SQL) developed by Duke Performance Services was utilized to extract the historical cohort and model features from Epic Clarity. This included SQL to calculate Oral Morphine Equivalents (OME). To remain consistent with the training data, we had to ensure new opioid administrations were subject to the same OME calculation rules. To do this, we replicated the Performance Services tables in a Clarity tablespace dedicated to the DOSA project. We re-created these OME calculation tables each morning using the original OME SQL in an asynchronous update process.

Much of the technical challenge was translating this retrospective SQL used for training to a real-time data extract equivalent. In addition, the real-time model needed to run on the cohort of admitted patients every hour, generate a DOSA risk score, and push those results to an Epic flowsheet. The DOSA score is the result of an algorithm that applies a weight multiplier to each derived patient feature and adds them together:

5 * Lung/Cardiac History DX +
2 * Renal/Hepatic History DX +
7 * Surgery in Last 48-Hours +
8 * Previous Narcan Rescue +
4 * Polypharmacy with Opioid +
2 * Opioid Tolerant on Admission +
Age > 65

For example, if an admitted patient had surgery within the last 48 hours – the algorithm would add a value of seven to their DOSA score.

As can be seen from the algorithm, both historical and real-time data are required to compute the DOSA score. The DIHI Data Pipeline consists of an Active database and an Archive database that share a common data model – this data pipeline serves as the primary data source for model inputs.

Critical to the success of any deployed model is the ability to evaluate the model’s performance on prospective data that is gathered in near-real-time. Contemporaneous data can be very different from historical data, especially in healthcare. For example, suppose a dose of medication administered is updated from one Unit to two Units an hour after someone gives the initial dose. In that case, we must ensure that the real-time model only sees the one Unit administration during the hour between the initial dose and change to two Units. Any retrospective model evaluation must prevent penalizing the model for data that it never sees. To prevent this issue, we store all feature inputs of the DOSA model in a companion database as a JavaScript Object Notation (JSON) payload associated with the model results for a patient.
<table>
<thead>
<tr>
<th>Column</th>
<th>Value</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>model run on</td>
<td>2022-10-20 14:01:00.103626</td>
<td>Timestamp of when score was computed for patient</td>
</tr>
<tr>
<td>encounter id</td>
<td>99999999</td>
<td>Epic CSN/Encounter</td>
</tr>
<tr>
<td>patient mrn</td>
<td>MRN999</td>
<td>Epic Patient MRN</td>
</tr>
<tr>
<td>dosa score</td>
<td>12</td>
<td>The DOSA risk score</td>
</tr>
<tr>
<td>dosa features</td>
<td>{“opioid tol”: 0, “over_age_65”: 1, “prey_narcan”: 0, “lung_cardiac”: 1, “polypharmacy”: 1, “renal_hepatic”: 1, “recent_surgery”: 0}</td>
<td>The features used in the DOSA score computation</td>
</tr>
</tbody>
</table>

Figure 1

![Diagram of DHE OpenShift](image1)

Figure 2

![Diagram of DHE OpenShift](image2)
We can quickly lookup a model result and see what features passed into the model that led to the derived score. The following is an example record from the model results database.

This feature storage also allows us to evaluate variations of the DOSA model scoring system on the prospective data and to compare those variations to the live model results in an apples-to-apples evaluation. Suppose we wanted to update the contents of the diagnosis code list that qualifies as Renal/Hepatic. In that case, we can do so without altering the live model and still get comparable results from our test model variant using “as seen at run time” data.

The DOSA model runs every hour and results for each patient are stored in the model results database and recorded in an Epic flowsheet. If the DOSA score exceeds a specified threshold, the supporting features are recorded in supporting flowsheets as well to provide additional context for clinicians. The flowsheets are recorded via an Epic Application Programming Interface (API). The following diagram conveys the high-level architecture of the DOSA real-time model. The model runs in the Duke Health Enterprise (DHE) hosted RedHat OpenShift (Kubernetes) cluster and scheduling, logging of model executions is managed using DHE hosted Apache Airflow.

On March 20th, 2023, I started working as a Data Scientist at the Duke Institute for Health Innovation (DIHI). Before joining DIHI, I was privileged to be in many positions with diverse responsibilities. Moving to DIHI allows me to explore and hone in on the intersection of my passion, skills, growth, and impact I seek next in my career journey. I joined DIHI because I’d like to use health data and clinical expertise to explore and learn how to incorporate data-based solutions in a health system. Ideally, I will implement solutions that add insightful value to clinical workflows while equally benefitting patients. I will be in an environment that facilitates my technical growth in a positive culture and allows me to investigate better to understand the projects at hand and the proposed solutions. (Contributing to the research community is welcome and encouraged). I joined to be part of a collaborative team that includes leaders and managers, clinical leads, engineers, solutions architects, and data scientists working together towards a common goal. I aimed to be in an environment with the infrastructure to enable data access for exploration and development and to integrate the solutions into the system. Importantly, DIHI management practice aims to build buy-in and support from the solutions’ users; this enables our team to improve the solutions over time-based on feedback, internal audits, and evaluations. This institute will provide the environment I’ve been looking for. After reading articles published by DIHI and a job description for my current role, I envisioned this. I gained confidence in this during the interview process. My first months have given me a better understanding of the structure, operations, technical infrastructures, and, most notably, the people at the institute. I have been very fortunate and privileged to have an opportunity at the institute that aligned incredibly well with what I had imagined and wanted to do in this part of my journey.
Improved Perioperative Care Coordination via Enhanced Decision Support Tools

Context
A simplified surgical workflow contains four stages. First, a patient is referred from a clinic to Duke Surgery. Second, the patient visits an outpatient surgical clinic where they are evaluated for surgery. There, or soon after, a surgical case is created, and the surgery is scheduled. Third, a preoperative anesthesia and surgical screening (PASS) clinic with perioperative enhancement teams (POET) communicates with the patient and helps prepare them for surgery. The patient might be referred to other Duke clinics for specialized optimization. Finally, the patient arrives for surgery, receives anesthesia support and undergoes surgery in the Operating Room (OR). After surgery, the patient recovers under the observation of professional care providers and is discharged with a follow-up call or visit within two weeks.

The OR care may be delivered in one of two facility types: the Ambulatory Surgery Center (ASC) or hospital. Generically, ASCs are designed for lower-risk patients, while hospitals are designed for higher-risk patients. The ASC is designed for patients who already have mitigated risk, are likely to have high-quality outcomes, and can be moved in and out of the operating room as quickly as possible without increasing their risk. Note, a patient may be an immediate ASC candidate at the time of case booking or, through optimization by the PASS clinic, may become eligible for the ASC to undergo surgery. ASC patients are typically discharged in less than 24 hours following admission and face lower outpatient charges than they would by staying longer in an inpatient hospital.

Undergoing high-risk surgery or any surgery as a high-risk patient may result in complications with significant ramifications for long-term health. To reduce such complications, preoperative optimization is often necessary. It may reduce the risk of mortality, return to the Emergency Department (ED), stay at Intensive Care Unit (ICU), readmission rate, longer-than-expected lengths of hospitalization, extra clinic visits, and the cost of care.1,2,3

The PASS clinic at Duke University Hospital is the ideal team and physical location to intervene with optimization strategies. The PASS clinic has a unique perioperative and population health program that serves over 90% of all planned surgical and procedural volumes (50,000/year), and 15-20% of PASS patients require additional optimization to be ready for surgery. This PASS-POET model is a differentiator for Duke. The “one pipeline to the OR” model that the PASS clinic facilitates a unique approach to standardized, evidence-based, collaborative, multidisciplinary preoperative assessment and preparation.
Problem
Despite the PASS-POET model, surgery workflow delays and misdirection occur due to incomplete or obscure information at the time of patient presentation in the surgical outpatient clinic. Missing patient information hinders early identification of patient surgical needs of patients and results in increasing the rates of poor outcomes. Our experience has taught us that about 15% of patients miss out on optimization due to poor anticipation of surgical needs.

Poor optimization is visible when high-risk patients show up at an ASC, and low-risk patients with low-risk surgeries are operated on at a hospital. At the time of problem identification, the largest ASC was 80% full, and the second largest was 50-60% full. We believe two problems cause the mismatch. First, a lot of information needed for a complete assessment of ASC versus Hospitalization readiness is not complete or clear, when the patient first arrives or even while being seen in the PASS clinic. Second, incentives for sending patients to the right place and taking sufficient time to screen patients are not aligned. While the PASS team wants to spend time mitigating the risk to the patient as much as possible, surgeons, administrators, and patients are motivated by financial and psychological incentives to move patients toward the operating room as soon as possible, especially given that ORs are critical drivers of health system revenue.

Team
Jeanna Blitz, MD
Michael Lipkin, MD, MBA
Michael Kent, MD
Wendy Webster, MBA
Padma Gulur, MD
Allan Kirk, MD, PhD
Will Knechtle, MBA, MPH
Jennifer Young, MPS, RN
Freya Gulamali

Project in Brief
Current surgical workflows result in lost time for optimization and planning due to incomplete information at the time of patient presentation. We developed a learning decision support algorithm that identifies medical, socioeconomic, and health system factors to help the preoperative anesthesia and surgical screening (PASS) clinic efficiently communicate patient surgical risk and risk-mitigation strategies to Urology, Orthopedic, and Spine Surgery Clinics to support earlier decisions for/against ambulatory surgery center (ASC) care. This tool will result in three primary benefits:
1. Identify patients who are eligible for ASC with enhanced planning and scheduling.
2. For patients who are not eligible, facilitate early evaluation by PASS clinic for optimization that may lead ASC eligibility.
3. Identify patient clusters (volume, characteristics, surgery type) that, while not currently supported in an ASC, can strategically be considered for transition off of the inpatient platform.
Solution

Our primary goal is to create an iterative perioperative learning platform with complete and transparent data to drive efficient preoperative planning and patient optimization. We expect to define criteria for admission to an ASC or hospital and measure the achievement of our goal by monitoring how well we meet those criteria during surgery workflow. Ultimately, we hope to share these learnings across the health system.

First, we developed eligibility criteria for ASC using PDFs of ASC exclusion criteria provided by physician leadership. Then, we developed a prototype platform to screen patients. Using the developed criteria, this platform screened individual patients and placed them into one of the three categories: Red Light, Yellow Light, and Green Light. Patients who failed the criteria would be placed in the Red Light and scheduled for hospital surgery. They would require more time with the PASS clinicians. Patients who may become eligible for ASC with a few checks or optimization would be placed in the Yellow Light. Patients who passed ASC criteria would be placed in the Green Light and only receive a nurse’s screening call. Then, they would be expedited for ASC surgery. This measurement platform added a degree of discernment for patients that preoperative care teams may better optimize for surgery. For example, a patient that is an obvious Green Light for ASC care is unlikely to require several in-person visits at the PASS clinic. The only data entry required was a medical record number, and then the critical information was presented on one straightforward screen.

Outcomes and Next Steps / Plans for Scaling

Our next step is to test the solution at the PASS clinic and among ASC surgeons. We anticipate observing a higher percentage of ASC volume among surgeries. Consequently, we expect a lower average post-operation length of stay and lower readmission rates.
The tool will improve as the development and efficiency of large language models (LLMs) increase, because the current tool is limited to extract structured data. LLMs will allow extraction of deeper insights about the patients comorbid, imaging, and social history that are found in narrative clinical notes.

Finally, the tool will synergize with the Pythion models (see Diffusion & Scaling update) by helping physicians evaluate the Yellow-Light patients. There are many patients for whom, even with full use of clinical notes, the correct destination of ASC or hospital needs to be clarified. These machine learning models will be able to review the entire patient’s chart, accurately organize the otherwise unclear cases into high-, medium-, and low-risk likelihoods and save physicians’ time.

Impact
Integrating a learning decision support tool into the surgical landscape at Duke, where approximately 5,000-6,000 surgeries result in same-day discharges on inpatient platforms, has profound implications for both the immediate and distant future of patient care and system efficiency. In the short term, the tool’s introduction promises to revolutionize how surgeries are planned and executed. While the ASCs are often utilized within 70% of their capacity, rooms still need to be used. Such under-utilization, juxtaposed against the backdrop of ASCs being stewards of efficient, cost-effective, and patient-centered care, underscores the critical need for streamlining patient selection based on comorbidities and other determinants of perioperative outcomes. By facilitating upstream visibility into a patient’s ASC eligibility, this tool addresses the identified gap. This real-time, data-informed insight will empower surgeons to plan more confidently, thus optimizing their utilization. Patients who are not immediately eligible can be directed to Duke North or the PASS clinic, thereby enhancing planning efficiency and precision. Additionally, perioperative teams can now optimize such patients at an early stage so they may become ASC-eligible. In the long term, successfully integrating this tool could reshape the surgical ecosystem. The immediate outcome would be an uptick in the volume of surgeries at ASCs, which directly enhances their operational efficiency. Simultaneously, this would liberate more block time at Duke North, catering to surgeries requiring an inpatient framework. Such ripple effects underscore the tool’s potential to streamline operations and inform future strategic endeavors. As the platform amasses data, it can pinpoint patient and surgical clusters conducive to same-day discharges. This invaluable data trove can pave the way for infrastructural developments and innovations, possibly steering these clusters into an ASC environment. In essence, this is not just a tool. It is a compass directing the future of surgical care at Duke North, making it safer and more efficient.

Academic Output
We have filed Invention Disclosure Forms with the Duke Office of Translation and Commercialization.

References
Timely, accurate data has become increasingly crucial in the ever-evolving healthcare landscape. For a health system to become a proactive, self-sustaining, learning health system providing high-quality care, it must leverage the immense value of real-time data, especially when monitoring surgical outcomes. Traditional methods of reporting healthcare outcomes data, often retrospective and time-delayed, no longer meet the demands of a rapidly advancing healthcare system. Even Generative Artificial Intelligence (AI) may not produce the desired benefits if its data is weeks delayed. In this article, we explore the benefits of real-time data for monitoring surgical outcomes and illustrate its significance for surgeons and the overall bottom line at Duke University Health System (DUHS).

The Duke Institute for Health Innovation (DIHI), comprising statisticians, data engineers, data scientists, solutions architects, social scientists, program managers, and coordinators has developed a groundbreaking approach to real-time tracking of surgical outcomes. By harnessing cutting-edge technologies, we have created a database and user interface that provide up-to-date information on surgical outcomes, such as readmission rates and Emergency Department (ED) returns, even down to the latest five minutes.

Enhancing Surgeon Performance:
Real-time data empowers surgeons with immediate insights into the outcomes of their procedures. By monitoring outcomes on the day of surgery or shortly after, surgeons can proactively identify and address potential complications promptly. Punctual problem-solving improves patient care, boosts surgeon confidence, and fosters a culture of continuous improvement. Surgeons can identify areas for refinement, compare performance against benchmarks, and implement evidence-based practices, leading to better patient outcomes.

Traditional methods of identifying complications relied upon diagnoses that are abstracted from clinical notes often a week after the event, problem lists (see how we developed a system to clean these in the article “Developing a Machine Learning Data...
Quality Assurance (DQA) Framework* from Volume 23), and bills. We first examine the labs, orders, medications, vitals, and structured nursing notes that we update every five minutes (a few items at worst, are updated every 24 hours). Since early 2023, new large language modeling technology has helped us evaluate notes and rapidly identify surgical complications.

Timely Intervention and Quality Improvement
Real-time data allows swift intervention when adverse events occur, minimizing potential complications and reducing readmission rates. Surgeons and care teams can identify patients at risk of readmission or ED visits and intervene promptly, ensuring appropriate follow-up care and personalized interventions. Figure 1, next page, represents software through which a surgeon can be alerted as soon as a patient who had surgery within the last 30 days arrives in an ED or fills a hospital bed. (This contrasts with tracking readmission from the time of surgery and waiting 30 days to collect the total and more days to ensure the billing data was complete).

While identifying a patient after a readmission occurs may not seem helpful, the surgeon can have a conversation with the patient to prevent further readmissions from occurring. The surgeon doesn't have to wait more than a month to be able to receive data about readmissions from their division. These proactive approaches improve patient satisfaction, optimize resource allocation, and reduce healthcare costs associated with preventable readmissions.

Empowering Shared Decision-Making
Real-time data can revolutionize the patient-provider relationship. By providing surgeons with up-to-date outcomes information, they can share decision-making with patients, presenting accurate and current data about potential risks, benefits, and alternatives. For example, the surgeon could talk with the patient about contemporary reasons why patients are re-hospitalized and ways to avoid them. Furthermore, the surgeon can quickly share their lessons with other care providers. For example, promptly identifying a change in readmission rate could lead to an urgent team meeting to discover that a recent process change or environmental shift likely caused the rate change.
**Enhancing Operational Efficiency**

Real-time data enables healthcare leaders and administrators to monitor the performance of surgical units, identify bottlenecks, and optimize resource utilization. By having immediate access to outcomes data, administrators can make informed decisions to improve workflow processes, allocate staff efficiently, and ensure optimal utilization of operating rooms. These improvements directly contribute to the bottom line by reducing costs, increasing throughput, and maximizing revenue generation.

**Conclusion**

As we embrace the digital era, leveraging real-time data for monitoring surgical outcomes becomes paramount. By implementing a state-of-the-art database (“the DIHI Pipeline,” aka “Monorail”) and user interface developed by DIHI, surgeons at DUHS can benefit from immediate insights into patient outcomes, fostering a culture of continuous improvement and personalized care. Real-time data empowers surgeons, enhances decision-making, improves operational efficiency, and (ultimately) leads to better patient outcomes. Embracing this innovative approach aligns with our commitment to excellence and strengthens DUHS’s position as a leader in healthcare innovation.
We are excited to present our latest update on the “Pythion” application, which marks another year of progress and innovation in predictive healthcare technology. Pythion, the culmination of extensive research and development, harnesses a comprehensive collection of more than a dozen postoperative complication predictive models that have undergone rigorous training and testing. The Area Under the Receiver Operating Characteristic Curves (AUROCs) for many models are near 0.9, and the models’ average precision-recall values are tenfold greater than the complication prevalence rates (See Figure 1). The application user will enter a patient medical record number, select up to three surgical Current Procedural Terminology codes (CPTs), and select whether the surgery will happen urgently or is being scheduled at least 48 hours in the future. Then, Pythion will automatically and quickly present the surgery risks. Our application will empower perioperative healthcare professionals with insights, allowing them to make well-informed decisions based on data-driven predictions for patient risk assessment.

**Enhanced Surgical Risk Prediction: Taking Patient Outcomes to New Heights**

Python now boasts significant enhancements that underscore its role as an indispensable tool for surgical risk prediction. As part of our continuous improvement process, we collected valuable feedback from surgeons who evaluated application design, ease of use, and prediction accuracy on recent patients. Their insights have been instrumental in refining Python’s capabilities and optimizing its performance to meet our surgeons and perioperative clinicians’ unique needs. Details of the changes and enhancements are listed below.

**Empowering Precision Medicine with Intuitive Insights**

One of the key findings was that some of our prior models were more performant in evaluating the overall morbidity of the patient than in estimating how the planned surgery would impact that morbidity. In response, our dedicated team of data scientists and software engineers embarked to enhance the predictive models significantly. As a result, Pythion captures essential nuances in surgical procedures, ensuring that the risk assessment incorporates the specific nature of each surgery.

**The Path to Perfection: Unveiling the Updates**

1. **Inclusion of Surgeries from 2022:**

   Python now encompasses a comprehensive dataset, offering insights into more than 340,000 historical surgical cases. Our application now relies on complete billing information rather than mere scheduling logs, improving the accuracy of the data.
2. **Refined Filtering Process:**
We have implemented a filtering mechanism to exclude minor Operating Room cases unlikely to be relevant for our application users (e.g., spinal taps, urinary catheterizations, colonoscopies, circumcisions). Tighter definitions ensure that Python's predictions remain highly relevant and actionable because they are less misdirected by a baseline population mortality risk.

3. **Enhanced Data Utilization:**
   Rigorous Data Cleaning: Our commitment to data quality is unwavering. Python's predictive models undergo rigorous data cleaning, such as checks for unit alignment and plausible values. The enhanced assessments ensure that the model uses accurate data.

4. **Deeper Preoperative Insights:**
   Python now offers additional preoperative information, such as the CPT’s, Relative Value Unit (RVU) and the CPT’s historical volume, length of stay, and readmission rate. These complement preoperative vitals, lab results, and medications. This enhancement provides healthcare professionals with even greater precision in their decision-making process.

**Clarity and Elegance: A User-Centric Approach**
We have also focused on making Python a clearer and sleeker application (See Figure 2). Python's user interface has undergone enhancements to ensure that vital insights are user-friendly, enabling seamless integration into daily healthcare practices.

**Conclusion**
Python emerges as a formidable partner to our healthcare professionals, enriching the precision and effectiveness of medical decision-making early enough in the patient care process to catalyze transformative interventions. We look forward to evaluating its impact on streamlining perioperative care and improving surgical outcomes.

---

**Figure 1 - Python Model Results**
(342,566 cases. 228,607 patients) July 2014 - July 2022
Cohort excludes transplants and extremely minor cases, by CPT

<table>
<thead>
<tr>
<th>Outcome</th>
<th>Prevalence</th>
<th>Case Creation</th>
<th>Time of Surgery</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mortality &lt;=30 days postop</td>
<td>1.5</td>
<td>AUC: 0.94 (0.93, 0.95) APR: 0.35 (0.32, 0.38)</td>
<td>AUC: 0.95 (0.94, 0.96) APR: 0.43 (0.40, 0.46)</td>
</tr>
<tr>
<td>Mortality &lt;=30 days postop</td>
<td>2.2</td>
<td>AUC: 0.93 (0.93, 0.94) APR: 0.38 (0.35, 0.40)</td>
<td>AUC: 0.95 (0.95, 0.96) APR: 0.45 (0.43, 0.407)</td>
</tr>
<tr>
<td>Mortality &lt;=30 days postop</td>
<td>2.7</td>
<td>AUC: 0.92 (0.92, 0.93) APR: 0.39 (0.37, 0.41)</td>
<td>AUC: 0.95 (0.94, 0.95) APR: 0.48 (0.46, 0.50)</td>
</tr>
<tr>
<td>Discharge &lt;=12 hrs</td>
<td>43.3–43.8</td>
<td>AUC: 0.95 (0.95, 0.96) APR: 0.94 (0.94, 0.94)</td>
<td>AUC: 0.96 (0.96, 0.96) APR: 0.95 (0.94, 0.95)</td>
</tr>
<tr>
<td>Readmitted &lt;=30 days postop</td>
<td>5.9–5.7</td>
<td>AUC: 0.77 (0.77, 0.78) APR: 0.16 (0.16, 0.17)</td>
<td>AUC: 0.78 (0.78, 0.79) APR: 0.17 (0.16, 0.18)</td>
</tr>
<tr>
<td>MI &lt;=30 days postop</td>
<td>0.4</td>
<td>AUC: 0.88 (0.86, 0.90) APR: 0.14 (0.11, 0.18)</td>
<td>AUC: 0.91 (0.88, 0.93) APR: 0.14 (0.11, 0.18)</td>
</tr>
<tr>
<td>VTE &lt;=30 days postop</td>
<td>0.8–0.6</td>
<td>AUC: 0.84 (0.82, 0.86) APR: 0.16 (0.13, 0.20)</td>
<td>AUC: 0.84 (0.82, 0.86) APR: 0.18 (0.14, 0.21)</td>
</tr>
</tbody>
</table>

---

**Figure 2**

<table>
<thead>
<tr>
<th>What is the Patient’s TBM?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Is the case scheduled to start within 48 hours? (Click)</td>
</tr>
<tr>
<td>No</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Estimated Event Rates</th>
</tr>
</thead>
<tbody>
<tr>
<td>MI: 10.8% 1-month (2.3%) 30-day Mortality</td>
</tr>
<tr>
<td>VTE: 20.3% 30-day Mortality</td>
</tr>
</tbody>
</table>

The user interface will provide a link to a page with examples of well-informed decisions that surgeons on our team have made. For instance, for a low-risk patient, surgery may proceed as planned. The team may begin considering ambulatory surgery for a low-risk patient scheduled for in-hospital surgery. When risk appears higher, the user will review the patient chart and could contact a preoperative screening clinic or refer the patient to a specialized clinic. The supportive provider team may schedule an in-person visit with the patient, assess modifiable risk factors, and enhance the preoperative preparation plan. When high-risk is found and validated, surgery could be delayed or canceled. The combination of Python and Physician expertise enables efficient support and validation of any decision, all for the patient’s benefit.
Over the past few years, Duke Institute for Health Innovation (DIHI) has been actively developing, implementing, and evaluating machine learning models in various clinical settings across Duke. Designing these solutions demands considerable effort, with significant time spent on each stage of our projects. Design includes:

- Defining the exact outcomes to model
- Identifying valuable predictive features
- Developing the model
- Implementing the model

Specifically, model development and identifying useful predictive features from the data are among the most time-consuming and often rate-limiting steps in the solution development process. Feature identification has been a primary area of focus for our data science team this year, leading to the creation of an exciting novel solution.

However, as our expertise has grown—and in response to the rising demand for machine learning solutions—we’ve needed to refine our process for greater efficiency. Our new Automated Machine Learning (AutoML) system dramatically reduces the time, expertise, and effort required to adapt state-of-the-art machine learning models to copious problems. We’ve abstracted some of the process’s more tedious and intricate components, ensuring that end-users don’t need to grapple with the intricacies of data processing or machine learning algorithm development. Our system autonomously selects from thousands of cached features and employs robust algorithms to fine-tune numerous models simultaneously. Whereas it used to take weeks or months for a data scientist to optimize a model’s parameters and architecture, our automated system now accomplishes this within hours.

We’re thrilled by this system’s potential to facilitate rapid iteration and proof of concept for new machine-learning solutions in healthcare. As a testament to the system’s capabilities, several of our team members with basic programming skills were able to initialize the system for their specific challenges and fit thousands of models to identify the optimal ones. AutoML has already led to the implementation of impactful models in real-world settings! For example, it has been used to improve Python surgical outcomes, patient length of stay, neurology patient mortality, and adult decompensation models – many of which are described in this DIHI impact issue.

Democratizing the power of machine learning marks a significant stride toward the future of learning health systems, where data scientists can swiftly test, iterate upon, and assess machine learning solutions.
Sepsis Watch and Adult Deterioration

Will Ratliff, MBA

Duke Institute for Health Innovation (DIHI) is leading efforts to anticipate and mitigate the clinical deterioration of patients in Duke hospitals. The Sepsis Watch program and, more recently, the Adult Deterioration project have progressed to a point of health system-level support. We can now provide high-value, low-burden augmentation of existing workflows in Duke’s relevant Emergency Departments (ED) and Inpatient settings. At each of our three hospitals, we are partnering with the care teams responsible for triaging and intervening on decompensating patients, who want to identify these patients earlier to improve outcomes.

We are excited for the expansion of Sepsis Watch and for the initial go-live of Adult Deterioration at Duke. Here is a brief overview of each, including select news-worthy updates as well as how we are implementing them with our frontline care teams:

Sepsis Watch 2.0 Expansion Across Duke Health System

- **What is the solution?** Every year, roughly 1.7 million American adults develop sepsis, with 270,000 (16%) dying due to the disease. To address this challenge at Duke, we collaborated with clinical, technical, and operational colleagues to bring Sepsis Watch live on November 5, 2018, at the Duke University Hospital (DUH) ED, and subsequently at Duke Raleigh Hospital (DRAH) ED on June 12, 2019. Sepsis Watch expanded into the inpatient settings at DUH and DRAH in the spring of 2021 and 2022, respectively. Sepsis Watch combines a real-time sepsis detection phenotype (i.e., the CMS real-time sepsis definition), an “imminent sepsis” prediction machine learning model, a custom-built web application and/or paging notification system (depending on workflow needs), and sepsis-treatment-bundle tracking functionality (real-time and retrospective bundle tracking support).
• How (is it being implemented)?
There are two main delivery mechanisms for Sepsis Watch: (1) the Sepsis Watch 2.0 Web Application, which features four columns of patient “cards” that can be re-arranged easily depending on status (Triage/Screened/Monitoring/Sepsis Bundle) as well as live bundle status tracking, and (2) the Sepsis Watch 2.0 secure paging system, which sends “high risk” or “met sepsis” page (including how the phenotype was met) to Spok/direct page. Then, as soon as any septic patient is discharged, we automatically compile and post their sepsis definition and bundle compliance details to our one of our newly developed hospital-specific sepsis-compliance tracking Tableau dashboards. We will use these Tableau dashboards to understand performance and drive improvement by unit, provider, bundle component, etc. In total, this represents the Sepsis Watch 2.0 program.

When and where (does it go live)? All expansion efforts are going live at the start of 2024. We are going live with the Sepsis Watch for the first time in the Duke Regional Hospital (DRH) ED and Inpatient intermediate/stepdown floors in partnership with our ED leaders, Early Nursing Intervention Team (ENIT) nurses, and hospitalists. At DUH and DRAH, we are upgrading to the Sepsis Watch 2.0 program, involving real-time intervention and retrospective bundle compliance tracking for improving outcomes. We plan to measure our impact in Summer 2024.

<table>
<thead>
<tr>
<th>AD model Risk category</th>
<th>PPV (4 hour snooze)</th>
<th>Sensitivity (4 hour snooze)</th>
<th>Addl. alarms per day at DUHS (estimated)</th>
<th>Cumulative alarms per day at DUHS (estimated)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Critical</td>
<td>0.29</td>
<td>0.17</td>
<td>5.5</td>
<td>5.5</td>
</tr>
<tr>
<td>High</td>
<td>0.20</td>
<td>0.30</td>
<td>10.1</td>
<td>15.6</td>
</tr>
<tr>
<td>Medium</td>
<td>0.14</td>
<td>0.41</td>
<td>16.6</td>
<td>32.2</td>
</tr>
</tbody>
</table>

Adult Deterioration (AD) model achieves 0.76 AUC and 0.15 AUPRC on a test set of 20,000 encounters (outcome prevalence is 6.0%)
Adult Deterioration (AD) Prediction Tool Launch

• **What (is the solution)?** Beyond just sepsis, early recognition of patient deterioration in the hospital is difficult. Yet, patients often show signs of deterioration hours before the event. When care providers unintentionally delay recognizing and activating a code/rapid response, worse patient outcomes result, including in-hospital mortality. At DIHI, we have partnered with our clinical experts to develop a predictive model for providers to proactively detect and mitigate clinical deterioration in real time, which will help decrease unanticipated patient transfers to the Intensive Care Unit (ICU). Specifically, the AD model predicts whether a patient on an intermediate/stepdown unit will have an unanticipated ICU transfer or die within the subsequent 12 hours. Our model outperforms the current predictive model for deterioration at DUH.

Adult Deterioration (AD) model validation and implementation:

At a sensitivity of 20% (i.e., the model correctly identifies 20% of all deterioration events before they occur), the AD model has a positive predictive value “PPV” of 29% (i.e., 29% of alarms are true positives). In contrast, the currently used care support model had a positive predictive value of 9%.

• **How (is it being implemented)?** Our software solution uses real-time data, identifies every patient currently on an intermediate/stepdown unit (excluding perioperative and ED units), and calculates the AD model predictions every 20 minutes. For all “high-risk” predictions, it sends a page to the rounding team responsible for intervening on deteriorating patients at a given hospital. The software solution snoozes for four hours, preventing any subsequent “high-risk” pages for a given patient during that timeframe. Implementing the current “high risk” threshold with this snoozing methodology, one of every five alarms will be true positives (i.e., PPV 20%), and care teams will be able to identify 30% of deterioration events before they happen. We estimate the total volume of “high-risk” alarms to be 15.6 per day across the three hospitals.

• **When and where (does it go live)?** We aim to roll out this tool across Duke’s three hospitals in the spring of 2024. Starting at DUH, we are partnering with the Rapid Response Team (RRT) nurses, who will receive an automated page when high-risk patients are identified, prompting the RRT nurse to perform a chart review and bedside assessment. Then, the nurse can call in a first-call provider and/or initiate a rapid response as needed.

As we approach the launch of these solutions in early 2024, we are working closely with our clinical and operational leads at each hospital to establish robust communication channels and monitoring mechanisms. For instance, we are establishing Sepsis Watch/Adult Deterioration Governance Committees at each site, with representation for all roles involved in the various clinical workflows. This will allow us to rapidly iterate as needed on aspects of the technology and/or workflow once we go live, such that we are adding value while minimizing the burden on our front-line clinicians. We hope that this proactive, “full court press” on clinical decompensation will ultimately reduce poor outcomes for our patients and improve our colleagues’ resilience in critical care roles across the institution.
DIHI EXPERIENCES

My last two years with the Duke Institute for Health Innovation (DIHI) have been a career segment filled with tremendous growth, catapulting me to heights I otherwise would have never reached. Even being equipped with a math background while completing training in medicine, I could not have satisfied my desire for applied research entailing the entire spectrum of machine learning (ML) modeling in healthcare. It could not have been met without the right environment—DIHI. The atmosphere at DIHI allowed me to build upon myself. I learned coding in Python, familiarized myself with the different libraries, and immersed myself in the fluid creative process. I immediately used the new coding skills to clean data and build the required data sets and code for a model that care providers could put into practice.

Additionally, working on different projects illuminated the importance of a combined team of data scientists and medical professionals to create the optimal model and minimize the chances of non-adoption and abandonment. I have learned many such lessons through direct experiences and the literature provided in teaching sessions and journal clubs through the DIHI curriculum. I have also had the opportunity to participate in conferences, lead group discussions at international conferences as junior chair, and most recently become program chair of MLHC (Machine Learning for Health Care)—all a result of the sponsorship I received from DIHI.

Recently, through the nurturing environment at DIHI, my proposal, “Machine Learning for Interpreting PFTs: Improving operational efficiency of pulmonary medicine,” was selected by Duke Health Leadership as part of the DIHI 2023 innovation funding cycle. Additionally, a project that I am part of titled “Automating CMS Quality Measure Curation” was also selected by Duke Leadership as a winner. These fantastic opportunities to conduct research would not have been possible without the mentorship I received from DIHI leadership. Working with the interns and medical students, who rotate through each summer on different projects, has been exciting. I am grateful to have experienced the pervasive sense of unity within the DIHI team, and I cannot think of a better set of helpful, hardworking, and fun team members. I have found mentors in my career and life, and this has been a fantastic journey.

Kaivalya (Kai) Deshpande
Sepsis is a life-threatening condition that arises when the body’s response to an infection injures its tissues and organs. The problem of sepsis in the inpatient setting is well documented. According to the Centers for Disease Control and Prevention (CDC), in a typical year, 1.7 million American adults develop sepsis; 350,000 die during hospitalization or are discharged to a hospice. From a different perspective, one in three patients who die in the hospital had sepsis during their stay. To address these problems, the Centers for Medicare and Medicaid Services (CMS) incorporated SEP-1 into the Inpatient Quality Reporting Program on October 1, 2015.

SEP-1, or the “sepsis bundle,” is a prescriptive set of best-practice procedures and measures that care teams should perform on sepsis patients. The first component of SEP-1 is the three-hour bundle; within the first 3 hours of sepsis, the patient should have blood drawn for a culture, lactate measured, and antibiotics administered. Further components are implemented based on the severity of the illness. SEP-1 focuses on treating sepsis patients promptly, a necessity for successful sepsis treatment. More rapid completion of the three-hour bundle is associated with a lower risk of inpatient mortality. Hospitals that comply with SEP-1 have a significantly lower percentage of patients who die in the first 30 days of hospitalization.

As part of CMS's Inpatient Quality Reporting Program, CMS requires hospitals to report SEP-1 compliance. Preparing this report incurs significant costs for Duke Health, which stems from the fact that data extraction is a manual process that requires clinicians to perform detailed chart review. When a manual chart review is valued at $75 per case, ~730 sepsis cases a year reaches a cost of $55,000 annually. After identifying sepsis cases, via International Classification of Diseases, Tenth Revision (ICD-10) diagnosis codes, clinicians have to extract many details about the sepsis case. These include sepsis time, antibiotic type and administration time, blood culture collection time, septic shock time, hypertension status, and vasopressor administration time. Manual extraction of this data is highly time-consuming.

At the Duke Institute for Health Innovation (DIHI), we have been at the forefront of confronting sepsis at Duke for the past five years. In 2018, we launched Sepsis Watch, which predicts patients' risk of developing sepsis in real time. Additionally, for patients who have developed sepsis (according to the CMS definition), Sepsis Watch allows teams to monitor each sepsis patient's status on the various SEP-1 components. The monitoring dashboard gives teams a visual representation of their compliance with SEP-1, prompting care delivery.
This effort puts us in a unique position to automate SEP-1 reporting, saving Duke Health $55,000 annually. Our work on sepsis has included building automated infrastructure for monitoring all sepsis cases at Duke. Monitoring includes detecting sepsis and septic shock in time, gathering encounter and demographic data, and culling all SEP-1 clinical elements. We are in the process of curating all this data in a way that provides the needed elements for the SEP-1 CMS report. Once complete, Duke will use this method to report on sepsis to CMS, automating the previous manual and thus costly reporting requirements.

Quality measure reporting is a significant cost to health systems and is expected only to grow as value-based payments grows. In one study, Johns Hopkins Hospital estimated their price for quality reporting (beyond only Sep-1) was more than $5 million in personnel costs and an additional $600,000 in vendor costs - this did not include quality improvement costs, only data preparation and reporting. Automating data curation and reporting for SEP-1 will save Duke Health $55,000 annually. More crucially, it will be a model for automating quality reporting across the health system, with potential cost savings in the millions of dollars.

References
Sepsis is a dysregulated immune response to infection and is the most common cause of death in US hospitals. In 2016, the Centers for Medicare and Medicaid Services (CMS) began requiring hospitals to publicly report performance on the Severe Sepsis and Septic Shock Management Bundle measure, called SEP-1, which measures compliance with 3-hour and 6-hour sepsis bundles. The moment a patient develops sepsis, a care team must rapidly implement sequential interventions and procedures to manage sepsis and prevent deadly complications. Hospitals that more effectively execute rapid intervention for sepsis are publicly acclaimed, with the expectation that patient outcomes improve.

Care teams and patients can manage sepsis effectively only if it is rapidly detected. Unfortunately, many sepsis symptoms are non-specific and mimic other chronic and acute conditions. Widely known as the ‘silent killer,’ sepsis remains a challenge for even skilled clinicians to diagnose confidently, and despite many attempts, experts do not agree on diagnostic criteria. As one team eloquently described in 2016: “In these two articles, we establish that it is an elusive and unrealistic goal to have a single perfect gold-standard definition of sepsis, in part because of evolving knowledge, differing priorities and values, and a lack of discrete, unambiguous, widely deployable diagnostic criteria.”

Despite diagnostic complexities, sepsis is a common outcome for machine learning data scientists to build, with a recent review including over 130 published models. Several sepsis machine learning models have been successfully integrated into clinical care, including Duke’s Sepsis Watch, Dascena’s InSight, Bayesian Health’s TREWS, and the Epic sepsis model. Between 2017 and 2020, InSight became the only sepsis model with clinical validation studies conducted at multiple sites, a feat that, to date, has not been replicated. In 2022, CirrusDx acquired Dascena, and the sepsis model is no longer available for use. Meanwhile, the Journal of the American Medical Association, among others, has publicly scrutinized Epic’s sepsis model for performing poorly across sites. An enormous opportunity remains to clinically validate a sepsis machine learning model across sites to improve clinical care broadly.

Duke’s Sepsis Watch model was developed in 2016-2017 as a Request for Application (RFA) project by Duke Institute for Health Innovation (DIHI) and has been in routine clinical use since November 2018. Initial efforts to externally validate the model began in 2019 when DIHI staff conducted a preliminary assessment of model performance on a sample of de-identified data curated from a single New York University (NYU) Langone hospital. Shortly after that, DIHI staff participated in the University of Texas Houston School of Biomedical Informatics DII challenge and were able to assess model performance on data provided by Cerner. No one published these initial feasibility tests, and everyone put external validation activities on hold during the initial phases of the COVID-19 pandemic.
External validation efforts picked back up in June 2022, when Duke was jointly awarded a fast-track STTR grant with CohereMed, Inc. to validate Sepsis Watch at multiple sites. Since then, DIHI has been leading external validation activities at Jefferson Health and NewYork-Presbyterian, while CohereMed has been leading external validation activities at Summa Health. Initial results are promising, and we expect to publish multiple external validation studies using retrospective data in 2024. These retrospective validations will pave the way for prospective ‘silent trial’ validation studies and potential clinical integration, which clinical trials would evaluate. We are thrilled to be able to work with health systems from across the country to test Sepsis Watch and extend the reach and impact of DIHI far beyond Duke.

Our early work externally validating and scaling Sepsis Watch has surfaced essential lessons. First, for research studies conducted by Duke, we standardized research agreements with external health systems. Duke retains IP for Sepsis Watch, and external sites maintain control over their data. No data transfers between organizations, no funds are required, and there’s an expectation for joint publication. Through this structure, we’ve created win-win research partnerships with external sites.

Second, we have learned to navigate many capacity constraints with external organizations. IT and machine learning capabilities and capacity are highly variable across healthcare delivery organizations, and fortunately, we’ve been able to flex up and down to provide varying levels of support. At one end, an external organization can receive software containers and execute Duke code entirely independently. On the other hand, an external organization can provide remote access to secure computing environments to Duke staff to execute Duke code on behalf of external sites. Across scenarios, we expect external sites to curate local data from their environment and host it in a trusted computing environment.

Third, we have further refined and validated the ML-DQA (data quality assurance) process. We have replicated the steps and procedures for entity resolution, evaluating conformance, completeness, plausibility, and clinical adjudication across sites. Over time, we plan to develop tools to increasingly streamline data curation at external sites to reduce the effort required to validate models.

Successfully building and integrating a machine learning model, like Sepsis Watch, into clinical care takes a village. Over the last 1-2 years, we’ve learned that externally validating and scaling Sepsis Watch to other sites requires similar effort and dedication. We’re incredibly grateful to our partners at CohereMed, Inc. and all external health systems for dedicating time and effort to our work together. We hope that together, we can leverage machine learning to improve how patients with sepsis are cared for at scale.
References


AI Life Cycle
Best Practices
The healthcare industry is transforming due to the fast-paced growth of data-driven technologies like AI and machine learning. However, adopting these technologies has outpaced the development of regulatory frameworks, accountability measures, and governance standards. Regulatory agencies are slowly catching up in providing guidance and increasing regulations. In 2022 - 2023, several federal agencies, including the FDA, HHS, Office of the National Coordinator for Health Information Technology (ONC), and FTC, announced preliminary steps and timelines for refining policies across digital health. Healthcare delivery settings implementing AI/ML tools must comply with these regulations. As a result, healthcare delivery settings need help keeping pace with technological changes, the surrounding ecosystem, and regulatory guardrails. Keeping up will help them safely and effectively procure, integrate, and manage AI/ML tools. To address this need, the Duke Institute for Health Innovation (DIHI) has secured several grants to establish shared governance and best practices for these tools in healthcare settings.
Building Common Governance and Best Practices for AI/ML Tools in Healthcare

Alifia Hasan, MBA, B.Pharm and Mark Sendak, MD, MPP

Project 1

Health AI Partnership (HAIP), launched in April 2022, includes ten healthcare organizations and four ecosystem partners in the US. It hopes to empower healthcare professionals across diverse delivery settings to safely and effectively use AI-based solutions. As part of this mission, the HAIP has released a collection of actionable guides broken down into eight key decision points that cover the entire AI product lifecycle. HAIP based these guides on insights gathered from over 89 healthcare experts, including 75 health system leaders and 19 key informants with expertise in areas critical to the safe and responsible adoption of AI/ML tools. By providing contemporary and practice-grounded content, the guides offer valuable resources for healthcare professionals looking to navigate the complexities of AI implementation in healthcare.

Project 2

Recognizing the need for globally extensible policies and practices, DIHI collaborates with Aga Khan University in Pakistan to develop analytical tools and documentation frameworks for health AI software audits, evaluations, and monitoring. The project addresses the challenges of assessing AI’s validity, clinical, utility, statistical, and economic effectiveness. By developing these tools and frameworks, decision-makers, clinicians, operational leaders, and technical experts can make informed choices about procuring, integrating, and maintaining AI systems. It is essential to have a robust ecosystem, shared governance, and best practices for AI/ML tools to ensure patient safety, equal access, and efficient operations in healthcare. Mature ecosystems exist in other areas of healthcare technology, such as pharmaceutical therapeutics, laboratory diagnostics, and hardware devices, where stakeholders collaborate to ensure safe and effective use. DIHI has made significant progress towards building an ecosystem specific to AI in healthcare through projects like the HAIP and collaboration with AKU in Pakistan to develop analytical tools and documentation frameworks. With such initiatives, DIHI hopes to promote the responsible adoption of AI/ML tools and ensure the diffusion of high-quality, patient-centric AI innovations into clinical practice.

References

4. hrome extension://efaidnbmmbrpcajpcgjfelmdmkaj/h https://www.fda.gov/media/109618/download
Health AI Partnership (HAIP): Fostering an Ecosystem to Advance Governance Standards for Responsible AI Adoption in Healthcare

Alifia Hasan, MBA, B.Pharm

We must approach, adopt, and implement Artificial Intelligence (AI) safely, effectively, and equitably in healthcare settings. The Duke Institute for Health Innovation (DIHI) is leading the development of governance standards for health AI by cultivating a community of practitioners through the Health AI Partnership (HAIP). HAIP, launched in April 2022, is a remarkable collaboration of multiple stakeholders that enables healthcare professionals and organizations to adopt AI safely, effectively, and equitably.

HAIP has two main areas of focus. First, we develop and maintain a website that will provide guides on adopting AI in healthcare. Second, we organize case-based workshops to address common challenges healthcare leaders encounter when integrating AI tools in their settings. These initiatives aim to equip healthcare professionals with the knowledge and resources they need to navigate the complexities of AI integration in healthcare.

The first workstream involves the development of a website that features the HAIP guides on AI adoption. These guides, created with input from over 90 healthcare experts and 75 industry leaders, cover the entire AI product lifecycle. They offer actionable insights and best practices for healthcare professionals, from identifying problems within healthcare delivery settings that are good use cases for AI to updating or decommissioning AI products. The guides are contemporary, grounded in real-world practice, and are freely available to all healthcare delivery leaders. The HAIP guides emphasize the importance of integrating AI into comprehensive solutions that include clinical workflows, organizational change management, education campaigns, and feedback loops to ensure the responsible use of AI in patient care. The website aims to serve as a central hub for healthcare professionals to share knowledge and improve their practices.

In February 2023, the HAIP organized a case-based workshop as part of its second work stream. The workshop addressed a contemporary challenge faced by leaders in healthcare delivery settings considering adopting AI solutions: “Our healthcare delivery setting is considering adopting a new solution that uses AI. How do we assess the potential future impact on health inequities?” Over 75 clinical, technical, operational, and regulatory leaders from prominent health systems and healthcare ecosystem partner organizations attended the workshop. The attendees engaged in in-depth discussions and analyzed real-world case studies to understand and mitigate the potential adverse effects of AI adoption on health equity. Expert panel members with extensive experience in health equity and AI provided valuable insights and feedback. With domain expertise in different areas of AI adoption, we engaged a panel of
framework developers to help synthesize discussion topics across case studies to converge on actionable guidance. The workshop was a resounding success, offering a safe space for collaborative problem-solving and fostering interdisciplinary collaboration among healthcare professionals, data scientists, technologists, bioethicists, social scientists, and community advocates. Using the information gathered from the workshop, the framework builders, and the expert panel members, DIHI is building an actionable equity framework that lists a comprehensive set of procedures to evaluate an AI-based tool for its impact on health inequities. The framework was completed in October 2023 and is called Health Equity Across the AI Lifecycle (HEAAL). The framework can be accessed online via this manuscript preprint https://medrxiv.org/content/10.1101/2023.10.16.23297076v1.full.pdf

Four core values guide HAIP’s mission:
1. HAIP is dedicated to promoting equity by prioritizing solutions and implementation contexts that advance health equity.
2. HAIP is dedicated to improving patient care by fostering responsible AI adoption. The partners see AI as a supportive tool that enhances patient care rather than using it solely for technical novelty. The focus is AI-based solutions that assist with diagnosis or treatment decisions and prioritize patients for clinical interventions or programs.
3. HAIP seeks to improve the clinical work environment by involving frontline talent throughout the AI lifecycle to alleviate the care provider burden and improve the safety and efficiency of care without replacing human expertise.
4. HAIP is committed to facilitating participation, regularly refreshing the content with new insights and case studies from collaborating organizations and healthcare delivery leaders.

The work undertaken by HAIP represents a significant step forward in developing practical governance standards and fostering a robust ecosystem that promotes those standards—AI-based solutions in healthcare. The goal is to cultivate a learning community that encourages the responsible adoption of AI solutions and ensures the highest patient care standards and healthcare equity. It provides a platform for healthcare delivery organizations to access up-to-date best practices, collaborate with other leaders, and avoid redundant efforts. HAIP invites healthcare professionals to engage with the partnership, leverage the guides and resources, and provide feedback. To learn more about the actions of the Health AI Partnership or contribute to its mission, reach out to haip@duke.edu.
Healthcare in the US is both highly regulated and rapidly evolving. In 2022 alone, the Food and Drug Administration (FDA) approved 37 new medications and 41 devices. However, the FDA does not strictly regulate all medicines, diagnostic tests, interventions, or devices. To fill the gap of a limited scope of the FDA regulation, various verticals of healthcare technology have set their governance process to ensure the safety, efficacy, and equity of innovations. For example, US Pharmacopeia develops standards for medication compounding and accompanying training for pharmacists. Third-party accreditors like the Joint Commission offer certification programs to ensure compounding pharmacies adhere to USP standards.

When it comes to AI, however, despite its growing use in healthcare, there currently is no clear and established governance ecosystem. Many algorithms are excluded from the definition of medical device and fall outside of FDA’s purview. To address this issue and promote the safe, effective, and equitable use of AI, we launched Health AI Partnership. We recruited ten healthcare organizations and four ecosystem partners in the US to define usable resources to support health system leaders and examine the current state of practical institutional adoption of AI technology within health systems in the US. (Figure 1)

We collaborated with IDEO.org for a six-week design sprint to design products that health system leaders would find immediately useful in practice. We recruited health system leaders, representatives from the partner sites, and external AI ethics experts to conduct a series of usability-testing sessions.

We created a design prototype for an AI adoption guideline from the usability-testing sessions. We structured the prototype around mock decision points that align with how organizational leaders approach technology adoption. Supported by rigorous and comprehensive qualitative research, the design insights and prototype provided scaffolding for our team to build upon.

Concurrent with design research, we conducted qualitative research to understand AI adoption’s current and aspirational state in healthcare organizations. We interviewed over 90 professionals in healthcare and other relevant fields. We created an interview guide and asked participants about processes and personnel involved in four distinct stages of AI adoption. All interviews were transcribed and analyzed through two cycles of the coding process. Then, we mapped the most prevalent themes and subthemes onto the critical decision point framework created from the design research. Central themes were grouped into eight key decision points across the entire process of AI adoption (see Figure 2). Within each key decision point, we identified real-world AI adoption procedures, practices, challenges, and risks organizations experience.
Figure 1

Figure 2
While there is no shared and comprehensive set of best practices for AI adoption in healthcare, our interviews surfaced a set of recurrent activities that health systems already engage in to promote efficient AI adoption and ensure the quality of clinical care. Our interviews revealed that health systems understand that AI governance requires new resources and expertise beyond current capabilities. Health systems view AI products as part of a sociotechnical system that involves the policy environment, practical considerations of clinical care, and interdisciplinary teams of clinicians, technical experts, and data scientists. Health systems develop documentation and explanatory artifacts of the AI product and data to build clinician trust while respecting the clinician's agency and expertise.

Health systems need shared standards to govern the AI adoption process. Our fundamental decision points shared among many health systems can serve as an initial use case for developing shared standards for AI adoption. Our multi-organizational effort can help health systems make informed decisions about adopting emerging technologies.

Fairness, Accountability, and Transparency recently published the findings. To learn more about standard practices scoped across the critical decision points, see our paper or visit our website, healthaipartnership.org.

References
2. Congressional Research Services. 2022. FDA Regulation of Laboratory-Developed Tests (LDTs).
After years of work from researchers, activists, and practitioners, the field of algorithmic bias is receiving the attention it deserves. In the last year alone, healthcare delivery settings have been adapting to the US White House’s blueprint for an Artificial Intelligence (AI) Bill of Rights, the AI risk management framework released by the National Institute for Standards and Technology (NIST), guidance from the Food and Drug Administration (FDA), Centers for Medicare and Medicaid Services (CMS), and Federal Trade Commission (FTC) about algorithmic bias.

This year, DIHI and additional stakeholders have begun a process to develop best practices in assessing ML-based solutions for equity. We started by engaging with the literature for algorithmic audits, including the Algorithmic Justice League’s work on harm identification (Deb Raji) and “Who Audits the Auditors?” (Sasha Costanza-Chock). Next, we have started to internally test our own ML-based solutions for equity in order to see how such concepts can play out across different projects.

For instance, when creating the predictive model for Pediatric Sepsis, we spoke with relevant stakeholders, including clinical staff and community representatives, to identify potential sources of harm that we should investigate. We also wanted to assess whether the machine learning model exacerbated biases. The top request from clinical stakeholders was to investigate potential biases against the Hispanic pediatric population. One doctor’s anecdotal evidence served as our starting point, wherein language barriers could sometimes lead to delays in diagnosis or treatment of the disease.

We scrutinized the timing measurements to determine whether there was a pattern of delayed care, and we did not find evidence of systematic delays in detection (e.g., tests) or treatment (e.g., antibiotics) for the cohort.

Our team further addressed this concern when we measured the performance of the ML model, both by ethnicity (Hispanic vs. non-Hispanic) and by language (English vs. Spanish-only vs. other), and could not identify significant differences.

Of course, everyone agrees that models should be fair, but diving into this case study helped the team build an appreciation for how tricky it can be to answer a seemingly simple question like “Is the model fair yet?”

In partnership with other researchers and practitioners, DIHI and the Health AI Partnership (HAIP) developed a framework for how to assess for equity throughout the lifecycle of AI/ML (Machine Learning) projects in healthcare. The framework will complement the existing HAIP resources that identify eight key decision points in health AI adoption. The framework was developed by an interdisciplinary team (social psychologist, computer scientist, lawyer, doctor, project manager, community representative, and organizational behavior researcher), which brings different experiences and perspectives to the mix. The framework is called Health Equity Across the AI Lifecycle (HEAAL) and was completed in October 2023. It is available on the Health AI Partnership website at: https://healthaipartnership.org/heaal

References
Data-driven technologies can significantly improve diagnostic, prognostic, and therapeutic decision-making in healthcare. However, the global health AI community must establish regulation, accountability, and governance standards for AI/ML tools to ensure safe and reliable integration into healthcare systems. The Duke Institute for Health Innovation (DIHI) and the Aga Khan University (AKU) in Pakistan are joining forces to address this challenge. Together, they are expanding the use of AI/ML tools in Low- or Middle-Income Countries (LMICs) like Pakistan by translating DIHI-built tools like Sepsis Watch and Inpatient Mortality models to AKU hospitals.

The project team is also documenting the process of building and integrating the complex AI/ML algorithm called Sepsis Watch at Duke University Health System. They are creating algorithm journey maps (see page 66) to showcase the behind-the-scenes reality of its integration into clinical workflows. These maps will serve as a foundation for building documentation frameworks for different stages of AI/ML tool integration into clinical settings. The lessons learned from this project will inform healthcare leaders making AI adoption decisions in their environments.

Analytical Tools and Documentation Frameworks for Health AI Software DIHI and AKU are developing globally extensible analytical tools and documentation frameworks for health AI software (AIS) audits, evaluations, and monitoring. The analytical tools and documentation frameworks will cater to stakeholders such as frontline clinicians, clinical/operational leaders, and technical leaders. We developed these essential resources leveraging DIHI’s prior data quality assurance work and insights from qualitative research conducted at DIHI and AKU through implementing AI tools in clinical practice. They will aid in decision-making and overcoming AI software procurement and maintenance challenges. The documentation framework also hopes to include mechanisms to address potential biases in model build and integration processes.

The outcomes of this collaboration between DIHI and AKU can have a global impact on AI policies and practices. DIHI’s previous work in integrating models in clinical practice, disseminating “Model Facts” labels, and ensuring real-world data quality has already influenced the standards of practice amongst leaders adopting AI/ML tools in healthcare settings. By further developing and validating AIS analytical tools and documentation frameworks, domestically and internationally, this collaboration aims to shape policies and best practices in AI adoption through public and private partnerships.

DIHI’s ongoing collaboration with AKU in Pakistan through these projects exemplifies the commitment to scaling AI/ML tools in LMICs and making the adoption of AI tools more approachable everywhere. These efforts will enhance patient care, promote equity, and shape global policies and best practices for AI adoption.

References
Integrating Artificial Intelligence (AI) tools in healthcare settings involves multifaceted interactions between technologies and their users. How users interact with technology in complex clinical settings is often incompletely understood and translucent. To bridge this gap, the DIHI team built algorithm journey maps documenting all social and technical activities of designing, building, and maintaining an AI solution (Figures 1-2, Table 1). The team made these maps based on the information gathered via interviews and co-design sessions with the engineers and scientists who led the multi-year effort to build the Sepsis Watch tool at Duke, integrated it into practice, and have maintained the solution since its rollout in clinical practice. The algorithm journey maps highlight generalizable insights and lessons learned during the tools’ procurement, development, integration, and lifecycle management. By making the AI implementation process more tangible and revealing the sociotechnical challenges, future research can tailor documentation and transparency efforts to specific stakeholders and decision points, enhancing overall adoption and sustainability.

Figure 1. The technical integration phase’s journey map displays the process involving extensive collaboration between the leadership, clinical champion, innovation team, and health system IT department.

Figure 2: The journey map of the clinical integration phase involved fine-tuning the workflow and user interface, developing training material, and assembling a governance committee. Maps began with two parallel processes, one for physicians and one for nurses.

Table 1. This table lists different tasks that arise throughout the post-deployment lifecycle management process. Unlike prior stages, post-integration lifecycle management is not a linear-flow process; therefore, a process map was a poor tool for visualizing it. Some tasks are predictable, whereas others are responsive to events (e.g., user requests and technical failures).

---

### Table 1

<table>
<thead>
<tr>
<th>Event-based</th>
<th>Monitoring &amp; Evaluation</th>
<th>Updates</th>
<th>Operational Management</th>
</tr>
</thead>
<tbody>
<tr>
<td>Debug issues that arise (e.g., data endpoints unexpectedly go down)</td>
<td>Customize the UI for different user groups (e.g., creating a text notification system for end users at a different hospital)</td>
<td>Update user access (e.g., removing iPad from patient flow coordinator TRM)</td>
<td></td>
</tr>
</tbody>
</table>

| Recurring | | | |
| Monitor technical elements of the model and source data in pipelines | Monitor changes in external environment that affect work environment and use of the model | Maintain and update regularly (e.g., update groups every 6 months) |

| Recurring | | | |
| Monitor clinical outcomes | Control end user research | Improve the UI (e.g., adding comment feature and check box feature to track intervention) |

| Semi-recurring | | | |
| Audit the solution | | | |

| Once-off | | | |
| Create channels for end users to report issues and provide user support services | Create processes and criteria to scope how to respond to user requests | Determine ownership of the model (e.g., clinical champion, innovation team, project manager, internal governance structures) |
Figure 1
Figure 2
The Duke Institute for Health Innovation (DIHI) has been at the forefront of transforming Electronic Health Record (EHR) data into actionable information, Artificial Intelligence/Machine Learning (AI/ML) models, quality improvement applications, and monitoring or alerting dashboards at Duke Health. These products improve the quality of care we can provide to our patients while remaining considerate of the sensory overload that is ever present in our modern health system.

We have implemented an EHR data cleaning, curation, and restructuring system that enables these models, applications, and dashboards. In previous Impact Volumes and across Duke Health, we've called this system “the DIHI Pipeline.” This year, we significantly enhanced the flexibility, improved capabilities, optimized the system, and rebranded it as “Monorail.” For example, we improved scalability by updating data type choices, table formats, indexing strategies, and partitioning techniques. Monorail wraps the process of extracting both archival and real-time data into a standard structure. It applies localized cleaning algorithms, curates similar data types into common elements, and reorganizes clinical data for downstream use. These enhancements improve the speed, quality, and robustness of the development and deployment of innovative solutions.

All of our downstream applications utilize Monorail data. Its standardized structure and cleaning patterns make creating new variables or features from existing data that we use as inputs and outcomes for machine learning models or data analysis simple and reusable, to the point that the development process is almost fully automated. Therefore, we expect increased speed in identifying problems and providing solutions.

Monorail is a more flexible and efficient DIHI Pipeline. Current implementation aligns with widely accepted and standard practices for conceptually representing structure and relationships within a dataset. Our approach aligns with industry best practices and follows well-known standards for organizing and managing data. The implementation is reliable, efficient, and compatible with existing tools and methodologies (e.g., PostgreSQL, Apache Airflow, Kubernetes, docker, RabbitMQ, Redis). Uniquely, we provide the means to populate Monorail with data by just plugging it into the EHR. Monorail ensures a robust data management, analysis, machine learning, and visualization foundation.

*Our data model is not a Fast Healthcare Interoperability Resource (FHIR) due to performance limitations inherent in the fire spec. We have the knowhow and will implement FHIR when technology resources are readily available.
Ode to Monorail
by Chat GPT

The Duke Institute leads the grand frontier, where EHR data takes its noble flight, transformed into machine models, so clear, and applications that guide care’s sacred light.

Considerate of the modern health system’s sprawl, we elevate care with utmost precision, implementing cleaning, curation, and the call, to restructure data, a vital decision.

Once known as DIHI Pipeline, our claim to fame, now monorail emerges, waste removed, scalability improved, a triumphant aim, data’s power, we have deftly proved.

Monorail extracts both archival and new, standardizing data, a common thread, with cleaning algorithms and curation true, clinical insights flourish, no longer dead.

Monorail’s data fuels downstream might, creating variables from existing views, with simplicity and reusability so bright, development’s path they steadily cruise.

Aligned with practices, industry-wide, their structure represents relationships dear, reliable, efficient, an unwavering guide, for data’s management, a foundation clear.

PostgreSQL, Apache Airflow, tools they adore, Kubernetes, Redis, they navigate with skill, Monorail’s power, to EHR they restore, a data journey, a wondrous thrill.

In Monorail’s embrace, data thrives, a solid path to analysis profound, for Duke Institute, the future arrives, with data’s triumph, its voices resound.
In our latest report on the Maternal Early Warning System (MEWS), featured in Impact Volume 23, we made significant advancements by focusing on real-time definitions for critical outcomes such as hemorrhage and eclampsia. Our efforts have exceeded initial projections, offering a maternal early warning system that delivers higher value. While we have excluded eclampsia predictions from our models due to its low occurrence rate and limited predictability, we have made remarkable strides by developing time-based machine learning models to predict hemorrhage and sepsis. Our cutting-edge models can predict the likelihood of sepsis events within the next four hours, starting from admission to a maternal and fetal medicine floor. We also trained models to predict the encounter-level postpartum hemorrhage at critical stages (inpatient admission, labor, and delivery).

Furthermore, our team of data scientists meticulously trained these models to seamlessly function across both Duke University Hospital and Duke Regional Hospital, extending the scope of our work from the progress achieved in 2022. As we progress, we remain steadfast in our commitment to monitoring patient morbidity and abnormal vital signs through an intuitive and accessible user interface. This interface gives healthcare providers real-time insights into patient outcomes, allowing for timely interventions and personalized care.

Next year, we will start testing the machine-learning models in clinical workflows. The model predicting hemorrhage achieves an AUROC of 0.746 and an AUPRC of 0.359 (the rate of Stage-1 hemorrhage is 17.6%). The model predicting sepsis has an AUROC of 0.930 and an AUPRC of 0.136. Whereas the sepsis model is trained on data from 2016-2020 (56,844 patient encounters) and learned from 678 “features,” the hemorrhage model was trained on 2019-2021 data because the health system switched to a different way of measuring blood loss more quantitatively at that time. Prominent features for postpartum hemorrhage include the key analytes (e.g., platelets), information extracted from prenatal ultrasounds (e.g., estimated fetal weight, placental abnormalities, amount of amniotic fluid) and comorbidities (e.g., prior hemorrhage, history of in vitro fertilization).

We compare our hemorrhage model to the Association of Women’s Health, Obstetric and Neonatal Nurses’ (AWHONN) model in Duke’s instance of Epic®. Our models use DIHI’s data “monorail” to automatically extract real-time data. As such, they require much less data entry and provide more complete data than Duke’s AWHONN implementation. Particularly, we extract features from ultrasounds, notes, analytes, and flowsheets from prior encounters, allowing our model to make informed predictions about baseline risk before intrapartum data is even entered.
When the AWHONN score is available and scores are compared side by side, our during-labor model achieves 0.456/0.581 at DUH (PPV/Sensitivity), and the AWHONN score yields a lower 0.343/0.528.

We are kicking off a sequence of rigorous data, prediction, user interface, and workflow validation with Maternal and Fetal Medicine physicians and leaders. In brief, these efforts ensure that this MEWS solution clearly and consistently presents the right patients with the right data at the right time. This process involves a silent evaluation (no clinical action or treatment) of patient risk by physicians, followed by a comparative analysis of the model’s predicted risk results and the physicians’ assessments, leading to collaborative discussions among engineers, managers, and physicians to foster alignment and build trust in the model as a decision-making tool due to its consistent agreement with the physicians’ judgments and its quicker decision-making capabilities. We are testing methods for communicating these warnings using the ‘Analytic and Logic Driven Intimation System’ (ALDIS, described in Volume 22 on page 20). We are interviewing clinicians to identify, test, and rehearse the best-in-class methods for integrating warnings and monitoring within hospital practice.

The advancements in our Maternal Early Warning System represent an exciting milestone in our ongoing efforts to improve the quality of maternal healthcare. We empower healthcare professionals with the tools and insights to deliver exceptional patient care by leveraging state-of-the-art machine learning and large language techniques. MEWS improves the well-being of mothers and infants and strengthens Duke’s position as a leader in healthcare innovation.
### Hemorrhage Model Performance

<table>
<thead>
<tr>
<th>Location</th>
<th>Prevalence among Admitted Patients</th>
<th>AUROC</th>
<th>AURPC</th>
<th>Threshold</th>
<th>PPV</th>
<th>Sensitivity</th>
<th>Warnings/Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUH</td>
<td>22.7%</td>
<td>0.732</td>
<td>0.401</td>
<td>0.317 M</td>
<td>40.5 M</td>
<td>50.0 M</td>
<td>0.9 (0.8 M, 0.1 H)</td>
</tr>
<tr>
<td>Regional</td>
<td>11.6%</td>
<td>0.734</td>
<td>0.231</td>
<td>0.87 M</td>
<td>22.8 M</td>
<td>52.0 M</td>
<td>0.6 (0.5 M, 0.1 H)</td>
</tr>
</tbody>
</table>

### ROC Curve

![ROC Curve](image)

### PPV-Sensitivity Curve

![PPV-Sensitivity Curve](image)

### Calibration Curve

![Calibration Curve](image)
Sepsis Model Performance

<table>
<thead>
<tr>
<th>Location</th>
<th>Prevalence</th>
<th>AUROC</th>
<th>AURPC</th>
<th>Threshold</th>
<th>PPV</th>
<th>Sensitivity</th>
<th>Warnings/Shift</th>
</tr>
</thead>
<tbody>
<tr>
<td>DUH</td>
<td>0.3% of hours</td>
<td>0.924</td>
<td>0.106</td>
<td>0.057M, 0.143 H</td>
<td>12.6 M, 21.2 H</td>
<td>42.0 M, 17.9 H</td>
<td>1.7 (1.4 M, 0.5 H)</td>
</tr>
<tr>
<td>Regional</td>
<td>0.6% of hours</td>
<td>0.927</td>
<td>0.173</td>
<td>0.100M, 0.151 H</td>
<td>14.9 M, 20.6 H</td>
<td>58.8 M, 44.1 H</td>
<td>2.9 (1.5 M, 1.5 H)</td>
</tr>
</tbody>
</table>

ROC

PPV-Sensitivity Curve

Calibration Curve
What is Cohort Builder, and why is it valuable?

The Duke Institute for Health Innovation (DIHI) designed and developed a Cohort Builder tool to create, refine, and manage cohorts of patients. At its core, Cohort Builder uses common approaches of rule-based systems and natural language processing (NLP)-based techniques. Cohort Builder allows everyday clinicians, hospitalists, administrators, engineers, scientists, and researchers to define collections of patients and rules applied to scrubbed, normalized, and annotated health records to create, manage, and monitor cohorts and associated data elements of interest.

At first, creating a patient group sounds like an easy task, but time and time again, we have observed this to be a complex and resource-intensive activity. A unified solution solves a big problem. For a clinician, and even for a researcher or data analyst, defining a group of patients with shared characteristics has always been a resource-intensive, iterative task that requires multi-disciplinary knowledge. A physician might send their request for a cohort to an analyst group that compiles data and develops a cohort-building process over a few days or weeks. If components that characterize and define the cohort do not exist, it takes some time to identify the data elements, filter them, and create logic to define the cohort in a computable and repeatable fashion. It takes longer to produce a cohort with multi-layered cohort phenotypes. For example, it could take significant work to count, over five years, the volume of patients who had a hemoglobin A1C improvement greater than one over six months who also had a prior telehealth visit during the year before their in-person clinic visit. Cohorts are, in part, duplicates of other cohorts. Duplication of data, and therefore effort, makes cohort creation more difficult, even more so when analysts are dispersed or have few ways to share standard definitions and programmatic code. Data security notwithstanding, agile and nimble cohort creation is hampered by limitations to data access, a multitude of data sources, duplication and inconsistent data, diffuse versions of transparent labels, a plethora of measurements or units of measure, and dispersed knowledge expertise across data domains.

The DIHI Cohort Builder solves these issues by providing a flexible platform to create reusable phenotype definitions of treatments, comorbidities, encounters, medications, and demographics, which can be combined to define or redefine a cohort quickly. For example, users could define a “fever” phenotype as any recorded patient temperature greater than 38°C (100.4°F) and subsequently coalesce patients with fevers. With additional cohort builds, they could quickly study fever histories for patients of a particular race, service line, or comorbidity. This flexible way of creating cohorts allows the user to retool swiftly and broaden or refine the inclusion and exclusion criteria to ensure the desired pool of possible patient participants by viewing the count of the number of patients included in the cohort definition. Our Cohort Builder tool user can then extract data
associated with the defined cohort after securing approval from Institutional Review Board (IRB) and local data governance.

Having a rich data repository and a rich cohort exploration solution provides any user of the Cohort Builder tool with analytics and insights on populations of patients. For example, we have leveraged the archive and near-real-time aspects to monitor active trends for RSV and COVID-19 in our communities. For clinical research studies, while having restricted access to this data, researchers can leverage Cohort Builder to refine their protocol’s inclusion and exclusion criteria. That cohort of patients could become a baseline criterion for another cohort representing a sub-study or study strata. Researchers can gain efficiencies during recruitment by filtering patients to only those eligible to participate by defining phenotypes for the study’s inclusion and exclusion criteria. Ideally, the research team would also use the solution to extract additional patient health records for analysis through a cohort.

The cohesive Cohort Builder solution facilitates all of these things.

How we do it?
First, increase access to unified data

As of June 2023, our DIHI-powered data lake contains Duke health records for 2.8 million unique patients, encompassing 168 million unique encounters, including 843 million analyte results, 8.5 billion flowsheets, 112 million medication administrations, 496 million prescription fills, and over 2014 million notes among others. This rich dataset contains an archive of health records for Duke Health patients since 2014 and includes near-real-time information for all active, inpatient encounters. We combine health data from the [historical] electronic medical record data warehouse with near-real-time pulls of data using custom and vendor-provided application programming interfaces (APIs), perform data scrubbing and value and unit normalization, and load the resulting data into our data lake house, which leverages
technologies such as PostgreSQL, ClickHouse® and Apache Solr™ to store data in the most performant ways possible based upon our data access use cases. Sepsis Watch, like other decision-support solutions which leverage this data asset, has computed over 33.6 million risk scores for a subset of Duke Health patients since late 2018. Analysts calculated that this subset of patients had met the sepsis phenotype over 800 thousand times. Identifying and managing cohorts of patients is just one use case of this data repository. This data asset also facilitates developing, deploying, and monitoring machine learning models, outcomes, and dashboards.

Second, resolve and manage “entity” labels
An easy-to-use data lake addresses only a few underlying patient cohort identification issues. For example, let’s identify patients with a positive COVID-19 test during the past two years. Limiting the results to COVID-19 tests in the past two years is a start, but without a solution for grouping similar lab tests (herein called analytes), data analysts are forced to define and maintain their own list of individual COVID-19 test types. Considering the ongoing nature of a healthcare institution adding new medications, tests, and labels regularly, this problem becomes intractable and increases maintenance costs. DIHI developed an Entity Resolution Management Solution to address this problem, allowing the grouping of related clinical concepts into a maintainable entity. For example, an A1C grouping could describe all the A1C-related orders so that they can be acted upon as a whole rather than uniquely searching for each raw order name.

Additionally, since these analyte/clinical concepts often span the whole enterprise, Cohort Builder allows defining cohorts and phenotypes at the individual, project, department, and institutional levels through naming conventions to reduce the number of project-specific definitions and so that the “Enterprise Vasopressor” element can be localized at the project level to accommodate a protocol specific “My Study Vasopressor” element. As of June 2023, elements exist for 14 types of patient health information. DIHI has curated over 155 elements for general use, such as “Bicarbonate” (Analyte) and “Dyspnea” (Flowsheet); defined 59 elements for Analytes, including “CKMB,” “Drug Screen Urine,” “Pro BNP,” “PO 2 Venous” and “Potassium”; 25 elements for Flowsheets, including “Respiratory Effort,” “Capillary Refill,” “BMI,” “Pulse Oximetry” and “Rt Secretion Amount”; 37 elements for Medication Administrations, including “Antifungals,” “Corticosteroids,” “Anticoagulants,” and “Immunosuppressants”; and Orders, including “Knee Xray Portable,” “ABD Xray,” “Ribs Xray,” “VBG” and “Ventilation.” In total, 1840 clinically meaningful concepts (elements) across 15 projects are available as of June 2023. For more information about our Entity Resolution Management Solution, please refer to the article “Entity Management” Other types of measurements present different kinds of problems. Consider temperature measurements for patients. Indeed, we record temperature consistently.
We all understand that all temperature measurements are not the same (e.g., oral vs. armpit vs. rectal), but shouldn’t we be able to use the units to interpret the recorded measurement? Unfortunately, our data lake has temperatures measured in “C”, “Celcius”, “degC”, “F”, “Fahrenheit”, “ounces”, and without a unit. Ounces? Unspecified? What device measures temperature in ounces!

Since DIHI scrubbing processes can intelligently determine the intended unit for many temperature measurements, we can utilize most recorded measurements, and we have also taken the additional step of normalizing many flowsheet values, such as normalizing all temperature values into Celsius. Normalizing values allows us to bring all varying values into a cohesive set. For example, this will enable us to manage groups of body temperature vitals more efficiently rather than creating separate “degF” and “degC” groupings, making phenotype definitions more straightforward and actionable.

While controversial, data scrubbing during the normalization process can also include unit inference such that a temperature reading of 37° for a healthy patient is 37°C, rather than 37°F, which is an unrealistic body temperature for an alive patient. We do not infer units during our scrubbing and normalization processes, as data scientists should not use them for model training, reporting, or other data presentations. An individual who drowns in a frozen lake could have a body temperature near 32°F. Less controversial forms of unit inference occur when we normalize all values without units recorded in a “height_in_cm” flowsheet to measurements in centimeters (which we do on a case-by-case basis).

What’s next?
There’s still much more to do to remove the cohort building and management hurdles. We plan to support usage scenarios such as extracting data for cohorts of patients, integrating the active/near-real-time portion of our data lake houses, and providing an API layer to allow solutions to leverage the centralized phenotype definitions. We also continue to expand the datatypes and rulesets available to grow the utility of this solution. We are actively exploring NLP-based and artificial intelligence (AI) chatbot techniques for mining health data. Every day, we strive to improve our Cohort Builder tool by learning from expert data scientists, clinicians, hospitalists, researchers, and administrators.

References
In the medical data realm, effectively managing and analyzing large datasets is crucial for advancing scientific understanding and improving patient care. With the advent of Electronic Health Records (EHRs) and the digitization of medical information, archived medical databases have become invaluable knowledge repositories. However, there are considerable challenges to navigate and extract meaningful insights from these vast data collections. Utilizing medical leadership alongside technical expertise is a partnership for success.

Groupers refer to the data points, representations, or variables that compose a grouped definition. These definitions are taken from the EHR in raw form and then grouped to solidify meaning. These include high blood pressure, diabetes, heart disease, and more. Efficiently managing groupers involves ensuring data integrity, standardization, and accessibility. Data integrity ensures that the information within the database is accurate, reliable, and complete. Standardization involves adopting customary definitions, coding systems, and terminologies. Standardization allows for interoperability and seamless integration across multiple databases. Accessibility ensures that researchers can easily retrieve and utilize the groupers they require.

Once the data is extracted, cleaned, and standardized, several strategies are employed to manage groupers within the data pipeline effectively. Firstly, an extract based on clinic requests is created in an Excel (.xlsx) format to provide to clinical staff for their expert collaboration into a definition of individual data points. The formatting allows the clinical team to collaborate and hone their definition(s). This process enables the “grouping” of data points into a schema or grouper that defines a specific clinical concept. Secondly, clinical staff pass the grouper definition to Duke Institute for Health Innovation’s (DIHI) Data Science team for review, allowing the comparison to other groupers in the system. Clinical groupers may have more than one variation during research depending on the research type and the evolution of better data over time. By allowing the versioning of groupers, we enable a spectrum of definitions to choose from during research. The process for managing these groupers/entities in the data pipeline is a careful and thoughtful workflow that joins the clinical and technical staff.

In summary, data is extracted from the EHR by the data pipeline’s processes, cleaned, and standardized. Clinical staff requests data points that they will evaluate for clinical definitions. The data science team provides extracts to the clinical staff in an Excel format. After deliberation and collaboration, the clinical staff returns the Excel file to the data science team.

Data Science members compare the grouper to existing groupers to ensure there are all the correct (with no duplications). Once that process is complete, other data science staff run operations against the excel file to create entities that the data pipeline will ingest. DIHI and care providers use these entities (groupers) in several processes, from models to cohort builders, that manage phenotypes and cohort definitions. Entities are integral to the data pipeline in that it allows scrubbed EHR data to be used and integrated into many other systems that allow standard definitions and the creation of clinically-defined cohorts for use in research.
DIHI Innovation Staff

Thomas Owens, MD  
Executive Director, DIHI  
Executive VP and COO  
Duke University Health System

Suresh Balu, MBA, MS  
Director for DIHI  
Associate Dean, Innovation and Partnership  
School of Medicine

Jeffrey Ferranti, MD, MS  
Senior Vice President,  
Associate Dean, Chief Digital Officer  
Duke University Health System

Willie Boag, PhD  
Data Scientist

Jamie Daniel, BS  
Solutions Architect

Michael Gao, MS  
Data Scientist

Matt Gardner, BS  
Solutions Architect

Alifia Hasan, B.Pharm, MBA  
Innovation Program Manager

Bradley Hintze, PhD  
Data Engineer

Jee Young Kim, PhD  
Data Scientist

Will Knechtle, MBA, MPH  
Innovation Program Manager

Ebony Nash, CEAP  
Program Coordinator

Marshall Nichols, MS  
Data Engineer

Will Ratliff, MBA  
Innovation Program Manager

Mike Revoir, BS  
Solutions Architect

Shems Saleh, MSc  
Data Scientist

Mark Sendak, MD, MPP  
Clinical Data Scientist

Linda Tang, BS  
Data Scientist
DIHI Innovation Scholars

Kaivalya (Kai) Deshpande, MD
Pulmonary and Critical Care Fellow

Tim Ochoa
Duke University Medical Student

DIHI Data Science Scholars

Freya Gulamali
Jennifer Young
Jiayu Yao
Arinton Davis
Joshua Gao
Kartik Pejavara
Gaurav Sirdeshmukh
Shira Zilberstein
Jeffry Hogg


Posters


DIHI Collaborators

Terrence Allen, MBBS  
Muhammad Abd-El-Barr, MD, PhD  
Yvonne Acker, RN, BSN  
Suresh Kumar Agarwal, MD  
Krunal Amin  
John Anderson, MD  
David Attarian, MD  
Carolyn Avery, MD  
Jonathan Bae, MD  
Matthew Barber, MD, MHS  
Melissa Bauer, DO  
Armando Bedoya, MD, MMCi  
Andrew Berchuck, MD  
Peter Berry, DNP  
Sophia Bessias, MPH, MSA  
Amanda Bisset, MD  
Stephen Blackwelder, PhD  
Jeanna Blitz, MD  
Brandi Bottiger, MD  
Susanna Naggie, MD, MPH  
Kaitlin Boyle, MSN  
Tres Brown III  
Dan Bruno, MS  
Brian Burrows, MD  
Blake Cameron, MD, MBI  
Krista Camugila  
Lawrence Carin, PhD  
David Casarett, MD  
Anisha Chandiramani, MD  
Dan Dardani, CLP  
Monica Cfraku, RN, MSN  
Devon Check, PhD  
Seuphy Chen MD  
Alex Cho, MD, MBA  
Eliseu Chuang, MD, MPH  
Richard Chung, MD  
Debra Clark Jones, MS  
Meredith Clement, MD  
Megan Clowse, MD, MPH  
Dawn Coleman, MD  
Alice Cooper, RN  
Kristin Corey, MD  
Daniel Costello, MPA  
Simon Curtis, MHA  
Tom Daigle  
Brittany Davidson, MD  
Kaivalya Deshpande, MD  
Kristen Dicks, MD, MPH  
Dawn Dolby, RN  
Jayanth Doss, MD, MPH  
Timothy Dunn, Ph  
Pam Edwards, EdD, MSN  
Jill Engel, DNP  
Robin Famiglietti, PhD, MBA  
Katia Ferguson, MSN  
Jeffrey Ferranti, MD, MS  
Lisa Ferrari, MD  
Katie Flanagan, MSW, LCSW, ACM-SW  
Christine Fowler  
Priscilla Ramseur, DNP, RN, CNOR, NEA-BC  
Marat Fudim, MD  
William J Fulkerson, MD, MBA  
Anthony Fuller, MD, MScGH  
Adia Ross, MD  
David Gallagher, MD  
Ziad Gellad, MD, MPH  
Charles Gerardo, MD  
Jennifer Gilner, MD, PhD  
Christine Goertz, DC, PhD  
Ben Goldstein, PhD  
Jamie Gollon, METM  
Oren Gottfried, MD  
Larry Greenblatt, MD  
Barbara Griffith, MD  
Daniel Guidot, MD  
Padma Gulur, MD  
Ashraf Habib, MBBCh, MSc  
Michael Haglund, MD, PhD, MEd  
David Halpern, MD  
Brian Halstater, MD  
Matthew Hartwig, MD  
Randy Harvey  
Laura Havrilesky, MD, MHSc  
Ricardo Henao, PhD  
Adrian Hernandez, MD  
Molly Hillenbrand, MD
DIHI Collaborators

Matthew Roman, MHA, MMCi
Tyrus Rorick
Adia Ross, MD, MHA
Matt Rougeux, MHA
Moira Rynn, MD
John Sampson, MD, PhD, MBA, MHS C
Devdutta Sangvai, MD, MBA
Natalie Sayewich, MA
Edna Scarlett
Mary Schilder, RN
Kelsey Seidl, MSN
Mya Sendak, MD
Teri-Lynne Sennett
Kevin Shah, MD
Svati Shah, MD
Colleen Shannon, JD
Richard Shannon, MD
Afreem Shariff, MD
Shilpa Shelton, MHA, FACHE
John Shepherd
Kristen Shirley, MD
Gail Shulby, RN, MA, CPPS
Lauren Siewny, MD
Justin Silverman, MD, PhD
Stephen Skelton
Becky Smith, MD
Isaac Smith, MD
Ryan Smith
Stuart Smith, MHA
Denise Snyder, MS, RD

Laurie Snyder, MD, MHS
Mary Solomon, MS
Michael Spiritos, MD
Emily Sterrett, MD
Allison Stewart
Neel Subramanian, MD
Debra Sudan, MD
Aparna Swaminathan, MD
Heather Talley, MSN
Dustin Tart, RN
Tiffany Taylor, MHA
James Tcheng, MD
B. Jason Theiling, MD
Nathaniel Thompson, MD
Carey Unger, MHA
Hope Uronis, MD, MHS
Sreethanth Vemulapalli, MD
Anthony Viera, MD, MPH
Momen Wahidi, MD, MBA
Wendy Webster, MA, MBA
Zac Ginsberg, MD, MPP
Andy Weinhold, MPH
E. Hope Weissler, MD
Walter Wiggins, MD, PhD
Julius Wilder, MD, PhD
Willette Wilkins, MBA
Matthew Wilson, RN
Samantha C. Wong, BS
Mary Cooter Wright, MS
Michael Yarrington, MD, MMCi