HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION

BAHM BUSINESS SCHOOL ALLIANCE FOR HEALTH MANAGEMENT

Becca Nelson, Stanford University School of Medicine, Margaret C Nikolov, Clinical Excellence Research Center, Stanford University School of Medicine, and Kevin Schulman, Stanford University School of Medicine, Clinical Excellence Research Center, Graduate School of Business, Stanford University

Contact: kevin.schulman@stanford.edu

Abstract

What is the message? The United States has seen an increase in healthcare administrators, both absolutely and as a ratio to physicians, with a significant acceleration of these trends, since the 2010 passage of the Affordable Care Act.

What is the evidence? Analysis of data obtained from the U.S. Bureau of Labor Statistics Occupation Employment and Wage Statistics (OEWS) data.

Keywords: Healthcare Administration, Healthcare Workforce, Affordable Care Act

Timeline: Submitted June 30, 2026; accepted after review June 30, 2026.

Cite as: Becca Nelson, Margaret C Nikolov, Kevin Schulman. 2026. The Growth in Healthcare Administrators since 2004 in the United States. *Health Management, Policy and Innovation* (www.HMPI.org). Volume 11, Issue 2.

Introduction

Healthcare administration costs in the United States are the highest of any country,[1] yet understanding factors driving these costs remains poorly characterized. One open question is the growth of administrators in the healthcare system.[2] Since the passage of the Affordable Care Act (ACA) in 2010, the healthcare delivery system has been transformed by provider consolidation, the migration of physicians to an employment model, consolidation in the health insurance market, and the implementation of myriad value-based payment programs in the public and private sector.[3,4]

We sought to better characterize the growth of administrators in the U.S. relative to the growth of practicing physicians, and to assess whether the ACA and the subsequent changes to the healthcare market had an impact.

Methods

Data on healthcare administrators was obtained from the U.S. Bureau of Labor Statistics Occupation Employment and Wage Statistics (OEWS) data. Healthcare administrators were defined using the O*NET code 11-9111 for healthcare administrators. O*NET defines healthcare administrators as individuals who "Plan, direct, or coordinate medical and health services in hospitals, clinics, managed care organizations, public health agencies, or similar organizations." Physicians were defined using all O*NET codes that cover physicians which varied by year over the course of the dataset.

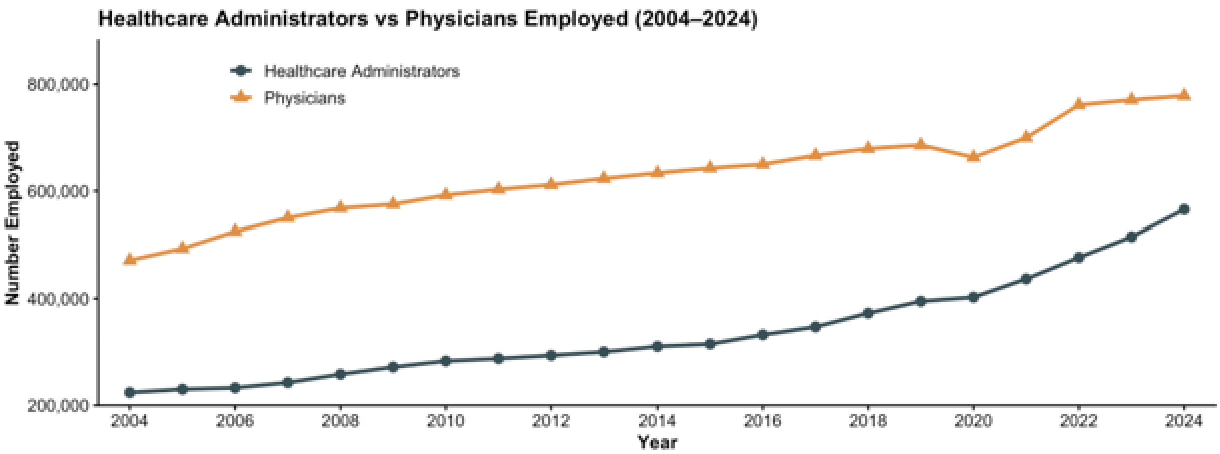
The ratio of healthcare administrators to physicians was computed for each year. Ratios pre-ACA (years 2004-2010) and post-ACA (2011-2024) were compared using one- and two-sample t-tests. Trend over time was assessed using simple linear regression, stratified by period (pre- vs post-ACA). Data analysis was conducted in RStudio v2025.05.1.

Results

The number of healthcare administrators increased from 224,070 in 2004 to 565,840 in 2024 while physicians saw an increase from 471,390 to 778,450 over the same period (Figure 1). Over this period, the ratio of healthcare administrators to physicians increased from 0.48 in 2004 to 0.73 in 2024.

Figure 1: Growth of Healthcare Administrators and Physicians 2004-2024 in the

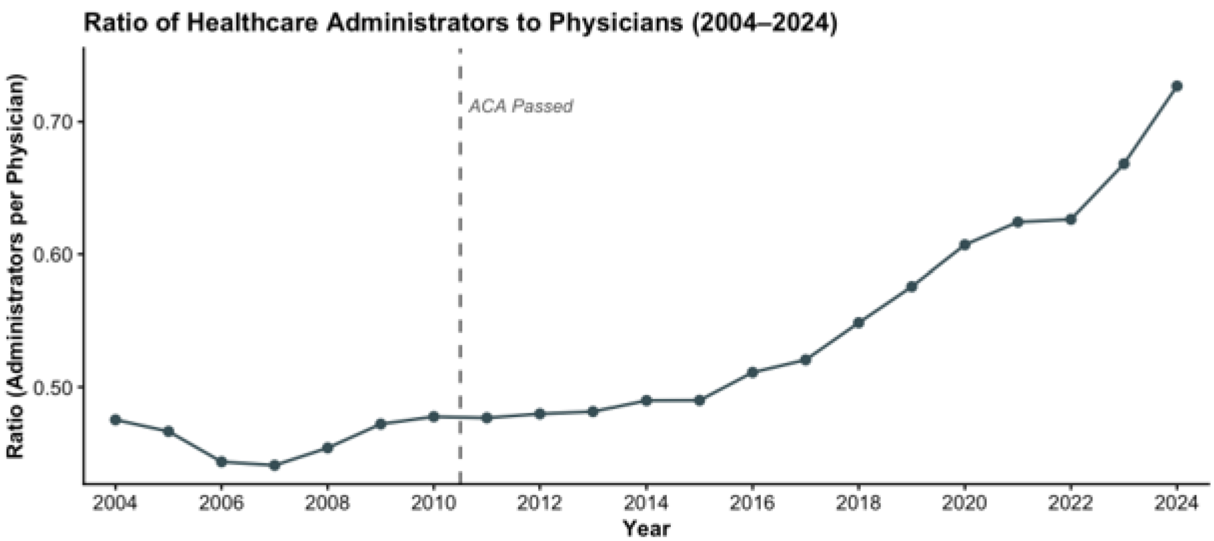
United States.



Analysis of U.S .Bureau of Labor Statistics Occupation Employment and Wage Statistics (OEWS) of the number of healthcare administrators and physicians over time. Healthcare administrators grew from 224,070 in 2004 to 565,840 in 2024 while physicians grew from 471,390 to 778,450 in the same period.

In examining changes in the market since the ACA, the mean ratio of healthcare administrators increased significantly from 0.46 (95% CI, 0.45-0.48) pre-ACA to 0.56 (95% CI, 0.51-0.61) post-ACA ($p < 0.01$) (Figure 2). Plotting the ratio of healthcare administrators to physicians over time, the trendline is constant pre-ACA ($\hat{\beta} = 0.001$ [95% CI, -0.01-0.01]), but is increasing over time post-ACA ($\hat{\beta} = 0.02$ [95% CI, 0.02-0.02]). (Figure 2)

Figure 2: Ratio of Healthcare Administrators to Physicians 2004-2024 in the United States.



Analysis of US Bureau of Labor Statistics Occupation Employment and Wage Statistics (OEWS) of the ratio of healthcare administrators to physicians over time. The dashed line indicates passage of the Affordable Care Act (2010). The ratio stayed relatively constant around 0.47 until 2011 when the ratio increased regularly until it reached a value of 0.73 in 2024. The trendline for the data points after the ACA from 2011 -2024 is increasing over time whereas there is no difference in the trendline from 0 pre-ACA.

Discussion

The data demonstrate an increase in healthcare administrators both absolutely and as a ratio to physicians, with a significant acceleration of these trends since the passage of the ACA in 2010. (Figure 1,2) The healthcare market has become more complicated over the last decade structurally on the delivery system side, and financially with myriad different payment models, each with distinct prior authorization and data reporting requirements. The result is the growth of non-clinical administrators across the system. Moreover, this growth is continuing (rather than reaching a new plateau). The implementation of technology during this period has not been associated with productivity gains seen elsewhere in the economy.[5] Unfortunately, this significant growth in expense and complexity has been associated with a decline in value in the delivery of services to patients.[6]

Reliable and reproducible data sources for the administrative workforce in healthcare are not readily available.[7,8] Data sources providing information about employment of health administrators are imperfect; OEWS data can be recoded frequently and may not include all administrative roles in healthcare systems, but the strength of the data are the survey sample of 1.1 million establishments and detail of 830 occupational categories.

The changes in the healthcare market since the passage of the ACA has led to a significant growth in the number of administrators across the healthcare system, and in the ratio of the number of administrators per physician.

References

1. Turner A, Miller G, Lowry E. High US health care spending: where is it all going? Commonwealth Fund. Published October 4, 2023. Accessed February 26, 2026. https://www.commonwealthfund.org/publications/issue-briefs/2023/oct/high-us-health-care-spending-where-is-it-all-going
2. Cutler DM. The good and bad news of health care employment. 2018;319(8):758-759. doi:10.1001/jama.2018.1054
3. Cooper Z, Gaynor M. Addressing hospital concentration and rising consolidation in the United States. One Percent Steps for Healthcare Reform. Accessed February 26, 2026. https://onepercentsteps.com/policy-briefs/addressing-hospital-concentration-and-rising-consolidation-in-the-united-states/
4. Schulman KA. A requiem for value-based care. *Health Manag Policy Innov.* 2024;9(3).
5. Huckman R, Chigurupati A, Cutler D. The IT transformation health care needs. *Harvard Business Review.* Published November 2017. Accessed February 26, 2026. https://hbr.org/2017/11/the-it-transformation-health-care-needs
6. Lescinsky H, Sahu M, Beauchamp M, et al. Exploring state-level change in health care value over three decades in the United States, 1991-2020. *Health Serv Res.* 2026;61(1):e70054. doi:10.1111/1475-6773.70054
7. Brooke Istvan, Heather Bello Thornhill, Karen E Joynt-Maddox, David Cutler, Ezekiel J Emanuel, Availability of consistent, reliable, and actionable public data on US hospital administrative expenses, *Health Affairs Scholar*, Volume 3, Issue 5, May 2025, qxaf069, https://doi.org/10.1093/haschl/qxaf069.

8. Drum K. Join Me On a Dive Down the Rabbit Hole of Health Care Admin Costs. Mother Jones. June 15, 2019. https://www.motherjones.com/kevin-drum/2019/06/join-me-on-a-dive-down-the-rabbit-hole-of-health-care-admin-costs/. Accessed March 5, 2026.

Southern California Permanente Medical Group: The Care Transformation Office (CTO) – Scaling Integrated Value-Based Care Through Strategic IT-Clinical Partnerships

(Stanford Graduate School of Business Case SM-403, May 7, 2026; 15 pages)

Authors: Sinjin Lee, Dinh Nguyen, Khang Nguyen, Ramin Davidoff, and Kevin A. Schulman

Southern California Permanente Medical Group (SCPMG), part of Kaiser Permanente, the largest nonprofit integrated health care system in the United States, faced the challenge of rising patient messages and the expectation of on-demand care, which added to the risk of physician burnout. Could SCPMG leverage the latest AI technologies to enhance integrated care delivery and streamline operations? And what would be the optimal approach to integrate AI-powered tools throughout Kaiser Permanente's extended network of physicians and care teams that served more than 12 million Kaiser members?

Kaiser Permanente had a long history of innovation—and disruption—in American health care, dating back to the organization's origins in 1933. This case study explores a new chapter in health care innovation, as doctors and clinics sought to improve patient care and provide quicker response times, particularly in managing low-acuity and chronic care cases. Hybrid care models, for instance, could help reduce the demand for in-person care, yet preserve Kaiser's focus on providing the highest quality of care to each patient and enhancing the patient experience. By exploring patient and clinician needs, Kaiser's investigations revealed new opportunities for future AI-driven innovations in triage, bot-based communications, predictive analysis, and clinical decision support. Kaiser acknowledged that successful integration of new processes and technologies would require deep integration between IT capabilities and clinical workflows—and an organizational culture aligned with the advantages AI could offer.

Learning Objective

The case study offers a framework for students to understand and discuss the integration of AI technologies in the health care sector. By discussing the Kaiser Permanente group's experiences, students can learn about one approach to scale integrated care delivery in the rapidly evolving health care landscape.

This material is available for download by current Stanford GSB students, faculty, and staff, as well as Stanford GSB alumni. For inquires, contact Kevin Schulman: kevin.schulman@stanford.edu

Lawton Robert Burns, The Wharton School, University of Pennsylvania

Contact: burnsl@wharton.upenn.edu

"Concise Introduction to Healthcare Management" (Elgar Publishing, 2026)

Two topics

Like many of you, I teach our introductory course on the healthcare ecosystem and how to manage it. I have done so for nearly four decades at various schools. I also have the privilege of serving as an editor of the Shortell & Kaluzny textbook (along with Betsy Bradley and Bryan Weiner).

Healthcare and management are two different fields: a huge industry and a diverse discipline. The *Healthcare* industry spans 25 different "verticals" along a value chain that contract as buyers and suppliers. *Management* textbooks focus on a core set of functions needed to run these companies.

Elgar asked me to write *A Concise Introduction to Healthcare Management* (not easy for a windbag like me). This volume combines these two topics into what is hopefully an accessible and engaging text, discussing healthcare management using seven different approaches. Consider them as alternative frames of reference and skillsets.

- ***One approach*** is to have students confront a series of dilemmas and paradoxes. Wrestling with these is a source of cognitive dissonance and a major source of learning, which lend themselves to the Socratic style of teaching. Healthcare is full of conundrums, which is why I start the book with them. They dispel some conventional wisdom that is often incorrect.
- ***A second approach*** identifies the major sources of complexity in healthcare. Management sage Peter Drucker famously referred to hospitals as the most complex human organization ever devised. Consider that hospitals are a matrix inside a dual hierarchy inside a conglomerate. It is crucial that students know and embrace this complexity. Healthcare has multiple professions, each with specialized knowledge, and each of which thinks they are the center of the universe. Such knowledge workers make the system productive. Drucker explained, "In a traditional workforce, the worker serves the system; in a knowledge workforce, the system must serve the worker." In practice, this means the typical organizational pyramid must be stood on its head.
- ***A third approach*** identifies the memes that dominate the fields of management and healthcare. These are ideas or phrases that become "sticky" by virtue of sheer mindless repetition (e.g., management buzzwords). They, unfortunately, substitute for rigorous analysis and critical thinking. Critical thinking as a vehicle to confront memes is an exercise which students find humorous; more importantly, they find it memorable and useful for discerning what is really going on around them. Your first task is to recognize a meme when you hear it; your second task is to know why it is (likely) bogus; your third task is to not repeat it to others; your fourth task is to stop wasting your time and hunt for another solution.
- **A *fourth approach*** tackles the dominant theme of integration and coordination. If THE big issue in healthcare is fragmentation (as many argue), then integration must be the solution. Integration and coordination are "must haves" in any management approach. The book explores the many management tools that have been deployed to bring them about. These include managing professionals (who need to be integrated), teams, integrated

delivery networks, and accountable care organizations. Unfortunately, the evidence base that supports the utility of integration and coordination needs reinforcement.

- ***A fifth approach*** is to identify the basic relationships that impact patient outcomes and firm success and, thus, need to be managed. These include relationships between people (e.g., patient-physician, physician-manager, scientist-manager) and between sectors (e.g., payer-provider, employer-payer). The first three of these occur at the micro-level and serve as the foundation for the rest of the healthcare ecosystem; the latter two illustrate the stakeholder relationships that need to be managed. Healthcare is more an enterprise that rests on human capital and relational capital than on technological capital. Relationships are key to hospital quality.
- ***A sixth approach*** concentrates on the important management skills that students should develop. These include learning, managing hairy problems, managing cost and quality, managing change, managing process (not structure), and focus. These topics have bedeviled the healthcare ecosystem for decades and pose perhaps the greatest challenge for students and managers.
- **A *seventh approach*** is to "disabuse" the student about some likely preconceived notions and persuade them that an idea or belief is mistaken. Here, the mistake is to believe (as many do) that new technology will solve healthcare's problems. I have seen this approach come and go under many guises. The book deconstructs many of the new technologies that currently fascinate everyone in the healthcare ecosystem. If healthcare is all about people and relationships, your task will be to figure out how to marry and harmonize that with technology – – what is called the "socio-technical" system approach.

This volume seeks to live up to the title of the Elgar series of which it is part: "Concise". It serves as a "pocket guide" (165 pages !!) to those looking for a quick introduction to the field and an overview of "what do I need to know to manage". It should thus appeal to executives, healthcare professionals, and students in professional programs. The book also features plenty of humor that makes it an enjoyable and memorable read.

Elman Amador Medina, Graduate School of Business, Stanford University; Steven Gong, Dashiell Miner, School of Medicine, Stanford University; Lauren Bilbo, Klara Klarowicz, Graduate School of Business, Stanford University; Sergio Mavridis, David Scheinker, Stefanos Zenios, Graduate School of Business, Stanford University; Kevin Schulman, Graduate School of Business, School of Medicine, Stanford University

HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION

BAHM BUSINESS SCHOOL ALLIANCE FOR HEALTH MANAGEMENT

Support: The Ludy Family Foundation, The Hirsch Family Foundation, The Mindshare Institute, and the Government, Business and Society Initiative of the Graduate School of Business, Stanford University.

Abstract

What is the message? Healthcare fraud has always been a major concern for health payers. While public debate often equates healthcare fraud with improper beneficiary enrollment, most documented losses arise from provider-side and organized criminal schemes. The structural complexity of the U.S. healthcare and lack of governance over the market has resulted in significant lags in response emerging artificial intelligence (AI)-enabled fraud. This paper proposes a technical framework to transition the industry from reactive fraud detection to proactive fraud prevention by adapting multi-layered defense mechanisms proven effective in the financial services sector.

What is the evidence? The proposed solution comprises a three-layer architecture integrated into claims processing: (1) Healthcare Identity Management, analogous to Know Your Customer (KYC) protocols, for rigorous provider and patient identity verification; (2) Continuous Claim Validation using “Zero Trust” principles, where each claim is assumed to be fraudulent until proven valid; and (3) AI/Machine Learning (ML)-Powered Real-Time Detection using advanced analytics and collaborative intelligence to identify aberrant patterns before payment authorization.

Keywords: Artificial Intelligence, Healthcare Fraud, Health Insurance

Timeline: May 31, 2026; accepted after review June 12, 2026.

Cite as: Elman Amador Medina, Steven Gong, Dashiell Miner, Lauren Bilbo, Klara Klarowicz, Sergio Mavridis, David Scheinker, Stefanos Zenios, Kevin Schulman. 2026. The Emerging Risk of AI-Enabled Healthcare Fraud. Health Management, Policy and Innovation (www.HMPI.org). Volume 11, Issue 2.

Introduction

Transaction fraud in claims and reimbursement is an unfortunate feature of the U.S. healthcare market. Improper payments, generally attributed to fraud and abuse, are thought to exceed $100 billion annually across the Medicare and Medicaid programs [1]. Fraud is criminal activity to steal money through the insurance claims process, while abuse describes non-criminal clinical practices or billing patterns. This paper focuses on the issue of payment fraud.

Healthcare fraud takes two distinct forms. Provider-side fraud occurs when clinicians, suppliers, or organized criminal groups bill payers. Beneficiary fraud occurs when individuals enroll in coverage or use services to which they are not entitled.

While recent policy debates, including the Medicaid eligibility provisions of the One Big Beautiful Bill Act of 2025, have centered on beneficiary fraud, enforcement data tell a different story [2]. The large majority of documented fraud losses come from provider-side schemes and organized criminal enterprises [3-5]. Several patterns of provider-side fraud have been identified: medical identity theft, billing for unnecessary services, billing for services not provided, upcoding, unbundling, and kickbacks [3]. Historically, fraud has been a particular issue in the Medicare program, where Medicare must pay for services within a fixed period of time. When it discovers that it paid fraudulent claims, it must “chase” after the biller for a recovery (the so-called pay and chase model). Medicare began to check for fraud prior to payment in 2011 [6], but fraudulent billing still persists [4].

AI-Enabled Fraud

These legacy issues of fraudulent billing are now facing new challenges with the emergence of artificial intelligence (AI). The convergence of generative AI and healthcare’s analog infrastructure is creating an unprecedented fraud escalation.

According to data security vendors, deepfake fraud[1] incidents surged globally by over 700% between early 2024 and early 2025, with North America experiencing a 1,100% increase over the same period [7]. Synthetic identity document fraud—in which AI-generated credentials combine real and fabricated personal data—rose 311% in North America during Q1 2025 compared to Q1 2024 [7].

These capabilities are already being weaponized against healthcare payment systems. Deepfake medical identities enable phantom billing schemes, fraudulent prescription access, and insurance claims filed under fabricated patient profiles. Voice cloning technology, which now requires as little as three to five seconds of audio to produce a convincing replica, has driven a 475% year-over-year increase in attacks targeting insurers [8]. Telemedicine platforms present particular vulnerabilities; the 2025 DOJ National Health Care Fraud Takedown included $1.17 billion in fraudulent telemedicine and genetic testing claims, with foreign actors using AI to generate fraudulent audio recordings of Medicare beneficiary consent [4]. In August 2025, the Department of Justice announced its first healthcare AI fraud enforcement action against Troy Health, Inc., which used an AI platform (Troy.ai) to fraudulently enroll over 300 Medicare beneficiaries in a single day without their knowledge or consent—a harbinger of how AI reduces the marginal cost of fraud execution [9]. CMS prevented over $4 billion in fraudulent claims from being paid in the lead up to the June 2025 enforcement action [10].

Nation State Threat Actors

Healthcare's vulnerability extends beyond financially motivated criminal enterprises to include advanced persistent threat (APT) groups operating with the tacit or explicit support of foreign governments. These groups can target intellectual property for theft and commit financial crimes.

Iranian state-sponsored group Pioneer Kitten provides network access to ransomware affiliates who deploy attacks against healthcare systems, creating a cooperative model between nation-state espionage and criminal extortion [11]. North Korean actors deployed the Maui ransomware specifically against U.S. hospitals between 2021 and 2022, with ransom payments funding the state's nuclear weapons program [12]. Chinese APT groups have targeted biomedical research institutions for COVID-19 vaccine intellectual property theft and deployed malware through medical imaging DICOM files [13]. Russian-linked APT29 (Cozy Bear), attributed to the Russian Foreign Intelligence Service (SVR), has conducted persistent espionage campaigns against healthcare and pharmaceutical organizations in the United States and Europe [14]. Ransomware attacks against healthcare organizations have increased 300% since 2015, and in 2024, healthcare reported the most cyber threat incidents of any sector: 458 ransomware incidents targeting the health sector [15][16].

In June 2026, Google sued a Chinese enterprise, Outsider Enterprise, for using Gemini's AI tools to create fake websites including websites that appear like government sites including the Postal Service and New York State's E-Z Pass system. Google reported that hundreds of thousands of people had been exposed to these fraudulent sites. Google was cooperating with the Federal Bureau of Investigation in the case, with wireless carriers to try to shut down the network. [17]

These APT groups could use AI technology to develop billing fraud schemes using advances in AI technology.

Data Security

The full scope of the 2024 Change Healthcare data breach included 192.7 million individuals, nearly two-thirds of the U.S. population, making it the largest healthcare data breach in history [18]. The breach exposed names, Social Security numbers, dates of birth, medical information, driver's licenses, passports, and financial payment card data [17]. The Change Healthcare attack is not an isolated incident but part of an accelerating pattern. In 2024, the healthcare sector experienced 742 breach incidents affecting over 276 million records [19]. Since 2020, more than 500 million individuals, exceeding the entire U.S. population, have had healthcare records stolen or compromised [19]. All of these identify data on individual patients can be available on the dark web to further develop claims fraud schemes.

Escalating Threats

These escalating threats—AI-enabled fraud, nation-state exploitation, and systemic breaches—succeed not because they are undetectable, but because the healthcare transaction infrastructure was never designed to address these challenges.

Several structural deficiencies in the current claims processing environment create the conditions that both enable and conceal fraud at scale:

Fragmented Processing: A single claim often passes through five to seven different, often disconnected, systems before adjudication [20]. Each handoff represents a potential point of data degradation or manipulation, making holistic transaction analysis difficult.

Lack of Standardization: While X12 is the standard format for electronic data interchange

(EDI), [21], each payer uses distinct "companion guides" governing specific implementation rules. This variability forces providers to manage dozens of slightly different workflows, creating operational friction that disincentivizes participation by smaller payers and creates loopholes that sophisticated fraudsters exploit [22].

Incomplete Data for Detection: Current fraud detection systems are often trained on public or siloed claims data lacking integration with clinical and practitioner data from Electronic Health Records (EHRs). This incomplete picture limits fraud prediction model accuracy, as payers rarely possess the information needed to algorithmically validate claims [23].

Limitations of Supervised Fraud Models: Supervised fraud detection models depend on court validation of previous fraud cases. If a model incorrectly processes a fraudulent claim as legitimate, similar fraudulent claims may subsequently evade detection.

Legend: ChatGPT completed CMS-1500 professional claim form, the standard form used by non-institutional providers and suppliers to submit claims to Medicare, using synthetic data. Source:

National Uniform Claim Committee (NUCC); study authors.

Together, these infrastructure gaps create a compounding vulnerability: fragmented systems prevent the holistic visibility needed to detect fraud, while the absence of standardized, real-time data pipelines makes it impossible to apply the kind of continuous transaction monitoring that has proven effective in financial services. The result is a system structurally locked into post-hoc detection.

The persistence of the reactive model in healthcare creates a systemic barrier to modernization. Payers are hesitant to adopt fully automated claims processing because, without intelligent safeguards, automation could scale fraudulent payments alongside legitimate ones. This creates a critical challenge: payers must either continue with expensive, labor-intensive manual reviews that are overwhelmed by volume, or invest in automated systems equipped to differentiate between legitimate and fraudulent claims in real time.

This paper proposes a strategic and technical framework to address this challenge by adapting proven, multi-layered defense mechanisms from the financial services sector.

Banking Sector Lessons

The financial services industry, shaped by decades of evolving fraud schemes, has developed multi-layered defenses that offer a blueprint to prepare healthcare for the emerging threat of AI-driven fraud mechanisms. These mechanisms have reduced electronic payment fraud losses to a few cents per $100 transacted—a level of integrity to which healthcare can aspire [24].

The banking industry has achieved a 6% to 10% improvement in fraud detection accuracy using AI-powered, real-time systems that analyze transactions in milliseconds [25]. These strategies have also led to the development of novel identity and transactions enterprises that are focused on fraud prevention (Plaid and Stripe).

Insurers in other domains have reported a 30% increase in fraud detection rates and returns on investment (ROI) exceeding 210% within the first year of implementing similar technologies [26]. Deloitte projects that comprehensive AI-driven approaches could save property and casualty insurers between $80 billion and $160 billion by 2032 [27].

Core Banking Anti-Fraud Mechanisms

Robust Customer Identification: Financial institutions use Know Your Customer (KYC) processes to verify identity and credentials of new customers, performing background checks and risk scoring to prevent bad actors from entering the system. This continuous process of identity verification forms the first line of defense against fraud.

Multi-Layer Payment Authorization ("Zero Trust"): No transaction is trusted by default. Layers of verification—including PINs, biometrics, one-time passwords, and geolocation matching—are dynamically applied based on transaction risk profiles.[2] This philosophy assumes every transaction may be fraudulent until proven otherwise through automated verifications.

Continuous Transaction Monitoring (Anti-Money Laundering Analytics): Banks employ real-time anti-money laundering (AML) analytics to detect suspicious transactions as they occur. Algorithms flag anomalies against expected behavior—such as unusual spending patterns, transactions from high-risk geographies, or rapid fund movements—triggering automated reviews or holds on funds.

Translation to Healthcare

These principles can be translated to the healthcare ecosystem:

KYC → Member Credentialing

One of the most familiar aspects of the US healthcare system is the need to copy or type the member health plan identification information into a provider registration system. This simple process highlights one of the most critical failures of the current healthcare system: it is based on an analog process of identification where data can easily be copied incorrectly, while data on the patient and the plan (and the plan contract with the provider) are not digitally linked.

The simple contrast between a typical health insurance card and a credit card is illustrative. The health insurance card carries no digital information; it even lacks the most rudimentary digital technology of barcodes.

Legend: Medicare insurance card and Visa card with embedded EMV chip.

Source:
https://www.medicare.gov/basics/get-started-with-medicare/using-medicare/your-medicare-card; Bachelot Pierre J-P, CC BY-SA 3.0 https://creativecommons.org/licenses/by-sa/3.0, via Wikimedia Commons

The scale of the gap between banking and healthcare payment security is well documented. The average cost to process a single primary care claim is $20.49 [28] and claims routinely traverse a multi-step adjudication process taking days to months. By contrast, the Visa payment network processes approximately 1,700 transactions per second [29], with AI-powered fraud detection analyzing over 500 features per transaction in real time and preventing $25 billion in fraud in 2019 alone [30]. Banking has developed the EMV (an acronym for Europay, Mastercard, and Visa) chip-enhanced payment security that establishes a unique digital key for each transaction. EMV chip technology reduced counterfeit payment fraud by 76% within three years of adoption [31,32], although this effort pushed fraud to card-not-present transactions [33]. Healthcare's estimated fraud rate of 3% to 10% of total spending, representing $59 billion to $84 billion across the U.S. healthcare system, persists in part because the sector lacks any comparable realtime digital verification mechanism [34].

Member credentialing also faces adoption obstacles that banking did not. Public resistance to digitally coded patient identity is well documented: HIPAA directed HHS to create a unique health identifier for individuals in 1996, but Congress has prohibited funding it since 1999 over privacy objections, and a 2019 House vote to lift the ban stalled in the Senate [35]. National survey data reflect the same sensitivity, with roughly seven in ten individuals reporting concerns about the privacy of their medical records, even as a similar majority supports electronic records and health information exchange [36]. Access is a second obstacle: 91% of U.S. adults own a smartphone and 78% have home broadband, with adoption lowest among older and lower-income adults, the populations most likely to depend on Medicare and Medicaid [37]. Digital member credentials will therefore require clear communication about what data they carry and how they are used, simple non-digital fallback processes for members who cannot or will not adopt them, and visible benefits such as faster eligibility verification at the point of care. Implementation of this solution would have to account for the challenges of vulnerable patient populations, patients with disabilities, and patients that lack full broadband connectivity. Of course, in-person verification of patient identification at the time of a visit is a robust security check.

KYB-Provider Credentialing and Enrollment: Currently, each health plan must develop a provider directory, creating duplication across plans and increasing paperwork burden for providers. A secure KYB transaction environment would establish secure digital credentials (National Provider Identifier [NPI], licensure, billing entity, billing location) for transaction platform entry. Since fraudulent claims submissions is the largest scale risk of AI-generated fraud, KYB efforts likely have the highest return on investment in developing ore robust identity schemes.

Prepayment audits and account verification would establish an additional level of security. Real-time screening of network applicants against the HHS Office of Inspector General's (OIG) List of Excluded Individuals/Entities (LEIE) [38] could further reduce entry of potentially fraudulent entities. The Centers for Medicare and Medicaid Services' (CMS) enhanced screening programs have already deactivated over 660,000 suspect enrollment records from 2011 to 2015, demonstrating the scale of this opportunity [39].

Zero-Trust Payments → Zero-Trust Claims: Payers need systems to validate each transaction as legitimate: verifying the provider, patient, and transaction itself. Banks validate credit card transactions through EMV (Europay, Mastercard, and Visa) chips using one-time codes. According to Visa, chip-enabled cards reduced card-present counterfeit payment fraud by 76% at chip-enabled merchants within three years of the 2015 U.S. EMV transition [32]. A zero-trust model for healthcare would require that high-risk claims clear additional automated and, when necessary, human checks before payment authorization. The GAO has noted that pre-payment audits are more cost-effective than post-payment recovery efforts [40].

AML Analytics → Claims Pattern Recognition: A digital backbone for the transaction system enables new tools to detect digital fraud, from unexpected surges in claims from a single NPI to ensuring that coded services reflect a provider's practice pattern. Graphical analysis of payer-provider relationships can be used to further enhance the effectiveness of this security layer.

Implementation Framework

This framework adapts the banking sector's three-layer architecture to healthcare fraud prevention, integrating defenses directly into claims workflows

Layer 1: Healthcare Identity Management (KYC/KYB)

The first layer prevents fraudulent actors from entering the system by strengthening provider and patient identity controls.

Enhanced Provider Credentialing: Implement rigorous, automated, continuous credentialing processes including: primary source verification of licenses through API calls to state medical boards; continuous background checks against federal and state exclusion lists (such as the OIG's LEIE, which contains over 660,000 excluded parties); and physical or virtual validation of practice addresses to expose phantom clinics [40]. Providers can be stratified by risk level, with higher-risk entities (e.g., durable medical equipment providers) facing enhanced due diligence.

Digital Provider and Patient ID Authentication: Introduce modern identity-proofing for claim submission, such as multi-factor authentication or digital certificates tied to provider identity. This prevents use of stolen billing credentials. Similarly, strengthening patient identification at the point of service, potentially through biometric-linked insurance cards or secure digital wallets, can reduce identity swapping and medical identity theft. REAL ID verification is a powerful identity mechanism now in place [41]. Patient identity verification should leverage existing FHIR-based standards to enable interoperable identity tokens that travel with the patient across providers and payers.

Unified Provider Directory Infrastructure: Rather than each health plan maintaining its own provider directory, a centralized or federated provider registry, modeled on the NPPES database but enhanced with real time credentialing status, would reduce duplication, lower administrative burden, and close gaps that enable phantom providers to enter the system undetected. Such infrastructure already has precedent in CMS's Provider Enrollment, Chain, and Ownership System (PECOS), though the current system lacks real time validation capabilities.

Layer 2: Continuous Claim Validation (Zero Trust)

This layer applies the "never trust, always verify" principle to every transaction.

Pre-Payment Review and Smart Adjudication: High-risk claims identified by analytics are automatically routed for pre-payment review or checked against additional data before disbursement. This approach shifts from "pay and chase" to proactive loss prevention.

Machine-Readable Digital Contracts: The technical foundation uses machine-readable standards (X12, FHIR) to encode provider agreements, fee schedules, and benefit plans into computable formats [21]. When a claim is submitted, it can be automatically validated against digital contract terms. For example, a system could automatically deny a second claim for a procedure that a contract specifies is covered only once annually or flag a claim for a service billed by a provider whose digital contract excludes that service. Research indicates that decision tree algorithms embedded in smart contracts can achieve high accuracy in detecting contractual fraud [41].

Clinical Plausibility Engine: Beyond contractual validation, claims should be assessed against clinical logic rules that reflect medical practice standards. For instance, a claim billing for a complex surgical procedure at a facility lacking the appropriate licensure or equipment, or a claim for an office visit submitted on the same day as an inpatient admission, represents clinically implausible activity that can be flagged automatically. Integration with EHR data, where available and permissible under HIPAA's "minimum necessary" standard, would further strengthen clinical validation without requiring manual chart review. Risk stratification engines would assign dynamic risk scores to each claim based on provider history, claim characteristics, patient demographics, geographic patterns, and clinical plausibility. Only claims exceeding a configurable risk threshold would require enhanced scrutiny, allowing the vast majority of legitimate claims to process at current or faster speeds.

Layer 3: Advanced Analytics and Risk Scoring (AI/ML)

This layer deploys advanced analytics to detect anomalous behavior in real time.

Predictive Modeling and AI: ML models trained on historical claims data generate real-time fraud risk scores for each new claim. High-risk claims are automatically flagged for intervention. These models continuously learn from confirmed fraud cases, adapting to new schemes and improving accuracy while reducing false positive rates over time. CMS's Fraud Prevention System has demonstrated greater than 3-to-1 ROI [42].

Graph and Outlier Analysis: Graph analytics visualizes and analyzes complex, often hidden networks connecting providers, patients, claims, and locations. This technique effectively uncovers collusion, such as kickback rings or organized fraud using multiple shell companies linked to a common address [43].

Collaborative Defense Network: The framework leverages a federated data model inspired by the Healthcare Fraud Prevention Partnership (HFPP). As of 2016, HFPP represented over 65% of covered lives in the U.S. and has since grown to over 300 partner organizations as of 2023 [44][45]. Participants contribute anonymized fraud pattern data and gain access to collective intelligence, enabling smaller payers to access enterprise-grade detection capabilities and larger payers to identify schemes spanning multiple networks.

The framework proposes expanding this federated model to include real time threat intelligence sharing, analogous to the financial sector's FS-ISAC (Financial Services Information Sharing and Analysis Center), enabling participating organizations to disseminate indicators of compromise as they emerge rather than in periodic retrospective reports. Network analysis can detect patterns that evade individual claim review, such as patient sharing rings, unusually concentrated referral patterns, and geographic clustering of high-cost services at addresses associated with minimal physical infrastructure.

Measuring Success

Effective implementation requires clearly defined performance metrics tracked across all three layers. Layer 1 (KYC) metrics include time to credentialing for new providers, percentage of enrollment applications flagged and confirmed as fraudulent, and reduction in phantom provider entries compared to baseline. Layer 2 (Zero Trust) metrics include percentage of claims validated against digital contract terms in real time, reduction in contractual fraud such as duplicate billing and out of scope services, and claim processing speed relative to pre implementation benchmarks. Layer 3 (AI/ML) metrics include fraud detection rate (true positive rate), false positive rate and its trend over time, dollar value of fraud prevented per dollar invested, and average time from claim submission to fraud identification. System wide metrics include overall improper payment rate for participating organizations versus nonparticipants, provider satisfaction and administrative burden metrics, and return on investment at each phase of implementation. These metrics should be independently audited and published to build the evidence base for broader adoption and to inform regulatory policy.

Risk Considerations

An aggressive fraud prevention program must be implemented thoughtfully to mitigate potential risks. Each layer of the framework carries potential unintended consequences for beneficiaries, providers, and payers. Acknowledging these risks explicitly and building mitigation into the design is as important as the fraud prevention itself. Provider risks are addressed in Section 6.3, payer risks in Section 6.4, and beneficiary risks in Section 6.5.

Regulatory and Legal Risks

All activities must comply with privacy laws including HIPAA and various state regulations, as well as prompt pay laws. While HIPAA permits use of protected health information (PHI) for fraud detection, a “minimum necessary” standard must be enforced. Pre-payment reviews must avoid violating claim payment timelines. To prevent False Claims Act liability, any automated denial process must include human oversight for ambiguous cases, a principle now required by law in states including California [46].

Cybersecurity and System Integrity

New analytics platforms and data-sharing networks expand cybersecurity risk. All systems require robust end-to-end encryption, multi-factor access controls, and regular penetration testing. If utilized, smart contracts require intensive code security audits to prevent exploitation of vulnerabilities [21].

False Positives and Provider Relations

The most significant operational risk is flagging legitimate claims (false positives), which can delay payments and damage provider relationships. This risk requires active management through:

- **Calibrating AI Thresholds:** Start with conservative settings flagging only high-confidence outliers, tuning to minimize provider friction initially while allowing threshold adjustment as models improve.
- **Transparent and Rapid Processes:** Establish clear, expedited appeals processes for providers to contest flagged claims.
- **Provider Education and Collaboration:** Communicate program goals and mechanics to network providers, framing it as collaborative effort to protect system integrity and accelerate payment for legitimate claims.

Pre-payment review also carries cash flow and workload consequences for providers. Physician practices already devote substantial resources to payer review processes: in the 2025 American Medical Association survey, physicians completed an average of 40 prior authorization requests per week, consuming roughly 13 hours of physician and staff time, and more than a quarter reported that these review processes led to serious adverse events for patients [47]. Fraud screening must not impose a comparable burden on honest providers. Risk-based exemptions offer one proven mitigation: under the Texas “gold card” law, providers whose requests are consistently approved earn exemptions from preauthorization review, a model that could similarly exempt consistently low-risk providers from enhanced claim scrutiny [48]. Clean claims from exempt providers should be paid within contractual and statutory prompt pay timelines, with flagging criteria disclosed to network providers and screening performance published as described in Section 5.4.

Implementation, Operational, and Payer Risks

AI system success depends on data quality. The project requires strong interdepartmental governance ensuring coordination between IT, compliance, claims, and provider relations. A phased rollout, starting with less complex claim types, and continuous performance monitoring are essential. For payers, the principal risks are financial and reputational: large technology investments may underdeliver if data quality is poor, and poorly calibrated models can create regulatory exposure and erode provider and member trust. Phased rollouts with published performance metrics (Section 5.4) help contain these risks. Payers also face model risk that is specific to an adversarial setting: fraud perpetrators actively probe and adapt to detection systems, so model performance degrades without continuous retraining, monitoring, and validation [49]. Governance should therefore include periodic independent model audits, documented retraining cycles, and fallback to human review when model confidence is low.

Risks to Beneficiaries

Beneficiaries ultimately bear the consequences of fraud prevention errors. A legitimate claim wrongly flagged can mean a delayed authorization, an unexpected bill, or an interruption in care. Any adverse action that affects access to care should therefore require human review before it takes effect, and patients should be held harmless while payer and provider disputes are resolved. Beneficiaries affected by a flagged claim should receive plain language notice and a simple appeal channel; payer review processes are already associated with care delays and treatment abandonment [50]. Stronger identity verification can also create access barriers for beneficiaries who lack smartphones, reliable internet access, or digital literacy; simple non-digital pathways to verify identity and receive care must be preserved. Aggregating identity, clinical, and financial data for fraud analytics raises privacy concerns beyond regulatory compliance, so programs should collect the minimum data necessary, disclose plainly how the data are used, and submit to independent audits. Finally, detection models should be monitored for disparate impact to ensure fraud controls do not systematically burden specific patient populations, such as those in underserved areas where documentation practices may differ.

Conclusion

The healthcare industry is increasingly vulnerable to threats of AI-enabled digital fraud, especially in the claims financial transaction system. Addressing this challenge will require transformation of the payment system to preserve the integrity of payment processes against this emerging risk by adopting a proactive prevention framework inspired by the financial services sector. A three-layer defense system built on rigorous identity management, zero-trust validation, and AI-powered analytics, can transform claims processing from a reactive, cost-intensive function into a proactive platform. Further leveraging machine-readable contracts as the foundation, payers can automate fraud detection at the point of submission, creating potential ROI that aligns financial incentives with technological modernization. Realizing these benefits will require managing the framework's risks to beneficiaries, providers, and payers with the same rigor as the effort applied to fraud prevention.

References

[1] U.S. Government Accountability Office. (2024, November). Medicare and Medicaid: Additional Actions Needed to Enhance Program Integrity and Save Billions (GAO-24-107487). https://www.gao.gov/products/gao-24-107487

[2] U.S. Department of Health and Human Services & Centers for Medicare & Medicaid Services. (2025). Trump Administration Prioritizes Affordability, Announcing Major Crackdown on Health

Care Fraud [Press release].
https://www.cms.gov/newsroom/press-releases/trump-administration-prioritizes-affordability-announcing-major-crackdown-health-care-fraud

[3] Centers for Medicare & Medicaid Services. (2016). Common Types of Health Care Fraud [Fact sheet]. https://www.cms.gov/files/document/overviewfwacommonfraudtypesfactsheet072616pdf

[4] U.S. Department of Justice. (2025, June 30). National Health Care Fraud Takedown Results in 324 Defendants Charged in Connection with Over $14.6 Billion in Alleged Fraud.
https://www.justice.gov/opa/pr/national-health-care-fraud-takedown-results-324-defendants-charged-connection-over-146

[5] National Health Care Anti-Fraud Association. (n.d.). The Challenge of Health Care Fraud.
https://www.nhcaa.org/tools-insights/about-health-care-fraud/the-challenge-of-health-care-fraud/

[6] California Health Advocates. (n.d.). Medicare Checks for Fraud Before Paying.
https://cahealthadvocates.org/medicare-checks-for-fraud-before-paying/

[7] Sumsub. (2025). Identity Fraud Report 2024–2025.
https://sumsub.com/blog/guides-reports/identity-fraud-report-2024-2025/

[8]). Insurance fraud increased by 19% from synthetic voice attacks in 2024. Fierce Healthcare.
https://www.fiercehealthcare.com/payers/insurance-fraud-increased-19-synthetic-voice-attacks-2024

[9] U.S. Department of Justice. (2025, August 20). Troy Health, Inc. Enters Non-Prosecution Agreement and Admits to Fraudulently Enrolling Medicare Beneficiaries and Identity Theft.
https://www.justice.gov/opa/pr/troy-health-inc-enters-non-prosecution-agreement

[10] Centers for Medicare & Medicaid Services. (2025, June). CMS Prevented Over $4 Billion in Fraudulent Claims. Referenced in the DOJ 2025 National Health Care Fraud Takedown announcement.

[11] CISA, FBI, and DC3. (2024, August 28). Iran-based Cyber Actors Enabling Ransomware Attacks on US Organizations (Advisory AA24-241A).

https://www.cisa.gov/news-events/cybersecurity-advisories/aa24-241a

[12] CISA, FBI, and U.S. Department of the Treasury. (2022, July 6). North Korean State-Sponsored Cyber Actors Use Maui Ransomware to Target the Healthcare and Public Health Sector (Advisory AA22-187A).
https://www.cisa.gov/news-events/cybersecurity-advisories/aa22-187a

[13] FBI and CISA. (2020, July 22). People's Republic of China (PRC) Targeting of COVID-19 Research Organizations.
https://www.cisa.gov/news-events/alerts/2020/07/22/peoples-republic-china-prc-targeting-covid-19-research-organizations

[14] National Cyber Security Centre (UK). (2020, July 16). Advisory: APT29 Targets COVID-19 Vaccine Development.
https://www.ncsc.gov.uk/news/advisory-apt29-targets-covid-19-vaccine-development

[15] U.S. Department of Health and Human Services, Health Sector Cybersecurity Coordination Center (HC3). (2024). Healthcare Sector Ransomware Trend Analysis [HC3 Threat Brief].

[16] Health Information Sharing and Analysis Center (Health-ISAC). (2025, February). 2025 Health Sector Cyber Threat Landscape Report.
https://health-isac.org/health-isac-2025-health-sector-cyber-threat-landscape/

[17] Kang C. Google Says Chinese Cybercrime Group Used Its A.I. in Scams. NY Times. June 12, 2026. https://www.nytimes.com/2026/06/12/technology/google-lawsuit-china-ai-scams.html

[18] U.S. Department of Health and Human Services, Office for Civil Rights. (2025, July 31). Change Healthcare Updated Breach Notification.
https://www.hhs.gov/hipaa/for-professionals/special-topics/change-healthcare-cybersecurity-incident-frequently-asked-questions/index.html

[19] HIPAA Journal. (2025). Healthcare Data Breach Statistics.
https://www.hipaajournal.com/healthcare-data-breach-statistics/

[20] Renfrow, J. (2019, April 10). 90% of claim denials are avoidable with help of tech tools.

Fierce Healthcare.
https://www.fiercehealthcare.com/payer/waystar-90-claim-denials-are-avoidable-help-technology

[21] Lauren Bilbo, Elman Amador Medina, Dashiell Miner, David Scheinker, Stefanos Zenios, Kevin Schulman. 2026. Developing a Computable Payer-Provider Contract. *Health Management, Policy and Innovation* (www.HMPI.org), Volume 11, Issue 1.

[22] Invene. (2025, August). 270/271, 835, 837: Decoding the 9 Key Healthcare EDI Transactions.
https://www.invene.com/blog/demystifying-healthcare-edi-the-9-critical-transactions-explained

[23] Straits Research. (2024). Healthcare Fraud Analytics Market Size, Global Trends, Share, Forecast to 2033. https://straitsresearch.com/report/healthcare-fraud-analytics-market

[24] The Nilson Report. (2022, December). Card Fraud Losses Reach $32.40 Billion (Issue 1230). Referenced in: Payments Dive. (2023, January 5). Card industry's fraud-fighting efforts pay off.
https://www.paymentsdive.com/news/card-industry-fraud-fighting-efforts-pay-off-nilson-report-credit-debit/639675/

[25] Flinders, M., Smalley, I., & Schneider, J. (n.d.). AI fraud detection in banking. IBM Think.
https://www.ibm.com/think/topics/ai-fraud-detection-in-banking

[26] Aloa. (2025, August 28). How AI is changing fraud detection in insurance.
https://aloa.co/blog/ai-is-changing-fraud-insurance

[27] Deloitte. (2025, April 24). Using AI to fight insurance fraud. Deloitte Insights.
https://www.deloitte.com/us/en/insights/industry/financial-services/financial-services-industry-predictions/2025/ai-to-fight-insurance-fraud.html

[28] Tseng, P., Kaplan, R. S., Richman, B. D., Shah, M. A., & Schulman, K. A. (2018). Administrative Costs Associated with Physician Billing and Insurance-Related Activities at an

Academic Health Care System. JAMA, 319(7), 691–697.

[29] Gillai, B., & Mendelson, H. (2020, November). Creating Value with Blockchain: A Value Chain Management Perspective. Stanford Graduate School of Business. https://www.gsb.stanford.edu/faculty-research/publications/creating-value-blockchain-value-chain-management-perspective

[30] Dar, N., et al. (2021, April 14). Technology-Led Shifts and Opportunities in Card-Based Payments. McKinsey & Company. https://www.mckinsey.com/industries/financial-services/our-insights/banking-matters/technology-led-shifts-and-opportunities-in-card-based-payments

[31] Bowman, C. (2022, February 9). What Are EMV Chips and Do They Make Credit Cards More Secure? CNET. https://www.cnet.com/personal-finance/credit-cards/what-are-credit-card-chips-and-are-they-more-secure/

[32] Visa. (2018, May). Counterfeit fraud at chip-enabled merchants down 76 percent. Reported in: American Bankers Association. (2018, May). Visa: Counterfeit Fraud Drops 76 Percent at Chip-Enabled Merchants. ABA Banking Journal. https://bankingjournal.aba.com/2018/05/visa-counterfeit-fraud-drops-76-percent-at-chip-enabled-merchants/

[33] Hayashi F. New Data on Card-Present and Card-Not-Present Fraud Rates in the United States. Federal Reserve Bank of Kansas City. February 25, 2026. https://www.kansascityfed.org/research/payments-system-research-briefings/new-data-on-card-present-and-card-not-present-fraud-rates-in-the-united-states/

[34] Shrank, W. H., et al. (2019). Waste in the US Health Care System: Estimated Costs and Potential for Savings. JAMA, 322(15), 1501–1509.

[35] Pifer, R. (2019, June 14). House Votes to Overturn Decades-Old Ban on National Patient Identifier. Healthcare Dive. https://www.healthcaredive.com/news/house-votes-to-overturn-decades-old-ban-on-national-pati

ent-identifier/556859/

[36] Patel, V., Hughes, P., Savage, L., & Barker, W. (2015, June). Individuals' Perceptions of the Privacy and Security of Medical Records (ONC Data Brief No. 27). Office of the National Coordinator for Health Information Technology. https://www.healthit.gov/sites/default/files/briefs/oncdatabrief27june2015privacyandsecurity.pdf

[37] McClain, C. (2026, January 8). What We Know About Internet Use, Smartphone Ownership and Digital Divides in the U.S. Pew Research Center. https://www.pewresearch.org/short-reads/2026/01/08/internet-use-smartphone-ownership-digital-divides-in-u-s/

[38] U.S. Department of Health and Human Services, Office of Inspector General. List of Excluded Individuals/Entities (LEIE) [updated monthly]. https://oig.hhs.gov/exclusions/

[39] U.S. Government Accountability Office. (2016, November). Medicare: Initial Results of Revised Process to Screen Providers and Suppliers, and Need for Objectives and Performance Measures (GAO-17-42). https://www.gao.gov/products/gao-17-42

[40] U.S. Government Accountability Office. (2016, April). Medicare: Claim Review Programs Could Be Improved with Additional Prepayment Reviews and Better Data (GAO-16-394). https://www.gao.gov/products/gao-16-394

[41] Amponsah, A. A., Adekoya, A. F., & Weyori, B. A. (2022). A novel fraud detection and prevention method for healthcare claim processing using machine learning and blockchain technology. Healthcare Analytics, 2, Article 100034. https://www.sciencedirect.com/science/article/pii/S2772662222000534

[42] U.S. Department of Health and Human Services, Office of Inspector General. (2016, September). The Fraud Prevention System's Adjustment Factors (Report No. A-01-15-00510). https://oig.hhs.gov/oas/reports/region1/11500510.pdf

[43] Chan, T. (2024, January 16). Data visualization and AI for healthcare fraud detection. Cambridge Intelligence Blog. https://cambridge-intelligence.com/medical-fraud-detection/

[44] Centers for Medicare & Medicaid Services. (2016). The Health Care Fraud and Abuse Control Program Protects Consumers and Taxpayers by Combating Health Care Fraud [Fact sheet]. https://www.cms.gov/newsroom/fact-sheets/health-care-fraud-and-abuse-control-program-protects-consumers-and-taxpayers-combating-health-care-0

[45] U.S. Department of Health and Human Services, Office of Inspector General and Department of Justice. (2024, December 6). Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2023. https://oig.hhs.gov/reports/all/2024/health-care-fraud-and-abuse-control-program-report-fiscal-year-2023/

[46] California Senate Bill 1120 (Becker). (2024). Health Care Coverage: Utilization Review. Signed September 28, 2024; effective January 1, 2025. https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=202320240SB1120

[47] American Medical Association. (2026). 2025 AMA Prior Authorization Physician Survey. https://www.ama-assn.org/system/files/prior-authorization-survey.pdf

[48] Texas Department of Insurance. (2025). Preauthorization Exemptions (House Bill 3459, 87th Legislature, 2021; updated by House Bill 3812, 89th Legislature, 2025). https://www.tdi.texas.gov/health/hb3459.html

[49] Lunghi, D., Simitsis, A., Caelen, O., & Bontempi, G. (2023). Adversarial Learning in Real-World Fraud Detection: Challenges and Perspectives. Second ACM Data Economy Workshop (DEC ’23). https://doi.org/10.1145/3600046.3600051

[50] Pickern JS. Prior authorizations and the adverse impact on continuity of care. Am J Manag Care. 2025 Apr;31(4):163-165.

HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION

BAHM BUSINESS SCHOOL ALLIANCE FOR HEALTH MANAGEMENT

[1] Deepfakes are synthetic audio, video, or images created with artificial intelligence to convincingly impersonate real people.

[2] Some of these security approaches may be challenging for elderly patients or patients with dementia.

Ron Varghese, Robbins Institute for Health Policy & Leadership, Baylor University

Abstract

What is the message? U.S. hospitals face fragmented operational pressures across capacity, staffing, resource coordination, and equitable access. This paper proposes the CARE Framework, Capacity Optimization, Adequate Workforce Alignment, Resource Coordination Across Hospitals, and Equitable Access to Care, as a practical structure for using AI Command Centers to improve hospital operations while accounting for human factors, source-data reliability, and financial sustainability.

What is the evidence? Evidence is drawn from U.S. and international hospital examples showing improvements in bed assignment, discharge efficiency, staffing prediction, transfer coordination, ICU strain, and financial performance after command-center implementation. The paper also uses healthcare management literature to distinguish AICCs from traditional dashboards and to explain implementation limits involving adaptive staffing, source-data reliability, and return-on-investment attribution.

Keywords: Artificial Intelligence, Hospital Operations

Timeline: Submitted May 11, 2026; accepted after review July 12, 2026.

Cite as: Ron Varghese. 2026. The CARE Framework: A Pathway for AI Command Centers to Streamline Hospital Operations in the United States. *Health Management, Policy and Innovation* (www.HMPI.org). Volume 11, Issue 2.

Current State of Hospital Operational Pressures

U.S. hospitals face increasing pressure to deliver high-quality care amidst a variety of complicated operational challenges. Since the COVID-19 pandemic, emergency department overcrowding has become more pronounced, resulting in longer wait times for patients before they receive treatment or are admitted to inpatient units [1]. This crowding contributes to the unprecedented strain on the healthcare workforce. Over 130,000 registered nurses have left the field since 2022, and one million nurses are projected to leave the workforce by 2030 [2]. 86,000 physicians are predicted to leave the field within the next decade as well [3]. With fewer staff available to manage patient care, hospitals experience delayed discharges, while the added workload on remaining personnel yields high burnout and fatigue.

In addition to staffing shortages, hospital supply chains have struggled to keep up with growing workload and patient surges. The shortage of ventilators and personal protective equipment during the COVID-19 peak, as well as other disease-outbreak surges with high patient overflows, highlight the challenges of coordinating resources among facilities, especially in rural or underserved areas [4]. These operational pressures exacerbate existing, persistent inequities in chronic disease outcomes and access to timely specialty care, particularly for Black, Hispanic, rural, and low-income populations, as fragmented systems limit visibility into patient needs and create additional access barriers [5]. Collectively, these issues lead to the overarching problem of fragmented hospital operations. If patient flow, staffing, and equipment are managed separately, system-wide inefficiencies compound, leading to poorer overall health outcomes for patients.

Current Approaches to Managing Hospital Operations

Against these operational pressures, hospitals traditionally have tried to improve efficiency through digital dashboards, departmental analytics, or lean management systems [6]. However, these tools often leave patient flow, staffing, and supply chain decisions disconnected. Administrators may still rely on spreadsheets, calls, or incomplete data, and dashboards that track beds, staffing, and equipment may not provide a coherent real-time view of hospital performance [7]. COVID-19 exposed these limits when hospitals struggled to predict surges and allocate resources quickly [8]. Because traditional approaches are often reactive to problems, hospitals need systems that anticipate and coordinate patient flow, workforce, and resources in real time. This drives the need for a comprehensive, data-driven system: the AI Command Center (AICC).

The AICC refers to a centralized operations model that combines real-time data, predictive analytics, escalation pathways, and assigned follow-up. Its inputs, outputs, and decision points differ by setting: a single hospital may use it for patient flow, a health system may use it for OR, clinic, and service coordination, and a multi-hospital network may use it for patient and resource routing. This paper explains how AICCs can be operationalized in hospital systems using the CARE Framework: Capacity Optimization, Adequate Workforce Alignment, Resource Coordination Across Hospitals, and Equitable Access to Care.

Using national and global examples, the paper describes each CARE pillar and outlines management actions that can be implemented through an AICC or pillar-specific AI-enabled operational tools: it explains how AICCs differ from ordinary reporting dashboards by linking prediction to operational decisions; describes each CARE pillar; and discusses implementation requirements, including human factors, data quality, financial risk, and the need to interpret reported savings cautiously.

The Usage of AI Command Centers in Hospital Operations

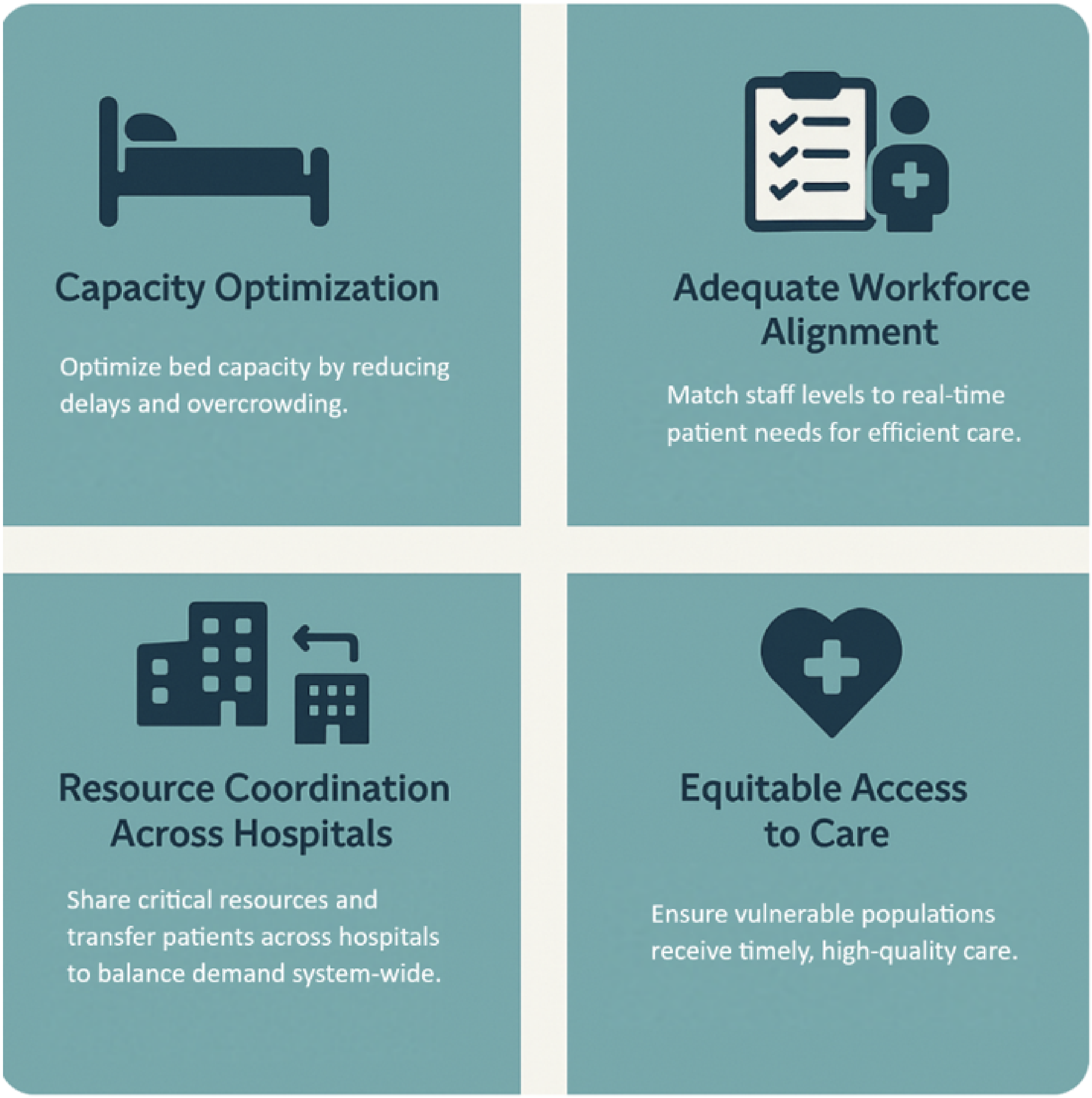
AI Command Centers (AICCs) offer hospitals a new way to help address operational challenges. Inspired by NASA's mission control center, which covers the operational logistics of the International Space Station, AICCs cover the operational logistics of hospitals by combining real-time and predictive analytics on patient flow, staff deployment, and resource availability [9]. In practice, an AICC functions as a centralized operations hub that integrates electronic health record, bed, staffing, and supply chain data into shared operational views. However, the AICC model is intended to extend beyond visualization. Rather than only displaying current conditions, an AICC pairs real-time visibility with predictive analytics, escalation pathways, and assigned operational follow-up. Its value comes from turning shared data into coordinated action, not simply monitoring hospital performance [7]. The AI contribution may be advanced machine learning, simpler predictive rules, or older coordination technology connected to faster operational decisions; therefore, the paper emphasizes what centers did and how they organized decisions, rather than treating reported savings as rigorous causal proof.

Despite these benefits, U.S. adoption remains limited. Of 6,093 U.S. hospitals, only 68 have a fully operational AICC [12]. About 1% of U.S. hospitals and roughly 300 hospitals globally have implemented AICCs [10]. Still, many hospitals use predictive AI tools that can serve as AICC building blocks. Seventy-one percent of U.S. hospitals use some predictive AI, and 86% of system-affiliated hospitals integrate AI tools with electronic health records [13]. Most applications remain confined to single functions such as scheduling, imaging, or billing, leaving AI in silos rather than providing the system-wide visibility and coordination AICCs can offer [14]. The CARE Framework demonstrates how AICCs can integrate predictive analytics across capacity, staffing, resources, and access.

The CARE Framework

This paper presents the CARE Framework, which includes Capacity Optimization, Adequate Workforce Alignment, Resource Coordination Across Hospitals, and Equitable Access to Care (Figure 1). Each pillar represents an operational decision area. The acronym CARE was selected because the pillars support patient care through more responsive and sustainable systems. Capacity Optimization focuses on forecasting bed capacity, shortening discharge times, and reducing emergency department bottlenecks. Adequate Workforce Alignment matches staffing resources to predicted patient acuity to avoid underuse or overuse. Resource Coordination Across Hospitals tracks, predicts, and redistributes critical resources during surges to support timely transfers and system-wide balance. Equitable Access to Care uses AICCs to identify gaps in service capacity and improve timely access across hospitals. The first two pillars are mainly intra-hospital or intra-system opportunities; resource coordination is explicitly inter-hospital; and equity can operate at both levels. Together, the pillars show how AICCs can convert fragmented operations into an integrated, predictive care model.

Figure 1: Key Components of the CARE Framework



Capacity Optimization

Capacity optimization is a foundation of effective hospital operations, ensuring that patients move smoothly from admission to discharge without unnecessary delays. In many U.S. hospitals, overcrowded emergency departments (ED) and prolonged boarding times disrupt the entire care continuum, increasing patient risk and reducing throughput efficiency. AICCs address this issue by using real-time dashboards and predictive modeling to track bed availability, anticipate discharges, and manage system-wide patient transfers. Here, inputs include ED boarding, staffed beds, surgery schedules, discharge readiness, and transport status; outputs include bed priorities, discharge forecasts, and bottleneck alerts.

Hospitals in Canada and the United Kingdom demonstrate how AICCs can be helpful in capacity management. At Canada's Humber River Hospital, command-center analytics reportedly supported discharge timing and bed turnover, creating 35 additional inpatient beds [15]. Humber River Hospital reduced acute conservable bed days by 52% and ED patient wait times by 23%, despite an 8% increase in ED visits [16]. Similarly, the UK's Bradford Teaching Hospitals used real-time data across an 800-bed network to help staff anticipate bottlenecks and place patients in the wards best suited to their care [17, 24]. Bradford reduced bed misallocations by 90% and increased early discharges from 35% to 54% [17]. These case-study outcomes are useful as implementation examples, but they should not be read as causal estimates of AICC impact.

Within the United States, AICCs have shown reported operational gains at Johns Hopkins and Tampa General hospitals. At Johns Hopkins, the AICC brought together access-line staff, bed placement nurses, transport staff, and admitting staff to coordinate incoming transfers, ED admissions, operating room status, staffed beds, and beds waiting to be cleaned [18]. They assigned ED patients to beds 38% faster (3.5 hours faster) after the admission decision [18]. The reported mechanism was faster bed assignment, transfer coordination, and earlier response to inpatient bottlenecks. Occupancy rates rose from 85% to 92% after the implementation of the AICC, with $16 million in incremental savings generated for the hospital. Another example includes Tampa General Hospital, where 30 new beds freed by eliminating 20,000 excess days in patient stay helped generate $40 million in annual savings [19]. These cases show how AICCs can tie capacity data to decisions, while the savings figures should be interpreted as reported operating outcomes rather than isolated causal effects.

Management Actions for Implementing Capacity Optimization:

For healthcare administrators and executives, capacity optimization should be managed using three key indicators: time-to-bed, ED boarding hours, and discharges by noon [20]. Leaders should review these metrics in a daily capacity huddle and assign clear ownership for actions such as accelerating discharges, opening surge beds, and prioritizing inpatient placement from the ED [21]. When these metrics are monitored consistently, hospitals can intervene earlier to reduce boarding and avoid system-wide congestion.

This pillar can be supported through a full AICC or through specific platforms such as TeleTracking or LeanTaaS, which apply AI to bed placement, discharge prediction, ED-to-inpatient throughput, and real-time capacity forecasting [22]. This pillar supports smoother patient flow, reduces bottlenecks, and promotes safer care delivery through more balanced use of hospital capacity.

Adequate Workforce Alignment

Adequate workforce alignment means the ability of hospitals to consistently match available personnel with the intensity and timing of patient care demands. In other words, it means having the right people in the right place at the right time to meet patient needs. This, in practice, is often harder to implement than it sounds. Traditional staffing methods struggle to adapt in environments where patient acuity can change within hours. AICCs address this challenge by using predictive analytics to anticipate census changes and coordinate workforce distribution across departments or entire hospital systems [9].

Hospitals in Australia and Britain illustrate the use of AICCs in predicting and adjusting healthcare personnel. Australia's Alfred Health has enhanced the accuracy of scheduling for nurses and minimized unexpected overtime through AICCs, which improved staff satisfaction [23]. Similarly, Britain's Bradford Royal Infirmary implemented AICCs to monitor patient flow and staffing pressure across units [24]. Although quantifiable outcomes are unavailable to the public for these hospitals, the following cases illustrate how AICCs can use available data to optimize the number of staff according to the current circumstances in a hospital. In this pillar, the main decision-point is whether leaders should adjust schedules, float-pool coverage, or redeployment before a unit is understaffed. These hospitals show how AICCs can contribute to properly utilizing the workforce and aiding staff well-being.

Duke University Hospital and the Cleveland Clinic are U.S. hospitals that have shown measurable results with workforce management. Duke Health reported a 95% accuracy rate in predicting staffing needs through using AICC-related tools, leading to $40 million in annual labor savings and a significant reduction in nurse burnout [25]. Similarly, Cleveland Clinic reported that AI-assisted staffing and scheduling reduced workforce response times and allowed supervisors to identify stress points before patient care is affected [26]. These instances show how AICCs help executives make evidence-based staffing interventions that influence staff well-being, patient care, and financial performance. This unique benefit of AICCs shows that staffing precision should be of importance for organizational sustainability.

Management Actions for Implementing Adequate Workforce Alignment

Healthcare leaders must ensure that their workforce is used effectively, starting with real-time monitoring of staff use and anticipation of overall patient surges and droughts via AICCs [10]. The administrative staff should conduct predictive staff meetings every morning to estimate changes in patient volumes and redeploy employees before they run out of capacity [27]. The AICC staffing data allows HR departments and nursing leaders to design float pools and redeploy nurses more efficiently, reducing premium labor costs and improving workforce utilization [28].

It is important to note that leaders should be aware of the human and clinical challenges involved in adaptive staffing. Redeployment is not interchangeable labor, especially in specialized units that require specific competencies, credentials, and orientation. For example, a nurse trained for one ICU environment may not be immediately prepared to work safely in another specialty ICU without orientation, cross-training, or supervisory support. Staff may also resist being flexed away from familiar units on short notice. Therefore, AICC-based staffing should be paired with competency mapping, cross-training, float-pool planning, and frontline feedback to ensure redeployment is safe and practical [29].

In practice, these workforce-alignment actions can be supported through an AICC or through workforce-focused platforms such as UKG and Qventus, which use AI to forecast staffing demand, optimize schedules, and enable real-time redeployment decisions during surges [30]. By incorporating these processes into everyday operations, leaders can maintain workforce preparedness and ensure staff satisfaction, where each staff member can provide the best care possible for patients.

Resource Coordination Across Hospitals

Resource coordination across hospitals is particularly helpful during periods of extreme strain, such as seasonal influenza surges or mass casualty events. During the COVID-19 pandemic, hospitals globally experienced severe operational strains due to shortages of ventilators, ICU beds, personal protective equipment, and infusion pumps [31]. Facilities operated in silos, where a given hospital could run out of required supplies, but nearby hospitals had idle supplies [32]. In this pillar, AICC inputs come from multiple facilities, and outputs include regional strain alerts, transfer recommendations, and resource-allocation priorities. AICCs can predict PPE shortages, manage ICU utilization, and facilitate quicker redistribution of essential resources or patient transfers across multiple hospitals. AICCs, especially those made by General Electric (GE), are data agnostic, meaning that hospitals with varying electronic health records (EHRs) can still feed their data to AICCs for combined analytics [10]. This interoperability enables hospitals with disparate data systems to cooperate efficiently in times of emergencies.

The Saudi Arabia Ministry of Health is a global example of inter-hospital AICCs effectively managing resource coordination in times of crises. AICCs were implemented in all 308 hospitals around the country [10]. The live visibility of ICU beds, available ventilators, and surgical resources were provided to all hospitals, allowing administrators to transfer patients or resources using inter-regional hospital networks. The decision-points were based on where a patient, ventilator, or surgical resource should move within the network. Compared with pre-COVID-19 baseline performance, the network reported a 27% decrease in hospital length of stay and a 60% reduction in surgical wait times, even during peak COVID-19 operational strain [33]. This case illustrates that AICCs may serve as the connective structure between hospitals where supply and demand are brought in line.

There are also examples of AICCs enabling effective resource coordination among U.S. hospitals. The AICC at Oregon Health and Science University received data from all 62 hospitals in Oregon to provide real-time information on ICU occupancy, ventilator utilization, N95 masks, and PPE distribution. Here, the input was statewide data rather than one hospital's bed board, which helped leaders see where strain was emerging. The system assisted in filling patient populations and alleviated stress on the ICUs of urban and rural facilities in Oregon during the COVID-19 pandemic [34]. On average, the number of days ICUs were stressed (i.e., 80% total ICU capacity or higher) across the country, was approximately 500 days during COVID, while the number of days of stressed ICUs in Oregon was 16 days [35]. Another example of resource coordination is Advent Health in Central Florida, where one AICC connected 11 hospitals and 18 emergency departments. According to this system, hospital-to-hospital lateral transfers increased by over 600% during the pandemic, allowing 2,500 patients to be reallocated to open locations with open ventilators and ICU units [36]. These examples show how AICCs can facilitate hospitals' ability to share strain during surges while sustaining care delivery.

Management Actions for Implementing Resource Coordination Across Hospitals

For healthcare administrators and executives seeking to manage resources better, data transparency across facilities is integral to gaining maximum benefit. Executives must think about and establish interoperability agreements to enable the exchange of information across hospitals, irrespective of the EHR platform, to identify real-time equipment capacity and availability [37]. The information produced by AICCs will help leadership teams model surge scenarios and know how to reallocate resources during future surges.

Platforms such as Palantir Technologies and WellSky can also support this pillar by strengthening cross-hospital data integration and providing analytics for regional patient transfer, bed capacity visibility, and resource distribution across facilities [19, 38]. The formation of regional logistics teams that move patients, supplies, medications, and staff across hospitals, can further improve resilience as sites approach capacity, allowing for better overall patient care [39]. That helps the AICC output lead to a transfer, resource move, or escalation rather than the creation of another report.

Equitable Access to Care

Increasing equitable access to care is essential to helping hospital systems fulfill their mission of serving patients with fairness and efficiency. Inequities in care tend to be concealed in disjointed hospital operations. These gaps can be seen with AICCs, where data can be integrated among different departments and locations. When wait times, discharges, and transfer approvals are stratified by geography, insurance, language, race, or service line, administrators can identify who is underserved and link that information to capacity decisions. AICCs helps address inequities in care by increasing access to care across diverse communities.

Examples from England and Australia suggest that AICCs can help decrease geographic and systemic inequality. The South Tees Hospitals in England used its AICC to help rural patients gain greater access to specialized services, performing 22% more surgical operations and providing access to 15% more ICU beds during periods of high strain [40]. In Australia, the New South Wales Hospital AICC connects metropolitan and rural hospitals to monitor care delays, reducing rural patient transfer times by 18% and increasing access to care for these patients [23]. These outcomes show how AICCs can bridge service gaps between urban and rural hospitals, ensuring that patients receive timely and equitable access to critical care.

AICCs in U.S. hospitals have also helped increase equitable access to care. Using predictive analytics, Deaconess Health System was able to add capacity without new buildings to serve 2,000 more patients annually, which helped treat underserved populations [41]. Similarly, implementing regional dashboards within Providence Swedish Health System helped in organizing staffing and capacity in many states, providing treatment to more than 5,000 rural patients in the recent year [42]. These U.S. hospitals show how data-driven coordination can expand hospital capacity and create a more inclusive healthcare network. The access benefit depends on whether leaders use the information to change routing, scheduling, transfer approval, or outreach decisions.

Management Actions for Implementing Equitable Access to Care

For healthcare leaders, advancing equity requires administrators to consider demographic and geographic data such as wait times, transfer rates, and patient outcomes within AICCs to identify disparities and emerging trends [43]. Equity analytics platforms such as Arcadia and Veradigm can support this work by using AI-enabled predictive analytics and stratified dashboards to surface disparities in wait times, transfer patterns, and outcomes by race, language, insurance status, and geography, making inequities visible and trackable over time [44].

Leadership teams can establish unified transfer and diversion protocols that automatically direct patients from lower resourced hospitals to those with available capacity, ensuring system-wide balance [45]. Executives should conduct regular reviews of equity scorecards and communicate progress transparently to both staff and the community to promote trust and shared responsibility in achieving equitable care.

Equity-focused expansion also requires financial planning; if an AICC increases transfers, admissions, or specialty access for underinsured or uninsured patients, the hospital may experience higher uncompensated care and not fully realize the financial benefits of the AICC. This does not weaken the ethical case for equitable access, but it does mean that leaders should monitor payer mix, uncompensated care, charity care, avoided transfers, and downstream reimbursement when evaluating AICC return on investment [46]. This is especially important because many hospitals already operate with limited margins and face pressure from labor costs, supply costs, Medicare and Medicaid underpayment, and rising uncompensated care.

Implementing and Operationalizing an AI Command Center

To utilize the CARE framework effectively, hospitals must first implement an AI Command Center (AICC). Generally, large academic or multi-hospital systems are best positioned to lead the adoption of AICCs due to the scale of data and coordination involved, as well as the significant capital investment required [9]. Medium to smaller-sized hospitals can participate by partnering with larger health system AICCs and feed their data virtually. It is important to note that an AICC is only as reliable as the source data entered by each facility. Bed status, staffing availability, discharge readiness, supply counts, and patient acuity can be delayed, incomplete, or shaped by local workflow pressures. Therefore, AICC implementation should include data governance, standardized definitions, audit checks, and leadership accountability before outputs are used for staffing, transfer, or resource decisions. This approach has been demonstrated to be effective through Oregon Health and Science University's AICC and the state's surrounding hospitals during the COVID-19 pandemic [35]. This type of collaborative approach ensures hospitals of all sizes benefit from AICC insights and move toward a connected data-driven healthcare network.

Although there are AI-enabled platforms discussed in this paper that support individual CARE pillars (e.g., TeleTracking, Qventus, and Palantir), GE Healthcare's Command Center model is positioned as a comprehensive solution that integrates these pillar-specific capabilities into a single, unified AI Command Center to operationalize the CARE framework across hospital operations [47]. This comparison is presented to distinguish between pillar-specific implementations and enterprise-wide command center models rather than to endorse a specific vendor.

Starting an AICC from the ground up will require a multidisciplinary effort involving administrators, IT specialists, executives, physicians, and nurses [9]. Early engagement of diverse staff ensures that the AICC reflects the hospital's unique operational challenges and addresses staff apprehension about AI. When clinicians and frontline teams participate in shaping the system, they develop a sense of ownership that promotes trust and confidence in technology. Likewise, involving executives and department leaders in this process creates an environment of collaboration where both clinical and administrative stakeholders are invested in the AICC's success.

Historically, the operational capabilities of AICCs are implemented in phases, with, for example, an initial focus on capacity management, then on staffing alignment, and subsequently, on other areas of operations covered under the CARE framework [48]. In the process of implementing AICCs, administrators and executives can learn from early adopters. For example, Johns Hopkins Hospital has hosted hundreds of peer visits and helped over 20 other institutions, such as Duke Health, New Haven Health, and HCA Healthcare, launch their own AICCs based on their experiences [18]. Such collaboration will accelerate adoption and help hospitals to turn data insight into system-wide improvements in operational performance.

Administrators must also plan for a significant investment; building a state-of-the-art AICC could cost roughly $10,000,000 or more [49]. Although upfront costs are significant, hospitals have reported substantial returns: Virginian Franciscan Health achieved a 12:1 ROI, Johns Hopkins generated $16 million in savings, and Tampa General and Duke Health each reported $40 million in savings [50-52, 25]. While these savings are promising, they should be interpreted cautiously. In many hospitals, command center implementation occurs alongside broader performance improvement work, process redesign, staffing changes, and leadership attention to throughput. Therefore, the impact of AICCs could be more safely interpreted as an enabling infrastructure that helps hospitals organize, accelerate, and sustain operational improvement efforts. In sum, launching an AICC is a significant investment, but the return may justify the cost when the center is paired with process redesign, reliable data, and clear decision ownership. Reported financial gains should still be attributed cautiously.

Conclusion

AI Command Centers (AICCs) can reshape how hospitals operate by creating interconnected systems that use real-time data and predictive analytics to improve patient flow, resource utilization, and overall care quality. Guided by the CARE framework, they bring structure and foresight to hospital management. When hospitals implement AICCs with reliable data, clear workflows, and human oversight, these centers can support more efficient and equitable care in an increasingly complex healthcare system.

References

1. Kurhayati K, Emaliyawati E, Trisyani Y. An updated scoping review of factors associated with length of stay in emergency department. Journal of Multidisciplinary Healthcare. 2025;18:3191-3203. https://doi.org/10.2147/JMDH.S525451
2. American Nurses Association. Nursing shortage fact sheet. 2022.
3. Spoehr C. New AAMC report shows continuing projected physician shortage. Association of American Medical Colleges. 2024. https://www.aamc.org/news/press-releases/new-aamc-report-shows-continuing-projected-physician-shortage
4. Sandhu P, Shah AB, Ahmad FB, et al. Emergency department and intensive care unit overcrowding and ventilator shortages in US hospitals during the COVID-19 pandemic, 2020-2021. Public Health Reports. 2022;137(4):796-802. https://doi.org/10.1177/00333549221091781
5. Ndugga N, Hill L, Artiga S. Key data on health and health care by race and ethnicity. KFF. 2024.
6. Lawal AK, Rotter T, Kinsman L, Sari N, Harrison L, Jeffery C, Kutz M, Khan MF, Flynn R. Lean management in health care: definition, concepts, methodology and effects reported: systematic review protocol. Systematic Reviews. 2014;3:103.
7. Rabiei R, Almasi S. Requirements and challenges of hospital dashboards: a systematic literature review. BMC Medical Informatics and Decision Making. 2022;22(1):287. https://doi.org/10.1186/s12911-022-02037-8
8. Filip R, Gheorghita Puscaselu R, Anchidin-Norocel L, Dimian M, Savage WK. Global

challenges to public health care systems during the COVID-19 pandemic: a review of pandemic measures and problems. Journal of Personalized Medicine. 2022;12(8):1295.

9. Small T, Baker L, Egizi A, Lesser P, Sindha I. Assessing the value of a command center model in large health care delivery systems. Deloitte Center for Government Insights. 2024.
10. GE HealthCare. AI command center outcomes in U.S. hospitals. GE HealthCare Reports. 2025.
11. OR Today Magazine. Tampa General Hospital and GE Healthcare's CareComm saves $40 million. OR Today. 2020.
12. American Hospital Association. Fast facts on U.S. hospitals, 2025. American Hospital Association. 2025.
13. Olsen E. More hospitals using predictive AI, but disparities persist: ASTP. Healthcare Dive. 2025.
14. Baten R. How are US hospitals adopting artificial intelligence? Early evidence from 2022. Health Affairs Scholar. 2024;2(10). https://doi.org/10.1093/haschl/qxae123
15. Humber River Health. Humber's Quality Command Centre helps deliver better, faster, and safer patient care: year one in review. GE HealthCare. 2019.
16. Analytics Informs. Command Center Analytics Revolutionize Healthcare. Analytics Magazine. 2018.
17. HTN Health. Case study: Bradford Teaching Hospitals NHS FT command centre. HTN Health Tech News. 2024.
18. Keating K. Capacity Command Center celebrates 5 years of improving patient safety, access. Johns Hopkins Medicine. 2021. https://www.hopkinsmedicine.org/news/articles/2021/03/capacity-command-center-celebrates-5-years-of-improving-patient-safety-access
19. Tampa General Hospital. Tampa General Hospital and GE Healthcare's CareComm saves $40 million, cuts 20,000 excess days and reduces length of stay. 2024.
20. Bailey R, Segon A, Garcia S, Kottewar S, Lu T, Tuazon N, Sanchez L, Gelfond JA, Bowling G. Increasing and sustaining discharges by noon: a multi-year process improvement project. BMC Health Services Research. 2024;24(1):478. https://doi.org/10.1186/s12913-024-10960-x
21. Al Harbi S, Aljohani B, Elmasry L, Baldovino FL, Raviz KB, Altowairqi L, Alshlowi S.

Streamlining patient flow and enhancing operational efficiency through case management implementation. BMJ Open Quality. 2024;13(1):e002484. https://doi.org/10.1136/bmjoq-2023-002484

22. Anumolu S, Samant M. Optimizing hospital bed capacity: how lean-AI integration prevents capacity constraints from multiplying demand. Intelligent Hospital. 2026;100048. https://doi.org/10.1016/j.inhs.2026.100048
23. Australian Digital Health Agency. DHA annual report. Australian Digital Health Agency. 2024.
24. NHS Confederation. Managing patient care through an AI-powered command centre. NHS Confederation. 2024. https://www.nhsconfed.org/case-studies/managing-patient-care-through-ai-powered-command-centre
25. Sohail S. Duke Health to deploy GE Healthcare's Hospital Pulse Tile for Command Center. Healthcare IT News. 2024.
26. Cleveland Clinic. How AI assists with staffing, scheduling and once-tedious tasks. Cleveland Clinic Consult QD. 2025.
27. Singh K, Nashwan AJ. Innovative forecasting models for nurse demand in modern healthcare systems. World Journal of Methodology. 2025;15(3):99162. https://doi.org/10.5662/wjm.v15.i3.99162
28. Knight S. How to master float pool management in hospitals. Shiftmed Blog. 2025. https://www.shiftmed.com/insights/knowledge-center/how-to-master-float-pool-management-in-hospitals/
29. American Nurses Association. Patient classification and acuity systems. 2017. https://www.nursingworld.org/globalassets/practiceandpolicy/work-environment/nurse-staffing/website-staffing-and-acuity-systems-pdf-final_2017.pdf
30. Hijazi S. AI-enabled hospital workflow automation and efficiency tools: the new operating system for modern health systems. Research2Guidance. 2025. https://research2guidance.com/ai-enabled-hospital-workflow-automation-efficiency-tools-the-new-operating-system-for-modern-health-systems/
31. Arabi YM, Myatra SN, Lobo SM. Surging ICU during COVID-19 pandemic: an overview. Current Opinion in Critical Care. 2022;28(6):638-644. https://doi.org/10.1097/MCC.0000000000001001

32. Peiris D, Feyer AM, Barnard J, Billot L, Bouckley T, Campain A, Cordery D, et al. Overcoming silos in health care systems through meso-level organisations: a case study of health reforms in New South Wales, Australia. The Lancet Regional Health – Western Pacific. 2024;44:101013. https://doi.org/10.1016/j.lanwpc.2024.101013
33. Saudi Ministry of Health. Ministry of Health reports advancements in Ada'a health indicators. Ministry of Health Saudi Arabia. 2024.
34. Wentz-Graff K. OHSU Mission Control offers modern, high-tech solution to historic challenge. OHSU News. 2019.
35. Radley D, Baumgartner J, Collins S. 2022 scorecard on state health system performance. The Commonwealth Fund. 2022.
36. Advent Health. AdventHealth Mission Control drives improved emergency department outcomes. AdventHealth. 2022.
37. Pylypchuk Y, Everson J. Interoperability and methods of exchange among hospitals in 2021. ASTP Health IT Data Brief. 2023.
38. WellSky. Care coordination software: WellSky. 2026. https://wellsky.com/care-coordination-software/
39. Zamani H, Parvaresh F, Isfahani MN. Optimizing patient flow logistics: strategic challenges, tactical solutions, and future directions. BMC Health Services Research. 2025;25:1382. https://doi.org/10.1186/s12913-025-13516-9
40. NHS England. NHS artificial intelligence (AI) giving patients better care and support. NHS England. 2024. https://www.england.nhs.uk/2024/12/nhs-ai-giving-patients-better-care-and-support/
41. Deaconess Health System. Deaconess Health System increases access to care with Capacity Management System. GE HealthCare Command Center. 2023.
42. Providence Swedish. Regional command center helps create capacity for 19,000 additional patients annually at Providence Puget Sound. GE HealthCare Command Center. 2025.
43. Boyer B, Huber K, Zimlichman E, Saunders R, McClellan M, Kahn C, Noach R, Salzberg C. Advancing the future of equitable access to health care: recommendations from international health care leaders. Health Affairs Scholar. 2024;2(8):qxae094. https://doi.org/10.1093/haschl/qxae094
44. Lavi S. Arcadia vs Veradigm (2026): pricing and hidden fees revealed. ITqlick. 2025. https://www.itqlick.com/compare/arcadia-analytics/allscripts-ehr

45. Richert ME, Diao G, Mancera A, Badesch B, Neupane M, Amirahmadi R, Sarzynski SH, et al. Statewide transfer coordination and patient transfer rates among hospitals during occupancy stress. JAMA Network Open. 2025;8(12):e2546002. https://doi.org/10.1001/jamanetworkopen.2025.46002
46. American Hospital Association. 2024 costs of caring: America's hospitals and health systems continue to face escalating operational costs and economic pressures as they care for patients and communities. American Hospital Association. 2025. https://www.aha.org/guidesreports/2025-04-28-2024-costs-caring
47. Oscarson N, Paxman E. Healthcare Operations Command Center ecosystem 2021. KLAS Research. 2021.
48. Oosterlinck K, Eveline D, Follet P, Djellouadji S. Measurable improvements in bed and staff capacity with hospital command centers. IQVIA. 2021.
49. Kettering Health. Kettering Health Network opens NASA-like command center. Strive. 2019. https://ketteringhealth.org/greene-memorial-hospital-receives-a-safety-grade-from-leapfrog-group/
50. Franciscan Health. Virginia Mason Franciscan Health Mission Control helps synchronize the entire patient experience. GE HealthCare Command Center. 2021.
51. Freeman G. Hopkins Command Center improves quality with coordination. Clinician. 2017.
52. Roth M. Artificial intelligence delivers $40M in real savings to Tampa General. HealthLeaders Media. 2020.

Alekhya Kommuru and Divyansh Agarwal, Stanford Graduate School of Business

Contact: div117@stanford.edu

Abstract

What is the message? India's Ayushman Bharat Digital Mission (ABDM), launched in 2021, has built a foundational digital health infrastructure at unprecedented scale, with over 880 million unique health identifiers, 500,000+ registered facilities, and 1,000+ integrated private companies. Yet the pace of engagement has not been in line with

HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION
BAHM BUSINESS SCHOOL ALLIANCE FOR HEALTH MANAGEMENT

enrolment: active record linkage and cross-facility retrieval remain far below the thresholds needed to generate clinical value. This paper argues that moving from infrastructure to impact requires a three-pronged approach: treating insurance penetration as the binding constraint and formalizing claims through National Health Claims Exchange (NHCX), investing in education and training to convert passive registrants into active users, and restructuring incentives through a combination of performance-linked payments and national mandate frameworks that close the gap between NHA's convening role and the regulatory authority held by Insurance Regulatory and Development Authority of India (IRDAI) and the Ministry of Health and Family Welfare.

What is the evidence? Drawing on 15 primary interviews between January and March 2026 with ecosystem participants spanning government bodies, development organizations, digital health startups, and large healthcare companies, alongside government dashboard data and published literature, this paper examines adoption patterns across public and private sectors, presents case studies from multiple states illustrating diverse implementation paths, and identifies the structural incentive misalignments that constrain the transition to digitally enabled care in India.

Keywords: Digital Health, Health Policy

Timeline: Submitted May 25, 2026; accepted after review July 10, 2026.

Cite as: Alekhya Kommuru, Divyansh Agarwal. 2026. From Infrastructure to Impact: Operationalizing India's Ayushman Bharat Digital Mission. Health Management, Policy and Innovation (www.HMPI.org). Volume 11, Issue 2.

Introduction

India's healthcare system faces structural challenges of extraordinary scale. With a population of 1.4 billion, a doctor-to-patient ratio of approximately 1:1,000 nationally (dropping to 1:4,500 in rural areas), and out-of-pocket expenditures historically representing nearly half of total health spending, the gap between supply and demand cannot be addressed through incremental reform (1). The government of India recognized that technology would need to function as a structural solution rather than an administrative convenience.

The Ayushman Bharat Digital Mission (ABDM) was launched nationally in September 2021. Building on the India Stack, which comprises Aadhaar biometric identity for 1.4 billion people, the Unified Payments Interface (UPI), which processed 117 billion transactions in 2023, and the Jan Dhan financial inclusion program, ABDM sought to do for healthcare what UPI had done for payments: create a shared, interoperable infrastructure layer that reduces friction, enables private innovation, and makes longitudinal care coordination possible at population scale. (2,3)

A prior paper examined ABDM's architecture and its foundations in India's digital public infrastructure strategy. If you are new to ABDM and its components, we strongly suggest reading it first. (4) This paper examines what has happened since: how ABDM is being operationalized on the ground, what adoption patterns have emerged across public and private sectors, and what structural changes are needed to move from infrastructure to impact.

This paper draws on 15 semi-structured interviews conducted between January and March 2026 with participants across India's public and private healthcare ecosystem. Interviewees included representatives from government bodies, development organizations, digital health startups, and large healthcare companies, spanning both frontline practitioners and central policy-level participants. Interview insights were synthesized through iterative discussion between the co-authors and reviewed by Professor Kevin Schulman, drawing on his expertise in digital health system design. These primary insights were triangulated against government dashboard data, peer-reviewed literature, and published policy documents throughout the analysis.

ABDM Architecture and Adoption

Architecture

ABDM's architecture comprises five interconnected components: ABHA (Ayushman Bharat Health Account), a unique 14-digit health identifier; the Health Professional Registry (HPR); the Health Facility Registry (HFR); a patient-controlled personal health record (PHR) application; and the Unified Health Interface (UHI), an open protocol for healthcare transactions. Two critical gateways have since been added: NHCX (National Health Claims Exchange) for digital insurance claims processing, and the Health Information Exchange and Consent Manager (HIE-CM), which enables patients to grant and revoke consent for cross-facility data sharing. A defining architectural choice, federated rather than centralized storage, means patient data resides with facilities and is shareable with consent across the network, rather than aggregated in a central repository. (4)

As of May 2026, ABDM's enrollment figures are substantial: 880+ million ABHA accounts, 500,000+ registered health facilities, and 860,000+ enrolled healthcare professionals. However, these figures require careful interpretation (5)(6). The CoWIN vaccination platform seeded approximately 130 million initial accounts without explicit separate consent, and Pradhan Mantri Jan Arogya Yojana (PM-JAY, a health insurance scheme offered by the Indian Government) integration has been the dominant driver of ABHA creation. A re-application anomaly, in which large numbers of duplicate creations appear in dashboard data, suggests the true denominator of unique active patients is materially lower than headline counts imply. (6)

ABDM distinguishes three engagement milestones: M1 (ABHA creation), M2 (record linkage within a facility), and M3 (cross-facility record retrieval). M1 has achieved mass scale. M2 activity is anchored to workflows like QR-based appointment check-in followed by prescription delivery to a health locker. M3, the milestone that would make ABDM genuinely valuable for care coordination, remains extremely low. This is the use case where the system's full value lies: a patient presenting at a new facility can consent to share records from prior encounters, enabling a clinician to make informed decisions without redundant testing. (5) (6)

Adoption Challenges: Consent, Labs, and Pharmacies

The steepest drop-off in the ABDM adoption funnel occurs between M2 and M3. Based on research and interviews with people familiar with ground-level execution, the consent framework emerged as a persistent friction point. When a provider wishes to retrieve records from another facility, the patient must locate a notification in their PHR app, review time-bound consent parameters, and approve the request. In practice, this loop has proven difficult to close in high-volume settings: patients are frequently unaware that a notification has been sent, the interface presents a learning curve for first-time users, and in public hospitals that provide free diagnostics, clinicians often find it more expedient to reorder tests than to navigate the retrieval workflow. A promising design direction suggested by stakeholders is to simplify consent to a QR scan plus biometric or PIN confirmation at the point of care, analogous to a UPI transaction, while preserving the privacy safeguards that make the consent architecture trustworthy. The "Scan and Share" initiative has shown meaningful progress in reducing registration wait times, though implementation teams have noted a task-shifting dynamic at high-volume facilities where congestion, rather than eliminated, is redistributed to specialist referral desks.

Three broader implementation challenges compound these ground-level friction points. First, interoperability between ABDM's open standards and the large installed base of legacy Hospital Management Information Systems remains unresolved. Many facilities run proprietary software with no ABDM-compliant API layer, and integration costs for mid-sized providers are estimated at INR 30-50 lakh. Second, provider workflow burden is a persistent barrier: ABDM's value to the patient is real, but the workflow steps required of a physician during a two-to-five-minute consultation in a high-volume public facility can add time rather than save it. Third, data governance under the 2023 Digital Personal Data Protection Act (DPDPA) introduces compliance obligations like consent artifact storage, data localization requirements, and cross-entity sharing rules, all of which create legal uncertainty for private health technology companies building on ABDM infrastructure.

In the short term, labs and pharmacies provide a great opportunity to close this gap. However, laboratory adoption has been constrained by a circular incentive gap: incumbent Laboratory Management Information System (LMIS) vendors have been slow to build ABDM-enabled versions without demand from labs, while labs have had limited incentive to request integration without patient-side pull. Breaking this dynamic would bring diagnostic results into the ecosystem. Pharmacy integration presents a similarly clear pathway: a viable digital prescription-retrieval workflow would give pharmacies a meaningful role and create a natural forcing function for broader provider and patient engagement.

Multimodal State-Level Implementation

States have been leading the charge in deeper ABDM adoption and implementation. Across the 28 states and 8 union territories, three implementation archetypes have emerged, matching policy levers to local conditions. These archetypes are not mutually exclusive and several states show elements of more than one.

Private-Sector Engagement: Karnataka, Maharashtra, Gujarat

These states have a relatively concentrated urban private market and a vendor ecosystem that has built ABDM-compliant Hospital Management Information Systems (HMIS). Karnataka has achieved 93.9% Health Facility Registry (HFR) verification and meaningful engagement from private clinics and hospitals, supported by professional networks and targeted field engagement that generated trust signals cascading to smaller providers (7). Maharashtra's Mumbai PATH pilot, IMA partnerships, and ward-level field teams have begun to onboard small clinics, although providers report few tangible incentives once field teams withdraw. Adoption in this archetype tends to begin with appointment booking and basic registration rather than full clinical record integration, and the model's replicability is constrained by its dependence on concentrated urban markets (8).

Public-Infrastructure-Led: Tamil Nadu, Uttar Pradesh, Kerala, Rajasthan

Uttar Pradesh has rolled out Scan and Share at scale, with more than 22.7 million tokens issued, supported by IMA-UP WhatsApp peer groups and a strong state program management unit (9). Tamil Nadu has paired the Makkalai Thedi Maruthuvam (healthcare at your doorstep) community's Non-Communicable Disease (NCD) outreach program with eSanjeevani, and is thus recording one of the highest volumes nationally (10). Kerala's 'eHealth Kerala' UHID platform covers more than 600 institutions and leverages the highest level of frontline digital literacy in the country, but faces identifier friction between UHID and ABHA and uneven module adoption (11). Rajasthan's iHMS auto-enrolls ABHA at the front desk, but an ABHA-to-EHR linkage rate of below 20% illustrates what happens when enrollment-oriented incentives are not balanced by usage-oriented accountability. The shared challenge across these states is that records are produced at outreach or front-desk events but are not consistently retrieved at subsequent facility visits.

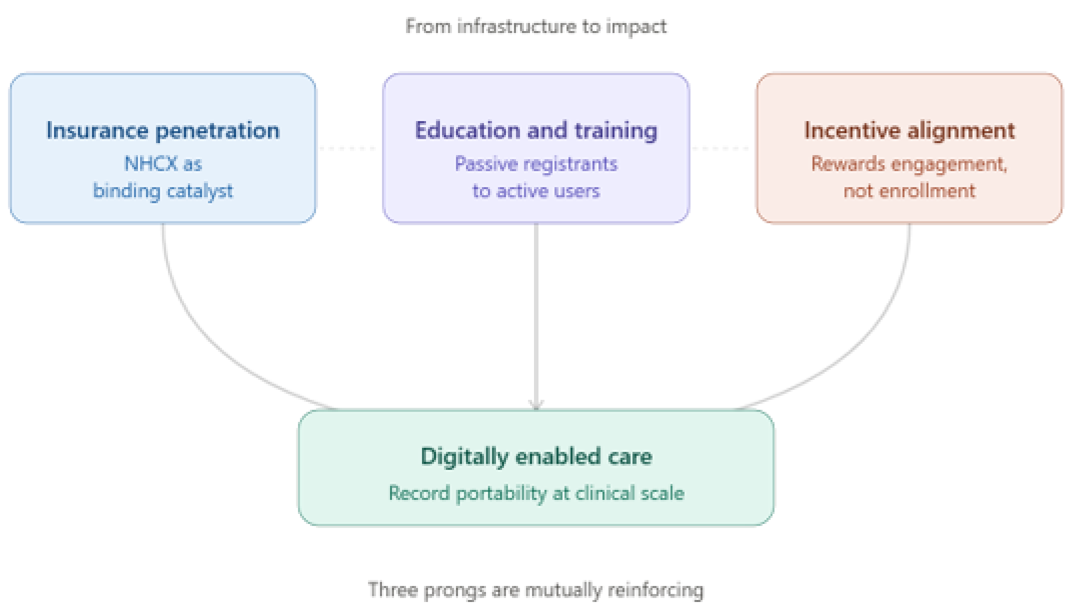
Telemedicine-Integrated: Andhra Pradesh, Telangana, Madhya Pradesh

The shared anchor in this archetype is eSanjeevani teleconsultation at Ayushman Bharat Health and Wellness Centres (AB-HWC), where community health officers provide assisted consultations that generate ABHA-linked digital prescriptions. Andhra Pradesh and Telangana have built hub-and-spoke arrangements with ABHA-linked NCD screening, and Telangana reports more than 98% AB-HWC coverage (12) (13). Madhya Pradesh ranks among the top five contributor states nationally for eSanjeevani volumes (14). The model reaches populations that no self-service digital application could serve, but its characteristic failure mode is episodic value: teleconsultations produce records that are not consistently retrieved at subsequent in-person encounters, and the continuity loop that would make ABDM clinically meaningful for chronic disease management remains largely untapped. The digital divide here is a design constraint that requires workflow solutions, not just technology solutions.

Across all active archetypes, two patterns repeat: adoption begins at registration rather than at clinical record use, and engagement plateaus once field teams withdraw unless a downstream incentive sustains it.

The Way Forward: A Three-Pronged Approach

Based on our interviews with people closer to the ground and to the central policy level, and drawing learnings from research across similar globally successful models within healthcare, we suggest a three pronged approach to move ABDM from infrastructure to impact.



1. Insurance Penetration as the Binding Constraint

A central reason ABDM's clinical value remains latent is that the underlying payment system does not yet reward digitization. India's health spending is dominated by out-of-pocket payments at the point of care, roughly 48.8% of total health expenditure, against 40.8% government spending and only 5.8% private health insurance (8). When patients pay cash and providers maintain cash-based bookkeeping, neither side gains immediate value from electronic claims, structured records, or interoperable history.

This is why NHCX matters disproportionately. As insurance penetration rises, whether through PM-JAY expansion, employer-sponsored plans, or individual cover sold under IRDAI's expanded mandates, the share of payments that must flow through a digital adjudication path grows. Each additional insurance-paid encounter is, by construction, an encounter that benefits from ABDM compliance: a digital claim, a structured discharge summary, and a verifiable record. The current cash dominance will loosen as insurance share rises, and we recommend directing policy levers at accelerating that shift.

ABDM's most persistent structural challenge on the provider side is that integration requires upfront investment with uncertain commercial return and no regulatory mandate. For payers, NHCX offers a clear commercial incentive: digital claims processing reduces fraud, accelerates the approximately 45-day average settlement cycle, and lowers administrative costs. We recommend that IRDAI require all health-insurance claims above a defined threshold to be processed through NHCX within a fixed transition window, creating an immediate mandate for insurer participation and, by extension, pressure on empaneled facilities to comply. Making NHCX integration a condition for PM-JAY empanelment would create a mandate-adjacent incentive structure that drives provider-side adoption.

2. Education and Training

A substantial share of ABDM's adoption gap is not technological but behavioral. Patients who have registered for ABHA do not understand what it does, providers who have onboarded facilities do not use digital workflows as part of the consultation, and frontline health workers who generate records through assisted channels are not trained to close the loop with patients on record retrieval. We recommend addressing this by investing across three levels.

At the patient level, the 880+ million individuals with ABHA accounts represent a registered population that has never been systematically informed about the value of their health account. We recommend a targeted national awareness program, delivered through Aarogya Setu, PM-JAY communications, and community health workers, to convert passive registrants into active PHR users.

At the provider level, the gap between facility registration and clinical workflow adoption is a training problem as much as an incentive problem. Short structured modules integrated into continuing medical education, focused on the three most common ABDM workflows (QR check-in, prescription push, and consent-based record retrieval) would reduce the friction that currently makes paper feel faster.

At the system level, community health officers at AB-HWCs, who are already the primary interface between ABDM and rural populations, need training on closing the post-visit loop: explaining to patients that their teleconsultation has generated a digital record, how to access it, and why it matters for their next encounter.

3. Incentive Alignment: Sticks, Carrots, and Mandates

The third prong addresses the structural reality that ABDM's current incentive architecture rewards enrollment over engagement and leaves participation voluntary for actors whose involvement is necessary for the system to function. We propose action on both incentive reform and mandate mechanisms.

In terms of reforming existing incentives, one solution could be changing the INR 20 per ABHA payment structure to reward active record linkage rather than mere account creation. Performance-linked payments to facilities based on M2 and M3 transaction volumes, modeled on the incentive logic of PM-JAY empanelment payments, would shift provider behavior toward the outcomes that generate clinical value. For private health technology companies, a tiered compliance framework that provides preferential access to ABDM's digital sandbox, faster certification timelines, and API rate allowances in exchange for demonstrated M3 transaction volumes would create a commercial reason to invest in the harder adoption problems.

Mandate mechanisms offer a complementary path and as discussed earlier, processing health insurance claims through NHCX and making ABDM compliance at M2 level a condition of PM-JAY empanelment, can be helpful levers. Besides these, the National Medical Commission, operating through state medical councils, could require HPR registration and verification as a precondition for license renewal, addressing the bottleneck that currently leaves a significant share of registered providers unverified and limiting the clinical trust that depends on a credible provider directory.

The National Health Authority is a convening and implementation body, not a regulator; it cannot, on its own authority, compel facilities or insurers to adopt the standards it publishes. Genuine adoption inflection will therefore require either NHA partnerships with regulators that already hold mandate authority, or a national-level mandate channeled through Parliament or the Cabinet. A national mandate framework that defines a minimum compliance floor while preserving state discretion over implementation is the policy architecture we propose.

Each of the above addresses a distinct failure mode in the current adoption trajectory, and the three are mutually reinforcing: progress on insurance penetration creates the financial incentive for providers to engage; education and training equips both patients and providers to use the system once the incentive exists; and incentive alignment through government policy ensures the system does not remain voluntary for actors whose participation is structurally necessary.

Lastly, two segments — labs and pharmacies — offer a near-term opportunity to accelerate this dynamic without waiting for broader mandate frameworks. We recommend targeted API support and onboarding incentives directed at LMIS vendors, combined with a viable digital prescription-retrieval workflow for pharmacies, to bring diagnostic results and medication records into the ecosystem and create natural forcing functions for both provider and patient engagement.

Conclusion

ABDM represents a genuine architectural achievement. In five years, India has built the infrastructure for a federated digital health ecosystem at a scale no comparable country has achieved. The central challenge is now the transition from infrastructure to impact, a transition that requires behavioral change among all ecosystem stakeholders, each operating under different incentive structures. The history of digital health implementation globally, from HITECH in the United States to NHS Digital in the United Kingdom, suggests that infrastructure build-out is the easier part. Adoption is harder, and clinical impact is hardest of all.

The three-pronged approach outlined above addresses each of the structural failure modes that currently prevent ABDM from delivering clinical value at scale. None of these interventions is technically complex; all of them require political will and cross-institutional coordination. It's also important to mention here that ABDM is ultimately intended to address the provider shortage in India through the development of a platform for digital services for patients. M2/M3 data exchange should be seen as the prerequisite building blocks of these high-quality services, not as the destination for the program. India's extraordinary physician shortage means the value of longitudinal records, which enable a stretched clinical workforce to deliver more effective care in shorter encounters, is higher here than almost anywhere. The architecture exists. The task now is activation.

References

1. World Bank. "Out-of-Pocket Expenditure (% of Current Health Expenditure) – India." World Bank Open Data. Accessed May 10, 2026.
2. National Payments Corporation of India. UPI Product Statistics. NPCI.
3. Government of India. PM Jan Dhan Yojana. PMJDY. https://pmjdy.gov.in/ (2021).
4. Narayan A, Bhushan I, Schulman K. India's evolving digital health strategy. npj Digital Medicine. 2024;7:284.
5. Ayushman Bharat Digital Mission, "ABDM-Insights," accessed May 13, 2026, https://dashboard.abdm.gov.in/abdm/
6. Mishra, Udaya Shankar, Suryakant Yadav, and William Joe. "The Ayushman Bharat Digital

Mission of India: An Assessment." Health Systems & Reform 10, no. 2 (2024): 2392290. https://doi.org/10.1080/23288604.2024.2392290.

7. Business Standard, October 2, 2024; https://www.business-standard.com/india-news/karnataka-tops-in-verified-health-facility-registry-tn-odisha-at-bottom-124100200701_1.html
8. Kaul, Varun, and Tanu Gupta. "Private Sector Engagement for Digitizing Health Care in India." PATH, March 22, 2023.
9. NITI for States – UP best practices document on Scan and Share and HMIS
10. BMC Primary Care peer-reviewed study, Dec 2025; https://link.springer.com/article/10.1186/s12875-025-03064-w
11. Discover Public Health peer-reviewed study, Jul 2025; https://link.springer.com/article/10.1186/s12982-025-00829-7
12. PMC study (Sood et al., 2025, Oxford Open Digital Health); https://pmc.ncbi.nlm.nih.gov/articles/PMC12558045/
13. PIB press release; https://www.pib.gov.in/PressReleasePage.aspx?PRID=1830743
14. NFHS-5 (National Family Health Survey 2019-21)

Other sources

- Press Information Bureau of India. NHA and C-DAC Launch e-Sushrut HMIS. PRID=2156603. January 2026.
- Ministry of Health and Family Welfare. eSanjeevani. esanjeevani.mohfw.gov.in. Accessed 2024.
- Government of India. Digital Personal Data Protection Act 2023. Ministry of Electronics and Information Technology, 2023.
- Narayan A, Schulman K. Access Health India and the Ayushman Bharat Digital Mission. Stanford GSB Case SM-359. 2022.
- Wood S, Rosengaus L, Thompson R, Patel K, Schulman K. The Digitally Enabled Care Framework: Leveraging Technology to Enhance the Physician-Patient Relationship. NEJM Catalyst Innovations in Care Delivery. 2024;5(10).

Acknowledgements

HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION

BAHM BUSINESS SCHOOL ALLIANCE FOR HEALTH MANAGEMENT

The authors conducted over 15 interviews for this research across India's public and private healthcare ecosystem, from frontline practitioners to state and central policy makers, from start founders to large health system executives. We thank all our interviewees for their time and candid feedback. We also have deep gratitude to Professor Kevin Schulman (Stanford University) for this support, insights and guidance throughout this journey. Interviews were conducted between January and March 2026. No specific funding was received for this work.

Kalgi Modi, Johns Hopkins University, Carey Business School

Contact: kmodi4@jh.edu

Abstract

What is the message? Using artificial intelligence (AI)-enabled clinical decision-support systems without changing workflows, decision authority, and governance structures can raise cognitive and administrative workload. Evidence shows that when AI is added to existing processes instead of being integrated into redesigned workflows, the burden on clinicians and the risk of burnout increase.

What is the evidence? Health management and health informatics research, including multi-hospital studies of AI-enhanced electronic health record systems and systematic reviews, show associations between AI implementation, increased cognitive burden, and higher burnout risk when AI is added to existing processes rather than integrated through thoughtful management redesign. [1-9] This article argues that AI-enabled clinical decision support should be viewed primarily as an organizational design challenge rather than a technology challenge. AI alters workflows, decision authority, and incentives, and these changes often lead to increased burnout risk when not accompanied by redesign. Finally, the article outlines practical leadership actions to connect AI implementation with workforce sustainability and patient care.

Keywords: Artificial Intelligence, Administrative Burden, Clinical Decision Support

Systems, Organizational Design

Timeline: Submitted May 11, 2026; accepted after review July 2, 2026.

Cite as: Kalgi Modi. 2026. When AI Becomes Extra Work: Redesigning Clinical Decision Support Without Clinician Burnout. Health Management, Policy and Innovation (www.HMPI.org). Volume 11, Issue 2.

The Problem AI Was Supposed to Solve

In many hospitals today, clinicians interact with screens much more than they review schedules or patient histories before a visit. AI-enabled clinical decision-support systems introduce continuous prompts, alerts, and recommendations that interrupt patient encounters and disrupt the flow of care. Rather than quietly supporting clinical reasoning, these systems now compete for attention, shaping how clinicians focus, patients experience care during clinical encounters, and leaders interpret performance.

AI-enabled clinical decision-support systems (AI-CDSS) were created to address major challenges in clinical practice. In this article, AI-CDSS refers to clinical decision-support tools that healthcare institutions approve; these tools are integrated into healthcare workflows and include systems for risk prediction, supporting triage, assisting with clinical documentation, generating alerts, offering imaging help, and recommending treatments. AI-CDSS does not refer to fully autonomous clinical AI systems, consumer health chatbots, or experimental generative AI tools that operate outside of formal care delivery settings.

Clinical practice challenges include increasing documentation demands, more complex patients, limited time during consultations, and the need to follow more clinical guidelines. AI-CDSS promised to help ease these pressures by assisting with tasks like flagging abnormal lab results, identifying patients at risk of deterioration, recommending evidence-based treatments, and prioritizing clinical tasks under time constraints. Early research appeared to support many of these expectations: evaluations of AI-supported triage and risk-stratification systems reported improved throughput, better prioritization of high-risk patients, and earlier identification of clinical deterioration. [2][10]

However, real-world use has shown a different set of challenges. In practice, many AI-CDSS implementations have introduced new layers of interaction rather than replacing existing work.[5][13] Clinicians must verify algorithm-generated risk scores, acknowledge alerts even when they are obvious, and explain any deviations from AI recommendations. In inpatient and intensive care settings, this can lead to dozens of alerts per shift, many of which require documentation regardless of their importance. What was intended to reduce cognitive burden often redistributes it into new, less visible forms of work.

These effects worsen when multiple AI tools, often created by different vendors and yet part of the same electronic health record (EHS) system, run simultaneously without coordination. Separate systems can create overlapping alerts, conflicting recommendations, or redundant documentation obligations. Instead of acting as a unified support system, uncoordinated AI tools divide attention and slow down decision-making, creating more challenges than any single tool used on its own.

Importantly, the AI-CDSS discussed here are not experimental or unregulated technologies. They are institutionally approved, regulator-compliant systems deployed within controlled electronic health record environments. Yet even within these mature settings, health systems face growing incentives to adopt AI, including competitive pressures, policy expectations, and organizational goals related to innovation. This can create pressure to implement systems before workflows are fully redesigned. The continuing challenges in these regulated settings therefore suggest that the integration of approved AI systems into clinical processes, not gaps in regulation or underdeveloped technology, are the core issue.

The shift from rule-based alerts to AI-enabled clinical decision support has increased the volume and complexity of interactions between clinicians and systems. [1][13] If workflows and decision-making authority are not redesigned, these systems may lead to more monitoring and verification work instead of a decrease in tasks. Reducing workload relies not just on better algorithms but also on rethinking workflows and decision-making structures.

These challenges mainly reflect leadership decisions regarding workflow design, task distribution, and governance. Research on successful health information technology implementation consistently highlights the importance of multidisciplinary pre-deployment workflow reviews that include clinicians, nurses, and IT teams.[6][15] When organizations introduce AI without adjusting the workflow, they may add verification and documentation tasks instead of reducing them. Some increase in workload is expected during technological transitions. The key challenge for leadership is to stop temporary complexity from becoming permanent.

From Support to Surveillance: How AI Changes Clinical Work

Before examining burnout as a design outcome, it is important to understand how AI-enabled clinical decision support systems change everyday clinical work. This essay takes a socio-technical view to explore these changes. Socio-technical theory focuses on how work outcomes result from the interaction among people, tasks, technologies, and organizational structures, rather than from technology alone. In clinical settings, this means that AI changes not just the availability of information, but also responsibility, coordination, and professional judgment.

One consequence of these changes is the creation of new forms of work that often remain invisible in traditional productivity measures. One of the most significant changes brought about by AI-CDSS is the increase in verification work. Verification work involves the extra mental and administrative effort needed to confirm, acknowledge, or reject AI-generated recommendations before taking clinical action. For instance, when an AI-CDSS flags a patient as “high risk” for deterioration, clinicians often must check the underlying data, acknowledge the alert, and record whether they agree or disagree with the recommendation even when the clinical assessment is already clear. What used to be internal clinical judgment is now an external task shaped by system prompts, documentation requirements, and organizational expectations. Similar findings have been reported in studies of EHR-based clinical decision support, where clinicians spent additional time reviewing alerts and documenting responses even when patient management did not change. [5][13]

Verification work fragments clinical attention during patient interactions where communication, reassurance, and trust are critical. Increased focus on alerts and documentation can lower eye contact, disrupt communication, and impact the perceived quality of care and patient satisfaction. In high-acuity settings, constant alarms and interruptions can raise anxiety for patients and families. These signals often indicate possible instability, even if the alerts do not require immediate clinical response. [14] While AI-CDSS aims to improve safety, poorly integrated systems can unintentionally divert attention from the patient. However, when designed effectively, these systems can lessen documentation demands and create more room for meaningful human interaction. The design of AI-CDSS affects not only clinician workload, but also patient experience and trust.

Alert volume intensifies this burden. AI-CDSS often generate alerts more frequently than earlier rule-based models because they continuously update risk predictions. McDonald et al. found that clinicians exposed to higher alert frequencies reported significantly higher perceived workload and emotional exhaustion, even with improved alert accuracy. In other words, better predictions did not lead to a better work experience when alert volume increased. The study showed that the cumulative effect of frequent alerts, not their individual usefulness, was strongly linked to burnout symptoms.[4]

As organizations work to ensure proper use of AI-CDSS, monitoring grows alongside decision support. They track override rates, review compliance dashboards, and monitor deviations from AI recommendations, which may lead to additional documentation requirements. In some organizations, clinicians feel that they need to explain more when they ignore AI guidance than when they follow it, especially if overrides are monitored or checked.[9][18]

These dynamics suggest that AI implementation can shift portions of clinical work from interpersonal coordination toward system monitoring and compliance activities. Tasks that were once handled through team discussions or professional judgment are now managed through individual interactions with automated systems. Studies on social and technical factors reveal that when technology alters relationships among people, tasks, and tools without changing workflows, coordination declines and cognitive demand increases.[3]

Empirical studies support this pattern. Research on AI-enabled documentation and decision-support tools found higher burnout scores after implementation, even with small efficiency gains. For instance, while the time spent on documentation per note decreased, clinicians reported greater mental fatigue due to the sustained need to interact with and monitor the system.[4][9] These analyses suggest that improvements in organizational performance and clinician experience do not always go hand in hand. This is especially true when new tasks are added without removing existing responsibilities for an already constrained workforce.

Clinician resistance to AI should not be seen as irrational or technophobic. Often, skepticism comes from valid concerns about workflow disruption, accountability, and professional autonomy. When AI adds more verification, monitoring, and documentation without easing other demands, resistance is a reasonable reaction to poorly designed systems.

Burnout as a Design Outcome, Not a Personal Failure

Burnout related to AI-enabled clinical decision-support systems should not be seen as a personal failure to adjust to technology. Studies on health information technology, including AI-enabled clinical decision support in electronic health records, consistently show that dissatisfaction comes more from the design, implementation, and management of these systems in clinical work than from the technology itself. Here, “health IT” specifically refers to officially approved digital systems, including AI-CDSS, that influence clinical decision-making, documentation, and care coordination. [13]

Across studies of AI-enabled clinical decision support and broader health information technology implementations, the sources of dissatisfaction are remarkably consistent despite differences in technologies, clinical settings, and implementation strategies. Three repeated issues stand out: misaligned workflows, unclear decision authority, and misaligned incentives. For instance, when AI-CDSS generates alerts but does not eliminate existing documentation obligations, clinicians experience workflow misalignment, with tasks piling up rather than being replaced. When it is unclear who is responsible for decisions, whether responsibility lies with the clinician, the organization, or the algorithm, clinicians may respond by increasing documentation or relying more heavily on AI recommendations to reduce perceived accountability risk. When efficiency gains benefit the organization without visible reinvestment in staffing or reducing workloads, the incentives continue to diverge between leadership priorities and frontline experiences. Similar implementation challenges have been reported in various health information technology projects. When technology was adopted without changing workflows, users felt less satisfied and believed their workload increased. [13][15]

When organizations do not measure these effects, burnout is often misattributed to individual resilience rather than system design. Without metrics to capture cognitive demand, alert volume, or workflow fragmentation, organizational performance reviews often focus on algorithm accuracy, adoption, and utilization metrics. Meanwhile, the human impact stays mostly unseen.

Cognitive demand theory helps explain why these design choices matter in clinical settings. Cognitive strain increases when work involves frequent task switching, sustained attention to multiple information streams, and quick decision-making in the face of uncertainty. In AI-supported clinical contexts, such as inpatient monitoring, risk stratification, and triage, AI-CDSS can intensify all three conditions simultaneously, particularly when alerts, recommendations, and documentation requirements are layered onto existing workflows.

Table 1 uses the socio-technical model of health information technology. This model views clinical work as a mix of people, tasks, technologies, and organization. Instead of requiring readers to infer this system from previous studies, the table shows how misalignment in these areas manifests in AI-CDSS implementations. When technology doesn’t fit well with the workflow, when management policies focus more on compliance than on judgment, and when feedback loops center only on technical performance, burnout becomes a predictable organizational risk when these sources of cognitive demand persist over time.

Table 1. Socio-technical domains of AI-CDSS and related burnout drivers

Socio-technical Dimension	How AI-CDSS shapes work across this domain	Burnout / Cognitive Load Implications
Hardware & Software	Embedded AI models, alerts, dashboards within EHR systems	Increased interruptions, alert fatigue, task switching
Clinical Content	Risk scores, predictions, recommendations without context tailoring	Verification burden; reduced sense of professional judgment
Human-Computer Interface	Pop-ups, mandatory acknowledgements, override prompts	Fragmented attention; cognitive overload
People	Clinicians expected to adapt to AI outputs with minimal training	Stress, reduced autonomy, erosion of psychological safety
Workflow & Communication	AI layered onto existing workflows without task removal	Work intensification; longer task completion times
Internal Organizational Policies	Monitoring of compliance, override tracking, performance dashboards	Surveillance anxiety; defensive reliance on AI
External Rules & Regulations	Regulatory pressure to document adherence to guidelines	Increased documentation load; risk-averse behavior
Measurement & Monitoring	Focus on accuracy and utilization metrics	Burnout remains invisible and unmanaged

Adapted from the socio-technical framework suggested by Sittig and Singh (2015).

The persistence of burnout, despite ongoing technical improvements, highlights this. Even as algorithms improve, dissatisfaction may persist even as technical performance improves when workflows remain unchanged. This reflects the broader "health IT paradox," in which adoption rates increase while end-user satisfaction declines. [14] Technology alone can't fix the structural and cultural problems rooted in clinical work.

Decision Authority, Trust, and Automation Bias

Trust in AI-based clinical decision support involves more than just clinician faith in algorithms. Trust in healthcare decisions is shaped by confidence in clinicians, institutions, regulators, and, increasingly, digital technologies. In some cases, AI can increase skepticism by raising worries about surveillance, opacity, or loss of professional autonomy. In others, it may increase confidence by improving consistency and supporting decisions under pressure. These tensions apply to both highly interpretable ("glass-box") systems and more opaque ("black-box") models. While explainable AI may improve transparency and clinician understanding in some settings, both types of systems can still increase verification work, documentation burden, and override pressure when implementation outpaces workflow redesign. The challenge for leaders not only is to build trust in AI, but also to understand how trust is shared among clinicians, organizations, patients, and technology.

When it's unclear who holds responsibility, whether it's the clinician, the organization, or the AI system, clinicians may change their behavior out of caution. For instance, if moving away from AI guidance requires extra documentation or review, following the recommendation might seem safer than using one's own judgment. In these situations, relying on AI may reflect institutional accountability pressure as much as confidence in algorithm performance. Research on automation bias suggests that diagnostic errors can increase when clinicians rely heavily on AI recommendations under time pressure or uncertainty. [18]

These dynamics also impact patients and families. Some patients may see AI-supported decisions as more objective or based on data. Others might find them impersonal or hard to question. Trust thus becomes about relationships instead of just technical aspects. It is influenced by how clinicians explain their recommendations and how organizations convey the role of AI in care delivery.

As AI becomes embedded in routine workflows, skepticism may gradually shift toward greater reliance. [13] Without well-defined guidelines for questioning and overriding AI recommendations, sustained reliance may gradually increase the risk of automation bias and reduce independent review of recommendations.

Leadership is fundamental in changing this direction. Clarifying decision authority involves clearly stating when AI provides advice, when clinicians have the final say, and how overrides are assessed and safeguarded. Governance structures that create this balance foster measured trust, avoiding both underuse and over-reliance. [12]

Incentives, Organizational Outcomes, and the Burnout Gap

AI investments in healthcare are often backed by expected benefits like improved throughput, lower costs, better prioritization of high-risk patients, and operational efficiencies. However, evidence about actual system-level benefits is mixed and heavily depends on the implementation context, workflow integration, and organizational readiness. [2][10]

Economic evaluations of AI-enabled triage and prioritization tools suggest that organizational benefits and frontline workload do not always move in the same direction. [10] Patients may also experience these tradeoffs indirectly when implementation priorities emphasize operational metrics without equal attention to communication, continuity of care, and clinical capacity. When operational improvements are not accompanied by staffing support or workflow redesign, frontline clinicians may perceive AI implementation as increasing organizational demands without providing meaningful relief.

Several studies have found that organizational performance indicators and clinician experience may differ during implementation. Reported improvements in throughput or operational performance may occur alongside an increased cognitive burden when new responsibilities are added without taking away existing work. [4][10][16]

This gap between organizational incentives and frontline experience drives burnout risk. At the board level, this means viewing AI-CDSS investments not just as capital expenses for efficiency. They should also be viewed as investments in organizational redesign. This may require additional investments in staffing support, changes in workflow, training, and governance capacity. Over time, AI-CDSS might change how organizations measure performance. It could focus more on compliance metrics than on adaptive clinical judgment. This shift may influence which activities organizations reward, measure, and prioritize in clinical practice.

The REDESIGN Framework for AI-Enabled Decision Support

The operational challenges described earlier, such as verification burden, alert fatigue, unclear decision authority, and misaligned incentives, are not isolated issues. They result from introducing AI into clinical workflows without changing how work is organized and managed. Each element directly addresses a breakdown identified earlier, connecting diagnosis to intervention instead of offering a general set of recommendations. The REDESIGN framework translates these failures into seven leadership actions that address workflow, authority, governance, measurement, incentives, and professional judgment. [7]

Many redesign actions can be tested within current operational limits. This can be done through phased implementation, specific governance changes, and workflow review instead of needing significant technology investment. In 3 to 6 months, organizations can test targeted redesign efforts, assess their impact, and expand successful changes.

1. **Redesign workflows before deployment:** AI-CDSS should replace low-value work instead of adding new verification layers. Before implementation, leaders should map workflows, standardize documentation when possible, and pinpoint tasks that can be simplified, automated, or eliminated.
2. **Establish clear decision authority:** Leaders should clearly define who is responsible when AI recommendations are involved and when these fail. Governance structures should allow clinicians to easily override AI recommendations without adding unnecessary administrative work.
3. **Govern AI throughout its lifecycle:** AI governance should continue after deployment. Organizations should regularly check performance, unintended consequences, and changes in clinician experience over time. [12]
4. **Evaluate human impact alongside performance:** AI evaluation should go beyond accuracy and usage metrics. Leaders should watch alert burden, mental load, clinician fatigue, patient experience, workforce well-being, and operational outcomes. [17]
5. **Reinvest operational improvements into staffing and wellness:** When organizations observe operational improvements associated with AI implementation, leaders should consider reinvesting some of those gains into staffing, training, workflow redesign, or clinician well-being efforts. [16
6. **Normalize questioning and overrides:** Organizations should encourage questioning and proper overrides of AI recommendations. Creating psychological safety helps guard

against automation bias and maintains independent clinical judgment. [13]

7. **Establish data readiness before scaling AI:** AI systems depend on reliable, interoperable, and consistently structured data. Before expanding AI-CDSS across departments, organizations should evaluate data quality, integration standards, and governance processes to ensure recommendations remain accurate and comparable across settings.

The goal of redesign is not to remove oversight, but to make sure it is appropriate, meaningful, and matches clinical judgment instead of being redundant or too much.

A Phased "When" Roadmap

A phased implementation approach is more manageable when deployment, evaluation, and scale-up occur sequentially rather than simultaneously.

- **Pre-deployment:** Create governance structures with input from clinicians, nurses, IT teams, and operational leaders. Involving stakeholders early helps to improve workflows, incentives, and documentation practices before deployment.
- **Pilot phase:** Use factors like alert frequency, clinician feedback, override patterns, and perceived cognitive workload to spot unintended consequences early during implementation. [6]
- **Scale-up:** Connect demonstrated operational improvements to workforce support, workflow redesign, and ongoing governance review. [17]

Policy and Global Implications

The challenges discussed in this paper go beyond individual hospitals. Global digital health strategies focus more on governance, interoperability, workforce readiness, and human-centered implementation as essential for responsible AI use. The World Health Organization's Global Strategy on Digital Health 2020-2025 highlights many of these priorities, such as leadership capacity, digital governance, and workforce engagement. [19] These priorities reinforce the central argument of this paper: successful AI implementation depends not only on technological capability, but also on organizational design, governance, and human factors.

Conclusion

AI-CDSS should be viewed as an organizational change rather than merely a technology upgrade. Its impact depends less on model performance and more on how leaders redesign workflows, clarify decision-making authority, govern implementation, and evaluate human consequences. Similar to the transition from paper records to electronic systems, AI adoption may initially introduce complexity before longer-term benefits emerge. The challenge for leaders is to prevent temporary implementation burdens from turning into permanent features of clinical work. Organizations most likely to benefit from AI will not necessarily be those with the most advanced algorithms, but those most willing to rethink how clinical work is supported, governed, and experienced.

References

1. Topol E. Deep Medicine. New York: Basic Books; 2019.
2. Ash JS, et al. Artificial intelligence and the future of health-care delivery. The Lancet Digital Health. 2020;2(12):e664-e666.
3. Sittig DF, Singh H. A new socio-technical model for studying health information technology. Quality and Safety in Health Care. 2015;24(1):6-15.
4. McDonald M, et al. Impact of clinical decision-support alerts on clinician workload and burnout. Journal of the American Medical Informatics Association. 2022;29(3):467-476.
5. Koppel R, et al. Ten common types of electronic health record-related errors. Journal of the American Medical Association. 2008;299:1010-1012.
6. Carayon P, et al. Human factors and ergonomics in health care. Human Factors. 2018;60(5):679-692.
7. Verghese A, Johnson KB. Redesigning AI-enabled workflows to reduce cognitive burden. Health Management Policy & Innovation. 2024 (forthcoming).
8. Lee SI, et al. Predictive analytics and burnout: a systematic review. BMJ Open. 2021;11:e048567.
9. Lin SC, et al. Clinician trust and automation bias. Annals of Internal Medicine. 2020;172:777-785.
10. Patel VL, et al. Economic evaluation of AI-driven triage. Health Economics.

2023;32(7):1021-1034.
11. O'Connor S, et al. Governance models for clinical AI. NPJ Digital Medicine. 2022;5:71.
12. Denecke K. Ethical implications of artificial intelligence in health care. Frontiers in Public Health. 2021;9:635332.
13. Buntin MB, et al. The health-IT paradox: adoption versus satisfaction. Health Affairs. 2019;38(12):1999-2007.
14. Zhang Y, et al. Alert fatigue in intensive care unit settings. Critical Care Medicine. 2022;50(4):567-574.
15. Kim HJ, Lee JH. User-centered design for AI-based clinical decision support. International Journal of Medical Informatics. 2023;169:105041.
16. Patel A, et al. Reinvestment strategies for AI savings. Journal of Health Management. 2024;26(2):112-124.
17. Friedman CP, Wyatt JC. Evaluation methods for clinical informatics. Journal of the American Medical Informatics Association. 2020;27(4):596-603.
18. Rosen MA, et al. Automation bias in diagnostic decision-making. Radiology. 2019;293(2):373-380.
19. World Health Organization. Global Strategy on Digital Health 2020–2025. Geneva: WHO; 2020.

Brooke Shearon, Kristine Burnaska, and Erin Weber, DataSpring, powered by CAQH; Brooke Istvan, Graduate School of Business, Stanford University; and Kevin Schulman, School of Medicine and Graduate School of Business, Stanford University

Contact: BShearon@dataspring.com

Abstract

What is the message? Payer-provider contracting remains largely manual and document-based, creating inefficiencies and limiting transparency across the healthcare system. Modernizing contracting with standardized, structured, and more digital approaches could reduce administrative burden and better support value-based care.

What is the evidence? Findings are based on data from the 2025 CAQH Index survey, which includes health payers representing 63 percent of covered lives and over 600 provider organizations. DataSpring, powered by CAQH, collaborated with Stanford Medicine to better understand lags in contracting advancements and how the industry can push advancement.

Keywords: Payer-Provider Contracting, Administrative Burden, Digital Transformation, Value-Based Care

Timeline: Submitted May 4, 2026; accepted after review June 15, 2026.

Cite as: Brooke Shearon, Kristine Burnaska, Erin Weber, Brooke Istvan, Kevin Schulman. 2026. The State of Payer and Provider Contracting: Challenges and Opportunities. Health Management, Policy and Innovation (www.HMPI.org). Volume 11, Issue 2.

Prior Dissemination: *This article was published on the DataSpring, powered by CAQH, website on Friday, March 6, 2026. It was promoted on DataSpring, powered by CAQH, LinkedIn, on Monday, April 13, 2026.*

Introduction

Contracting between payers and providers quietly shapes how care is financed, how networks are built, and how organizations work together to move healthcare forward.[1] By defining the services providers deliver and how payers reimburse them, contracting ensures patients can access the care they need.[2] Yet new findings from the 2025 CAQH Index [3] show that this critical process still relies on various manual methods. Drawing on the industry's most comprehensive view of administrative transactions and contracting practices, DataSpring, powered by CAQH, partnered with Stanford Medicine to understand why contracting has not advanced at the same pace as other technological developments in the healthcare industry and what payers, providers, consultants, and policymakers can do to drive modernization.

The 2025 CAQH Index reveals a contracting environment that remains manual, fragmented, and misaligned with the industry's shift toward more connected, data-driven systems.[4] As other administrative workflows modernize,[3] contracting continues to rely on manual tools and individualized practices that create inconsistency across the industry that drives up costs for all.

These inconsistencies reflect differences in organizational capacity, resources, and priorities, which shape how teams approach contracting and influence an organization's willingness to change its processes. On the provider side, larger practices often approach contracting with more tools and support, while smaller practices face constraints that limit their ability to change. Payers and providers also prioritize different outcomes, making it harder to move toward shared processes or standardized formats.

This article examines why current variation exists in contracting and possible opportunities for improvement. By addressing fragmented payment arrangements, standardizing contracts, and simplifying payment processes, the industry can begin reducing the billing and insurance-related burden consuming a substantial share of United States healthcare spending without wholesale reform of coverage.[5],[6],[7],[8]

The Document-Driven Approach to Healthcare Contracting

Contracting across payers and providers, particularly agreements governing network participation, reimbursement terms, and associated administrative requirements, continues to rely on manual, document-based workflows that have changed little over time.

Exhibit 1: Tools Currently Used in Payer/Provider Contracting Strategies, Medical and Dental, 2025 CAQH Index

Response Option	Medical Payers (n=9)	Dental Payers (n=3)	Medical Providers (n=324)	Dental Providers (n=219)
Contract Management Software (e.g., Conga, Icertis, DocuSign CLM)	44.4%	N/A	15.4%	25.1%
Manual Spreadsheets (e.g., Excel, Google Sheets)	66.7%	100%	28.1%	26.9%
E-signature Platforms (e.g., DocuSign, Adobe Sign)	44.4%	33.3%	46.6%	40.2%
Provider Network Management Systems	78.7%	100%	18.8%	20.1%
CRM or Sales Tools (e.g., Salesforce, HubSpot)	33.3%	66.7%	2.8%	5.0%
Contract Analytics or AI Review Tools (e.g., Kira, ThoughtTrace)	22.2%	N/A	3.4%	5.9%
Payer/Provider Portals or Contract Submission Platforms	44.4%	66.7%	32.7%	32.0%
Homegrown or Custom-Built Contracting Solutions	55.6%	66.7%	4.3%	5.5%
Policy and Regulatory Compliance Tools	22.2%	66.7%	8.3%	11.0%
External vendors	56.5%	100%	14.5%	11.4%
Lawyers	66.7%	100%	7.1%	7.3%
Other (please specify)	N/A	N/A	2.2%	0.5%
We did not use any formal tools	N/A	N/A	13.0%	18.3%
Unsure	11.1%	N/A	17.0%	11.9%

This question was multi-select, and users were instructed to select all that apply. Percentages may add up to more than 100%, as respondents could select more than one response.

N/A = 0 respondents selected this option.

Two-thirds of medical payers and all dental payers report using spreadsheets and lawyers as primary tools for their contracting processes. Providers similarly report use of manual tools, with more than one quarter of medical and dental organizations using spreadsheets for core contracting tasks and more than 30 percent utilizing payer/provider portals.

Reliance on unstructured documents or external relationships makes it challenging to maintain consistent records, understand contract changes over time, or align contract terms with downstream administrative processes. Tools like spreadsheets do not capture reimbursement methodologies or structured regulatory language in ways that support automation, an important factor in reducing operational burden and lowering administrative costs.[9], [10]

Digital tools, such as contract execution through e-signature platforms, are being used to perform specific tasks, with over 40% of medical payers, medical providers, and dental providers reporting use. However, these systems do not transform the underlying content into structured data. Contracts largely remain as PDFs or static files that cannot be easily analyzed or integrated with claims or provider data systems, leaving contracting disconnected from other administrative workflows.[11],[12] Additionally, managing numerous health payer contracts, each with its own platform, format, and timeline, can create a nearly constant stream of administrative work for provider practices.[13], [14]

While more advanced solutions, such as provider network management software, contract management software, and contract submission platforms, are available to everyone, they tend to be used by larger organizations with more resources. As shown in Exhibit 2, 42% of hospitals and 31% of large generalists (practices with five or more providers) rely on contract management software compared to 14% of small generalists.

Differences Across Stakeholders and Drivers of Variation

Despite the overarching use of manual tools across the industry, approaches to contracting differ across organizations, shaped by variation in size, staffing, and administrative capacity.

Providers

Larger provider practices tend to approach contracting with more infrastructure behind them.

On average, they are about ten percentage points more likely to involve legal review, adopt digital tools such as e-signature platforms, and maintain contract management software than smaller practices across specialties. Larger practices also show higher willingness to adopt new contracting formats (on average, about 14 percentage points more than smaller practices), suggesting that scale and available resources may influence how organizations evaluate potential changes.

Smaller practices operate under different conditions. With fewer administrative resources, they often rely more heavily on external vendor support and portals to manage contracting tasks than larger practices (on average, about three percentage points more). These constraints also affect how they view modernization. Findings show that smaller practices report lower willingness to change their contracting format, particularly if a new approach would introduce additional burden for their teams.

Exhibit 2: Tools Currently Used in Payer/Provider Contracting Strategies by Medical Provider Specialty and Size, 2025 CAQH Index

Response Option	Generalists (n=100)		Specialists (n=99)		Behavioralists (n=113)		Hospitals (n=12)
Number of Physicians in Practice	<5 (n=71)	5+ (n=29)	<5 (n=74)	5+ (n=25)	<5 (n=97)	5+ (n=16)	
Contract Management Software (e.g., Conga, Icertis, DocuSign CLM)	14.1%	31%	13.5%	20%	9.3%	12.5%	41.7%
Manual Spreadsheets (e.g., Excel, Google Sheets)	22.5%	31%	28.4%	44%	18.6%	68.8%	41.7%
E-signature Platforms (e.g., DocuSign, Adobe Sign)	49.3%	58.6%	44.6%	48%	38.1%	62.5%	58.3%
Provider Network Management Systems	18.3%	17.2%	21.2%	24%	16.5%	31.3%	N/A
CRM or Sales Tools (e.g., Salesforce, HubSpot)	1.4%	3.4%	4.1%	8%	N/A	6.3%	8.3%
Contract Analytics or AI Review Tools (e.g., Kira, ThoughtTrace)	1.4%	10.3%	2.7%	8%	N/A	12.5%	8.3%
Payer/Provider Portals or Contract Submission Platforms	35.2%	31%	37.8%	32%	28.9%	43.8%	8.3%
Homegrown or Custom-Built Contracting Solutions	N/A	10.3%	5.4%	8%	1%	18.8%	8.3%
Policy and Regulatory Compliance Tools	5.6%	13.8%	9.5%	12%	3.1%	25%	16.7%
External vendors	16.9%	13.8%	14.9%	12%	15.5%	12.5%	N/A
Lawyers	7%	13.8%	1.4%	16%	5.2%	12.5%	16.7%
Other (please specify)	1.4%	N/A	4.1%	4%	2.1%	N/A	N/A
We did not use any formal tools	14.1%	N/A	14.9%	N/A	19.6%	N/A	16.7%
Unsure	22.5%	27.6%	16.2%	16%	12.4%	6.3 %	16.7%

This question was multi-select, and users were instructed to select all that apply. Percentages may add up to more than 100%, as respondents could select more than one response.

N/A = 0 respondents selected this option.

These patterns align with national research indicating that administrative burden falls disproportionately on smaller practices, who often face higher relative administrative costs and have fewer resources to dedicate to modernization efforts.[15], [16], [17] These dynamics make it harder to adopt new systems or participate in standardization efforts. They also contribute to the economic pressures pushing physicians into employed arrangements.[18]

Exhibit 3: Willingness to Change Payer/Provider Contracting Format, by Medical Provider Specialty and Size, 2025 CAQH Index

Response Option	Generalists (n=100)		Specialists (n=99)		Behavioralists (n=113)		Hospitals (n=12)
Number of Physicians in Practice	<5 (n=71)	5+ (n=29)	<5 (n=74)	5+ (n=25)	<5 (n=97)	5+ (n=16)	
5 – High, our contract structure is burdensome 4	26.8% (19)	37.9% (11)	27% (20)	40% (10)	14.4% (14)	31.3% (5)	33.3% (4)
3	38.0% (27)	31.0% (9)	47.3% (35)	40% (10)	49.4% (48)	62.5% (10)	33.3% (4)
2 1 – Low, we will not change even with positive return-on-investment (ROI)	35.2% (25)	31.0% (9)	25.7% (19)	20% (5)	36.1% (35)	6.3% (1)	33.3% (4)

In addition to size, contracting priorities vary by specialty. As shown in Exhibit 4, while hospital-based respondents and generalists cited value-based care or alternative payment models as key influences on their contracting strategies, other specialties placed greater emphasis on quality improvement and outcomes-based care. Additionally, generalists and specialties ranked cost containment and financial performance higher than behavioralists and hospitalists. However, behavioralists and hospitalists were more likely to express interest in changing their organizational or contracting structure, which may indicate that some specialties focus on improving financial performance within existing contracts while others are more inclined to pursue broader structural changes.

These findings indicate that smaller practices need modernization options that reduce net workload without adding additional burden. Larger organizations can use contract management tools to model terms and downstream impacts, but they still depend on payers to simplify reimbursement structures and align contract language with systems.

Exhibit 4: Factors Influencing Payer/Provider Contracting Strategies by Medical Provider Specialty and Size, 2025 CAQH Index

Response Option	Average Rank of Generalists (n=100)		Average Rank of Specialists (n=99)		Average Rank of Behavioralists (n=113)		Average Rank of Hospitals (n=12)
Number of Physicians in Practice	<5 (n=71)	5+ (n=29)	<5 (n=74)	5+ (n=25)	<5 (n=97)	5+ (n=16)	
Cost containment and financial performance	1.92	1.83	1.72	1.75	1.88	2	2.17
Network adequacy and access requirements	2.05	2.33	1.95	2	2.36	2	2.25
Quality improvement and outcomes-based care	1.92	1.38	2	1.4	1.71	1.86	2.13
Regulatory and compliance requirements	2.05	2.38	2	2	1.96	2	1
Value-based care or alternative payment model alignment	1.84	2	1.94	2.5	2.11	2	1
Market competitiveness and differentiation	2.09	2.75	2.38	1.67	1.82	2.13	1.75
Provider/payer relationship management and collaboration	2.19	1.58	1.74	2.25	2.05	1.33	2.2
Member/patient satisfaction and experience	2	1.9	2.27	2.38	1.91	2.6	2.5
Data-driven insights and performance analytics	2.2	2	2	2	2.33	1.67	2
Other (please specify)	1.8	N/A	N/A	2.33	1.83	N/A	N/A
I am unsure of my organization's payer/provider contracting strategy.	1.27	1	1.50	1	1.25	1	1

Due to the small sample size, payer data cannot be reported for this table.

Users were instructed to rank their top 3 factors, with "1" being their organization's top priority.

N/A = 0 respondents selected this option

Payers

Payers approach contracting through a different operational lens, influenced by the scale and complexity of their systems. Most medical and dental payers involve their legal teams throughout the contracting process. More than half of medical payers emphasize maintaining alignment between contracting and operational functions such as claims, payments, and provider data. This reflects the scale and complexity of payer operations but may not align directly with smaller provider needs.[19]

Payers recognize a clear connection between contracting complexity and administrative performance. More than half of medical payers (56%) and one third of dental payers said they would change their contracting format to improve operational efficiency. This highlights a clear opportunity to simplify workflows and reduce administrative burden. Payers are also motivated to pursue contracting changes if those investments lead to better negotiated terms or reimbursement rates (one third of medical and dental payers).

By contrast, medical and dental providers were primarily concerned with reducing the burden on their contract negotiation (46% and 53%, respectively) and showed substantially less interest in negotiated terms (3% of medical providers and 2% of dental providers). Overall, these findings suggest that payers view contracting as a strategic lever to improve both administrative efficiency and financial performance, rather than solely as a transactional process.

Exhibit 5: Biggest Incentive to Change Contracting Format, Medical and Dental, 2025 CAQH Index

Response Option	Medical Payers* (n=9)	Dental Payers (n=3)	Medical Providers (n=324)	Dental Providers (n=219)
Less burden on our contract negotiations team	11.1% (1)	N/A	46.3% (150)	53.4% (117)
Better negotiated terms or reimbursement rates	33.3% (3)	33.3% (1)	2.5% (8)	1.8% (4)
Opportunity for improved efficiency of back-end processes (e.g., claims, payments, provider data)	55.6% (5)	33.3% (1)	6.2% (20)	3.2% (7)
Improved compliance with regulatory or accreditation requirements	11.1% (1)	N/A	4.3% (14)	2.3% (5)
Faster contract turnaround time	11.1% (1)	33.3% (1)	3.1% (10)	2.3% (5)
Enhanced visibility and tracking across contract lifecycle	11.1% (1)	N/A	8.0% (26)	6.8% (15)
Greater alignment with value-based care models or alternative payment	N/A	N/A	10.8% (35)	11.9% (26)
Other (please specify)	N/A	N/A	1.5% (5)	1.8% (4)
Unsure	22.2% (2)	N/A	17.3% (56)	15.5% (34)

* Due to plans mistakenly erroneously responding to more than one response option, the total percentage exceeds 100.

N/A = 0 respondents selected this option

Based on the findings, payers recognize the link between contract complexity and administrative performance. To move the industry forward, contracts should be designed as structured, computable data that can flow through claims, provider data, and payment integrity systems over the life of the provider relationship, not just as PDFs.

Lessons Learned and Next Steps

Overall, contracting can become more consistent, predictable, and easier to navigate. Through stakeholder collaboration and element standardization, the industry can align and address underlying sources of variation and reduce administrative burden. Industry workgroups can help define technical guidance to facilitate the use of structured data and workflows. Greater consistency in areas like reimbursement methods, renewal terms, and data-sharing expectations would create a more uniform foundation to work from.[20] Workgroups can also serve as a forum to discuss implementation considerations, including how modernization investments are prioritized and operationalized across stakeholders with varying needs and capabilities. Developing computable contracts could offer a pathway to improved efficiency for health payers and providers across the healthcare system.[21]

To ensure that improvements are effective, efforts should reflect differences across provider types, including primary care, specialty care, and hospital-based organizations. Given the varying contracting needs and capacities across stakeholders, a single, standardized approach may not provide meaningful outcomes, especially for smaller practices. These organizations often face a disproportionate administrative load on already burdened staff, and tools that simplify documents or offer ready-made templates could help modernize the contracting process, making it more feasible. Larger organizations may also benefit from clearer structures, making it easier to align contracting with downstream systems and reducing the time spent reconciling contract terms with operational processes.[22]

For consultants and government agencies, these findings point to a need for end-to-end contracting transformation, not only better rate negotiations. While negotiated outcomes are driven by market dynamics, investments in contracting capabilities can support greater standardization, improved data quality, and more efficient contracting processes. For policymakers in particular, standardizing core contracting elements such as reimbursement methodologies, renewal terms, and data sharing expectations can help advance policy goals related to network adequacy, payment integrity, and provider directory accuracy.

Driving meaningful change requires stakeholders to prioritize modernization efforts, invest in tools and processes that reduce administrative complexity, and collaborate to create a more consistent, data-driven contracting environment. Stronger alignment between payers and providers can reduce administrative steps, increase predictability, and clarify expectations across organizations, regardless of size or stakeholder. By understanding the costs tied to manual processes, organizations can prioritize changes with meaningful impact and become better positioned to support the industry’s shift toward more connected and data-driven systems.[3], [23]

References

[1] American Medical Association. Payor contracting for health care 101 [Internet]. Chicago (IL): American Medical Association; [cited 2026 Mar 6]. Available from: https://edhub.ama-assn.org/steps-forward/video-player/18947118

[2] Medical Group Management Association. Medical practice payer contracting glossary [Internet]. Englewood (CO): MGMA; [cited 2026 Mar 6]. Available from: https://www.mgma.com/payer-contracting-glossary

[3] CAQH. The CAQH Index report [Internet]. Washington (DC): CAQH; [cited 2026 Mar 6]. Available from: https://www.caqh.org/insights/caqh-index-report

[4] Centers for Medicare & Medicaid Services. Application programming interfaces (APIs) and relevant standards and implementation guides (IGs) [Internet]. Baltimore (MD): CMS; [cited 2026 Mar 6]. Available from: https://www.cms.gov/priorities/burden-reduction/overview/interoperability/implementation-guides-and-standards/application-programming-interfaces-apis-and-relevant-standards-and-implementation-guides-igs

[5] Jiwani A, Himmelstein D, Woolhandler S, Kahn JG. Billing and insurance-related administrative costs in United States’ health care: synthesis of micro-costing evidence. BioMed Central (BMC) Health Services Research. 2014;14:556.

[6] Turner A, Miller G, Lowry E. High U.S. health care spending: where is it all going? [Internet]. New York (NY): The Commonwealth Fund; 2023 Oct 4 [cited 2026 Mar 6]. Available from: https://www.commonwealthfund.org/publications/issue-briefs/2023/oct/high-us-health-care-spending-where-is-it-all-going

[7] Scheinker D, Richman BD, Milstein A, Schulman KA. Reducing administrative costs in US

health care: assessing single payer and its alternatives. Health Services Research. 2021;56(4):615-625.

[8] Tseng P, Kaplan RS, Richman BD, Shah MA, Schulman KA. Administrative costs associated with physician billing and insurance-related activities at an academic health care system. The Journal of the American Medical Association (JAMA). 2018;319(7):691-697.

[9] Sirion. Understanding healthcare contract compliance: what you need to know [Internet]. London: Sirion; [cited Mar 6]. Available from: https://www.sirion.ai/library/contract-management/healthcare-contract-compliance/

[10] NTRACTS, The hidden causes of spreadsheets in physician contract management [Internet]. Brentwood (TN): NTRACTS; 2025 Nov 12 [cited Mar 6]. Available from: https://www.ntracts.com/blog/the-hidden-costs-of-spreadsheets-in-physician-contract-management/

[11] SolumHealth. Payer contract management: a complete guide for healthcare practices [Internet]. Solum Health; [cited 2026 Mar 6]. Available from: https://www.getsolum.com/glossary/payer-contract-management-guide

[12] Aroris. Understanding the revenue cycle in healthcare: payer contracts [Internet]. Aroris; 2025 Jul 23 [cited 2026 Mar 6]. Available from: https://www.arorishealth.com/blog/revenue-cycle-in-healthcare/

[13] CAQH. The hidden causes of inaccurate provider directories [Internet]. Washington (DC): CAQH; [cited 2026 Mar 16]. Available from: https://www.caqh.org/hubfs/43908627/drupal/explorations/CAQH-hidden-causes-provider-directories-whitepaper.pdf

[14] American Medical Association. 2024 AMA prior authorization physician survey [Internet]. Chicago (IL): American Medical Association; 2025 May 22 [cited 2026 Mar 6]. Available from: https://www.ama-assn.org/system/files/prior-authorization-survey.pdf

[15] O'Malley AS, Gupta DP, Rich RC, Lewis C. Administrative burden in primary care: causes and

potential solutions [Internet]. New York (NY): The Commonwealth Fund; 2025 [cited 2026 Mar 6]. Available from: https://www.commonwealthfund.org/publications/issue-briefs/2025/oct/administrative-burden-primary-care-causes-potential-solutions

[16] Scott BA. Reducing regulatory burden – a major opportunity for new administration [Internet]. Chicago (IL): American Medical Association; 2025 May 22 [cited 2026 Mar 6]. Available from: https://www.ama-assn.org/about/leadership/reducing-regulatory-burden-major-opportunity-new-administration

[17] Office of the National Coordinator for Health Information Technology. Strategy on reducing regulatory and administrative burden relating to the use of health IT and EHRs [Internet]. Washington (DC): ONC; [cited 2026 Mar 6]. Available from: https://www.healthit.gov/sites/default/files/page/2020-02/BurdenReport_0.pdf

[18] Physicians Advocacy Institute. PAI-Avalere report on physician employment trends and acquisitions of medical practices: 2019-2023 [Internet]. Chicago (IL): Physicians Advocacy Institute; 2024 Apr [cited 2026 Mar 6]. Available from: https://www.physiciansadvocacyinstitute.org/PAI-Research/PAI-Avalere-Study-on-Physician-Employment-Practice-Ownership-Trends-2019-2023

[19] Douglas C. Navigating the challenges of narrowing provider networks [Internet]. Waterloo (IA): VGM & Associates; 2025 Apr 1 [cited 2026 Mar 6]. Available from: https://www.vgm.com/communities/adapting-to-narrowing-provider-networks/

[20] CAQH CORE. Operating rules [Internet]. Washington (DC): CAQH; [cited 2026 Mar 16]. Available from: https://www.caqh.org/core/operating-rules

[21] Istvan B, Schulman KA, Zenios S. Addressing health care's administrative cost crisis. The Journal of the American Medical Association (JAMA). 2025;333(9):749-750.

[22] symplr. The real ROI of modernizing contract management in healthcare [Internet]. Houston (TX): symplr; 2025 Dec 3 [cited 2026 Mar 6]. Available from:

https://www.symplr.com/blog/contract-management-symplr-value-proposition?

[23] Sahni NR, Gupta P, Peterson M, Cutler DM. Active steps to reduce administrative spending associated with financial transactions in US health care. Health Affairs Scholar. 2023;1(5):qxad053.

Bryn M.C. Rediger, Boston University, Questrom School of Business

Contact: brediger@bu.edu

This paper won first place in the inaugural 2025/25 Will Mitchell Student Essay Contest.

Abstract

What is the message? The traditional healthcare "Iron Triangle" holds that improving cost, access, or quality often requires tradeoffs because healthcare resources are finite. At the same time, the Triple Aim framework argues that healthcare systems should pursue better patient experience, improved population health, and lower per capita costs simultaneously through better system design and the elimination of waste. This analysis argues that artificial intelligence (AI) can help advance that goal by functioning as administrative infrastructure that reduces waste, lowers coordination costs, and makes value-based care more operationally feasible. AI does not eliminate scarcity or remove the need for difficult allocation decisions. Rather, by reducing administrative complexity and improving coordination between payers and providers, AI may expand the feasible frontier of cost, access, and quality within existing resource constraints.

What is the evidence? Drawing on published estimates of administrative waste, prior authorization burden, and emerging evidence from AI-enabled claims validation and utilization management tools, this analysis examines how administrative infrastructure influences the relationship between cost, access, and quality. It synthesizes implementation studies, payer-provider applications, regulatory developments, and governance frameworks to explore how AI may reduce administrative friction embedded in

HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION

healthcare payment and delivery systems. The analysis builds on the insight that if administrative waste, rather than clinical scarcity alone, accounts for part of the traditional tradeoff between cost, access, and quality, then reducing that waste may allow healthcare systems to make progress across multiple dimensions simultaneously. In this context, AI is considered not as a substitute for clinical judgment or resource allocation, but as a potential mechanism for lowering the administrative costs of coordination, measurement, and payment.

Keywords: Artificial Intelligence, Administrative Costs, Administrative Infrastructure

Timeline: Submitted May 11, 2026; accepted after review July 3, 2026.

Cite as: Bryn M.C. Rediger. 2026. AI as Administrative Infrastructure: Reducing Waste and Expanding the Feasible Frontier of Cost, Access, and Quality. Health Management, Policy and Innovation (www.HMPI.org). Volume 11, Issue 2.

Administrative Fragmentation and the Cost of Paying for Care

The U.S. healthcare system is often described through the "Iron Triangle" of cost, access, and quality, a framework suggesting that meaningful improvement in one domain generally requires tradeoffs in another because healthcare resources are finite.[1] The framework remains valuable because it highlights a fundamental reality of health policy: scarcity cannot be eliminated. Yet many health system scholars have argued that not all observed tradeoffs arise from resource constraints alone.[2,3] Some reflect waste embedded within healthcare delivery and financing systems.[4]

The Triple Aim framework reframed healthcare improvement around the simultaneous pursuit of three objectives: improving the experience of care, improving population health, and reducing per capita costs.[2] Instead of treating these goals as inherently conflicting, the Triple Aim argues that better system design can create progress across all three domains at once. In a related analysis, scholars identified administrative complexity, failures of care coordination, and other forms of waste as major contributors to excess spending that does not improve patient outcomes.[5] Their central argument was that reducing waste, rather than reducing valuable care, represents one of the largest opportunities to improve system performance while preserving or enhancing quality.[5]

This distinction is important. If a meaningful share of healthcare spending is consumed by administrative friction rather than clinical care, then some apparent tradeoffs between cost, access, and quality may reflect inefficient system design rather than unavoidable scarcity. The question then is whether technology can reduce the administrative burden required to coordinate, monitor, and pay for care. Artificial intelligence may provide one mechanism for doing so.

Administrative activities account for an estimated 15% to 30% of total U.S. health spending—roughly $600 billion to $1 trillion annually—and as much as half of that burden may represent excess costs.[6] Hundreds of commercial insurers, dozens of self-insured employers, and multiple public programs each maintain distinct benefit designs, formularies, network rules, and documentation standards. Even the federal government, the largest purchaser, operates through 50 separate Medicaid programs and hundreds of Medicare Advantage plans, limiting its ability to impose uniform standards. Healthcare systems must therefore navigate thousands of plan-product combinations, driving hospital administrative costs to roughly 17% of total spending.[7] Each additional payer rule or benefit variation adds new workflows for staffing, coding, reconciliation, and dispute resolution.

Prior Authorization (PA) is one of the clearest manifestations of this administrative burden. Originally intended to promote evidence-based and cost-effective care, PA has evolved into a major driver of delayed access and clinician burnout. A national survey by the American Medical Association shows 93% of physicians report PA-related care delays, and 29% report serious adverse events linked to those delays.[8] A typical practice completes 39-45 PAs per physician each week, consuming roughly 13 staff hours.[8] Multiplied across thousands of practices and millions of requests, PA represents recurring coordination cost at scale.

The deeper issue is strategic: both payers and providers have rational incentives to add administrative capacity. Payers invest in denial algorithms to manage cost exposure; providers invest in coding, appeals, and revenue cycle optimization to protect revenue.[4] Reworking a denied claim is stated to cost practices an average of $25, while each upheld denial allows payers to avoid payment for high-cost services.[9] These incentives drive payers to automate cost control and providers to automate countermeasures—fueling an "automation arms race" in which both sides profit from escalation even as total system costs rise.[4] In practice, the U.S. healthcare system relies heavily on costly after-the-fact review and dispute resolution rather than proactive coordination and prevention.[4]

Viewed through the Triple Aim, this dynamic represents more than administrative inconvenience. It reflects resources devoted to managing fragmentation rather than improving patient experience, population health, or affordability. If a meaningful share of spending is absorbed by administrative complexity, then reducing those coordination costs may allow healthcare systems to achieve better outcomes with existing resources. Without infrastructure that lowers this burden, value-based care (VBC) remains conceptually attractive but operationally difficult to execute.[10] The result is a system that is not only expensive, but structurally geared toward administrative conflict rather than clinical collaboration.

AI as Administrative Infrastructure for Expanding the Iron Triangle

AI's potential contribution lies in reducing the administrative burden associated with coordination, measurement, and payment. As administrative infrastructure, it can support more efficient information exchange, performance monitoring, and collaboration across fragmented healthcare organizations.

Many AI discussions in healthcare focus on labor replacement. The more consequential role of AI is institutional: it functions as coordination infrastructure that reduces the administrative burden embedded in complex payment and delivery systems.[11,12] In fragmented multi-payer markets, the core challenge is the high cost of information exchange, monitoring, and enforcement.[4] In other words, AI can lower the costs of gathering information, negotiating terms, verifying documentation, and resolving disputes. These transaction costs manifest in hours spent reworking denied claims, the weekly staff time devoted to PA, and the resources required to keep pace with frequently changing payer rules.

AI may reduce friction between these fragmented stakeholders by enabling real-time, bidirectional coordination. Machine-learning-based claims editing tools, for example, are increasingly being deployed to review submissions against payer-specific rules before claims are transmitted, helping identify missing documentation or coding inconsistencies earlier in the revenue cycle.[13,14] Commercial platforms such as Waystar and prior authorization platforms such as Cohere Health demonstrate that these capabilities are already being used in practice.[15] More broadly, large insurers and health systems increasingly describe AI as a tool for automating manual, data-intensive processes such as prior authorization, claims administration, customer service, and documentation review.[14] Broader scalability, however, remains dependent on continued implementation of interoperability standards such as HL7 FHIR and CMS-0057-F. Rather than relying on retrospective correction and appeals, these approaches seek to shift administrative effort toward prevention and coordination earlier in the workflow.

AI may also reduce the burden of performance measurement, a longstanding bottleneck in VBC.[16] Effective management requires timely, reliable data on spending and outcomes, yet much of that information is buried in unstructured clinical notes, scanned documents, and inconsistently coded reports.[17] Analyses of Medicare Cost Reports and hospital financial filings show that even basic administrative categories can be hard to compare across institutions because of inconsistent definitions and reporting practices.[7] Specifically, over 30% of subcategories and nearly 50% of the dollars reported are mislabeled.[7] Natural language processing models can extract clinically relevant variables—such as diagnoses, disease severity, comorbidities, and treatment response—from narrative documentation and map them to standardized vocabularies.[17] This converts unstructured clinical notes into measurable data that can support quality monitoring, authorization validation, and contract performance tracking.[11] Combined with rules engines that encode coverage criteria, these systems can assess whether documentation supports medical necessity and flag gaps before a record reaches a human reviewer.[11] In effect, AI upgrades documentation from passive record-keeping to actionable infrastructure.

More ambitiously, AI infrastructure could help address several operational barriers that have historically limited the scalability of VBC. Moving from fee-for-service to outcomes-oriented models requires a way to monitor performance and manage risk without drowning organizations in new paperwork.[16] Despite decades of experimentation, fragmentation across payers, providers, benefit designs, and reporting requirements continues to create substantial administrative complexity.[4] If administrative and measurement costs can be substantially reduced, value-based arrangements may become more operationally attractive than traditional fee-for-service reimbursement models.

The collaboration between Regence and MultiCare Health System offers a glimpse of this future. Emerging frameworks that combine AI-enabled workflows with HL7 FHIR standards can support more efficient data exchange between providers and payers and may reduce delays associated with authorization and documentation processes.[18] Instead of functioning solely as a cost-control mechanism, PA can become a “shared clinical safety net” that supports appropriate care paths while minimizing avoidable delays.[16] When AI is treated as managerial infrastructure, it provides real-time data flows, quality dashboards, and predictive models needed to make shared-risk contracts operationally sustainable for both large healthcare systems and smaller practices.[16] Notably, UnitedHealth executives reported that nearly all current AI applications within the organization are administrative rather than clinical, reflecting the scale of opportunity associated with reducing transaction and coordination costs.[14] AI does not eliminate fragmentation, but it can lower the coordination costs that fragmentation imposes.[4] In this sense, AI does more than automate administrative tasks. By reducing administrative waste and improving coordination across fragmented healthcare organizations, it may expand the amount of quality and access that healthcare systems can achieve with existing resources, shifting the practical boundaries of the Iron Triangle itself.

Aligning Incentives to Make Value-Based Care Work

VBC is meant to reward improvements in health outcomes relative to cost, but many organizations find that their value-based contracts underperform or stall.[19] Even as adoption of alternative payment models has grown to about 50%, a substantial share of total payments remains predominantly fee-for-service or only weakly tied to quality.[20] A flood of quality measures, delayed data, and unclear attribution rules creates friction that discourages participation.[21]

AI may help address these operational barriers by reducing the burden of measurement, reporting, and oversight that has historically limited the scalability of complex value-based contracts.[4,16] Rather than layering additional static rules onto legacy workflows, AI-enabled platforms can interpret unstructured data, identify patterns across large datasets, and support routine administrative decisions.[22] By lowering the administrative costs of coordination and performance tracking, AI has the potential to make value-based arrangements more operationally sustainable.

Bundled payments may become more scalable under this model. Traditionally, authorizing an entire episode of care—such as a joint replacement or complex oncology regimen—requires a series of discrete approvals for each imaging study, procedure, or infusion.[16] Emerging frameworks that combine HL7 FHIR data standards, Clinical Quality Language, and AI-driven decision support allow payers and providers to define episode-specific authorization rules that can be applied using standardized clinical data.[13,22] This approach may reduce administrative burden for both parties and align day-to-day workflows with the goals of episode-based payment.

AI may also improve the monitoring of shared-savings contracts. Building on the documentation infrastructure described in Section II, AI-enabled platforms can incorporate extracted clinical variables into quality dashboards that track care patterns against coverage criteria and clinical guidelines.[17] Rather than relying exclusively on retrospective audits, organizations may be able to monitor clean-claim rates, risk-adjusted cost trends, and quality performance more continuously, allowing earlier identification of operational issues during an episode or contract year.[17]

The same measurement capabilities may also strengthen outcomes-based pharmaceutical contracts, which tie manufacturer payment to real-world performance such as reduced hospitalizations or sustained disease control.[16] These arrangements are increasingly important as specialty drugs account for a growing share of health spending, yet they are difficult to administer at scale because they require tracking outcomes across diverse populations and care settings.[12] Using the data extraction and monitoring infrastructure described above, AI systems can help identify predefined clinical endpoints—such as hospital admissions, treatment discontinuation, or adverse events—and flag cases relevant to contractual payment terms.[17] Human reviewers remain essential for interpretation, dispute resolution, and ethical oversight, but AI could reduce the manual burden by surfacing relevant cases and standardizing performance tracking. By embedding outcome monitoring into routine data flows, AI supports accountability without recreating the administrative friction that has historically limited these contracts.

AI may also make risk-sharing contracts more feasible for a broader set of providers, not only large integrated systems.[7,17] Effective population health management depends on identifying patients at highest risk who would benefit from proactive outreach and coordinated care.[23] Predictive models that integrate clinical, claims, and social determinants data have demonstrated improved risk stratification compared with approaches based primarily on demographics and diagnosis codes, enabling more targeted interventions such as care management, telehealth follow-up, or home-based services.[4,17] When combined with administrative tools that support claims review, documentation management, and authorization workflows, these capabilities may reduce the operational burden of managing risk and allow providers to focus greater attention on prevention and care coordination.

Reducing administrative burden, however, does not automatically create value. Waste elimination requires organizations to invest in new systems, redesign workflows, and change established patterns of care delivery. Yet the financial benefits of those improvements do not always accrue to the organizations making the investment. As Brent James has argued, many forms of healthcare waste persist because payment mechanisms create misaligned incentives, allowing savings generated through quality improvement and waste reduction to flow elsewhere in the system.[3] Administrative complexity is both an information problem and an incentive problem. The central barrier is therefore not only the speed or sophistication of technology. Many of the tools needed to reduce administrative burden already exist, but their impact depends on whether payment models, governance structures, and organizational incentives reward waste reduction rather than the preservation of existing revenue streams.[3] AI may lower the transaction costs associated with coordination, measurement, and oversight, but its impact will depend on whether payment models reward organizations for converting those efficiencies into better outcomes and lower total costs. The greatest opportunity arises when waste reduction is paired with value-based arrangements that allow providers, payers, and patients to share in the benefits of improved performance. Under those conditions, administrative savings can be reinvested into care redesign, prevention, and innovation rather than absorbed by the existing structure of the system.[24]

AI's strategic contribution lies in supporting coordination, measurement, and transparency across fragmented healthcare organizations.[25] By lowering the operational cost of coordination, AI may help make incentive alignment more feasible across a range of value-based payment models, allowing organizations to translate administrative efficiencies into improvements in cost, access, and quality.

Expanding the Iron Triangle

The Iron Triangle has long been used to describe the tradeoffs between cost, access, and quality.[1] At any given level of spending, healthcare systems face limits on how much access and quality they can deliver. Expanding coverage or increasing service intensity has traditionally required additional resources.[3] If a meaningful share of healthcare spending is absorbed by administrative complexity rather than patient care, reducing that waste may create opportunities to improve cost, access, and quality within existing resource constraints.

Reducing administrative waste may improve cost, access, and quality through several mechanisms:

Cost: Administrative activities such as prior authorization, claim correction, documentation review, and payment disputes consume substantial resources without directly improving patient outcomes.[4] To the extent that savings are captured and reinvested, they create opportunities to improve value without reducing access or quality.[24]

Access: Access is often constrained by both clinical capacity and administrative processes that delay care.[16,26] Prior authorization requirements frequently postpone treatment, create additional work for clinical staff, and discourage participation in some insurance networks.[8] Technologies that reduce administrative friction may improve the timeliness of care and allow clinicians to devote more time to direct patient services.[27] By reducing delays associated with approvals, documentation, and care coordination, AI may help patients receive appropriate care more efficiently.

Quality: High-quality care depends on timely information, consistent measurement, and effective coordination across care settings.[2] AI may support these functions by organizing clinical information, identifying documentation gaps, and facilitating performance monitoring.[12] These capabilities do not replace clinical judgment, but they may strengthen the infrastructure required to deliver evidence-based care and identify opportunities for intervention earlier in the course of treatment.

In this framework, gains in cost, access, and quality arise not from overcoming scarcity, but from reducing the waste and administrative complexity that consume resources without adding value. AI does not eliminate tradeoffs: resources remain finite, and ethical allocation still requires human judgement.[27] Its potential contribution lies in lowering the cost of managing those tradeoffs. By reducing the share of spending absorbed by administrative waste and improving coordination across fragmented healthcare organizations, AI may expand the feasible frontier of cost, access, and quality. In that sense, AI helps free capacity for care rather than allowing administrative friction to consume a disproportionate share of the system’s value.[12]

Governance: Preventing AI-Enabled Abuse

The same AI infrastructure that can shrink administrative waste can also be weaponized to deepen it if incentives and oversight are misaligned. Recent lawsuits illustrate how quickly AI can become a denial engine rather than a coordination tool.[11] One class action alleges that a major insurer used a proprietary algorithm to deny 300,000 claims in a matter of weeks, with an average review time of 1.2 seconds—raising concerns about whether meaningful clinical judgement was ever applied.[11] Similar cases accuse other large insurers of using AI tools to terminate coverage for post-acute care and skilled nursing facilities at scale with minimal human review.[27] These episodes illustrate AI-enabled abuse: the use of opaque algorithms to maximize short-term financial margins.[11,28]

The same capabilities that allow AI to reduce administrative waste can also make opportunistic behavior cheaper and easier if oversight is weak.[11] To ensure AI functions as a shared clinical safety net rather than an adversarial tool, governance must be treated as a strategic design problem built on five pillars.

1. **Algorithmic transparency and labeling**. The FDA's framework for AI/ML-based Software as a Medical Device emphasizes a "Predetermined Change Control Plan" specifying which aspects of an algorithm may change over time and how those changes will be controlled.[28] By analogy, AI systems used for claims and PA should be required to document their objective functions and key inputs: are they optimizing predicted cost, adherence to coverage criteria, or patient outcomes?[28] Choosing the wrong target—such as short-term cost instead of health need—risks baking structural bias into the system.[23]
2. **Public reporting of authorization and denial patterns**. Beginning in 2026, CMS's Interoperability and Prior Authorization Final Rule (CMS-0057-F) will require many payers to publicly report PA metrics including total requests, approval and denial rates, and average decision times, with the first reports due by March 31, 2026.[29] These metrics create a baseline for external oversight: stakeholders can compare payers on how often and how quickly they authorize care, and regulators can investigate outliers that may signal inappropriate use of automation.[26,27]
3. **Rigorous audits across the AI lifecycle**. Best practice for AI governance emphasizes

continuous real-world performance monitoring, post-market surveillance, and bias assessment.[28] For administrative AI, this should include sampling automated decisions, comparing AI-driven denials to clinician judgements, and analyzing impacts across patient groups.[27] Without consistent and transparent reporting, it is nearly impossible to tell whether AI reduces administrative waste or simply moves it elsewhere in the system.[7]

4. **Alignment with medical loss ratio (MLR) and related regulation**. The Affordable Care Act's MLR rules require most insurers to spend 80-85% of premiums on clinical services and quality improvement.[4] However, when premiums rise, administrative spending can still increase in absolute terms even if the percentage remains constant.[4] Regulators could refine MLR frameworks to distinguish between investments that durably reduce per-member administrative costs and strategies that merely sustain complex, revenue-preserving billing structures.[7] One option would be to require that any documented savings from reduced billing- and insurance-related expenses be directly returned to the system—through lower premiums, enhanced benefits, or reinvestment in value-enabling infrastructure.[7]
5. **Reinvestment of savings in VBC**. If even a fraction of the estimated 15-30% administrative share reflects excess cost, then reducing that burden through AI could unlock a substantial system dividend.[4] Without explicit policy, those savings are likely to be captured by intermediaries.[11,27] Large purchasers and regulators can require documented reinvestment of administrative savings into lower cost-sharing, expanded benefits, care management, primary care, or social supports for patients. CMS-0057-F's transparency provisions are an important first step; future value-based contracts can go further by tying participation to demonstrable reductions in administrative waste and clear reinvestment plans.[21,26]

Ultimately, governance determines whether AI becomes a tool for systemwide efficiency or a mechanism for automated abuse. Aligning AI deployment with transparent objectives, rigorous auditing, and regulatory incentives that reward true waste reduction is what transforms AI from a denial accelerator into the infrastructure that enables VBC.[7,16] Governance therefore determines whether AI expands the healthcare frontier or accelerates existing inefficiencies.

Global Implications

The Iron Triangle framework applies globally: all healthcare systems must balance cost, access, and quality under finite resources.[1] What differs across countries is how much of the health budget is absorbed by administering that balance.[21] In the U.S., fragmented payer structures and complex billing rules have transformed administrative oversight into a major cost center, consuming an estimated 15-30% of total spending.[6,30] Over time, these structures have become path dependent: once a system builds layered prior authorization, multi-payer adjudication, and adversarial claims review processes, they generate constituencies, workflows, and technologies that are costly to unwind.[4] AI in the U.S. context therefore functions largely as a tool to reduce coordination costs within an already complex institutional architecture.

For many low- and middle-income countries (LMICs), the opportunity is different—and potentially more transformative. As nations expand insurance coverage and digitize health records, they are still shaping the administrative backbone of their systems.[31] Where AI in the U.S. reduces coordination costs within an already fragmented architecture, LMICs can apply the same principles earlier—embedding interoperability before fragmentation becomes institutionalized. AI-enabled infrastructure therefore offers a “design-stage” advantage: interoperable data standards, automated eligibility verification, and real-time authorization protocols can be embedded before fragmented billing ecosystems become entrenched.[31] Rather than digitizing paper-heavy bureaucracies, countries can implement digital-first coordination models that minimize retrospective dispute resolution and reduce incentives for adversarial billing behavior.[31] In effect, AI provides a chance to avoid importing the administrative arms race that characterizes high-cost systems.

In centrally financed or single-payer environments, AI further strengthens procurement and integrity functions by reducing information asymmetry between purchasers and suppliers. Real-time analytics can benchmark prices, detect anomalous billing patterns, and monitor contractual outcomes, preserving limited public budgets for frontline care.[15,17]

However, adoption remains uneven: the U.S. alone accounts for 75% of total investments in generative AI, raising the risk that administrative efficiency gains—and the fiscal space they create—will accrue disproportionately to already well-resourced systems.[17]

Globally, AI's greatest strategic value lies not in replacing clinicians, but in re-engineering the administrative state. Countries that deploy AI to reduce coordination costs and embed transparency early may expand the feasible frontier of cost, access, and quality simultaneously. Those that deploy it primarily to automate cost denial risk entrenching inefficiency at digital speed. VBC is therefore not the endpoint of AI-enabled administrative redesign, but an early demonstration of a broader principle: when coordination becomes cheaper than conflict, healthcare payment itself can be reorganized around value rather than volume. In that sense, AI expands the economic frontier of how care can be financed, measured, and delivered.

References

1. Kissick W. *Medicine's Dilemmas: Infinite Needs Versus Finite Resources*. 1994.
2. Berwick DM, Nolan TW, Whittington J. The Triple Aim: Care, Health, And Cost. *Health Affairs*. 2008;27(3):759-769. doi:10.1377/hlthaff.27.3.759
3. Hearing on Reducing Health Care Costs: Eliminating Excess Health Care Spending and Improving Quality and Value for Patients. Published online July 17, 2018. https://www.help.senate.gov/imo/media/doc/James3.pdf
4. Kocher RP. Reducing Administrative Waste in the US Health Care System. *JAMA*. 2021;325(5):427-428. doi:10.1001/jama.2020.24767
5. Berwick DM, Hackbarth AD. Eliminating Waste in US Health Care. *JAMA*. 2012;307(14):1513-1516. doi:10.1001/jama.2012.362
6. Health Affairs Research Brief. The Role Of Administrative Waste In Excess US Health Spending. Published online October 6, 2022. doi:10.1377/hpb20220909.830296
7. Sahni NR, Istvan B, Bello Thornhill H, Joynt-Maddox KE, Cutler D, Emanuel EJ. Availability of consistent, reliable, and actionable public data on US hospital administrative expenses. *Health Aff Sch*. 2025;3(5). doi:10.1093/haschl/qxaf069
8. American Medical Association. *2024 AMA Prior Authorization Physician Survey*. AMA; 2024. https://www.ama-assn.org/system/files/prior-authorization-survey.pdf

9. Poland L, Harihara S. Claims Denials: A Step-by-Step Approach to Resolution. Journal of AHIMA. April 25, 2022. Accessed February 2, 2026. https://journal.ahima.org/page/claims-denials-a-step-by-step-approach-to-resolution
10. Tong N. Why insurers still see value-based care as the answer to high costs. June 1, 2026. http://www.modernhealthcare.com/insurance/mh-insurers-value-based-care-unitedhealth-cvs/
11. Bichey B. A Novel, AI-Based Method of Mitigating Prior Authorization Denials. AAO-HNS Bulletin. July 15, 2024. Accessed February 15, 2026. https://bulletin.entnet.org/business-of-medicine-practice-management/article/22914205/a-novel-aibased-method-of-mitigating-prior-authorization-denials
12. McKinsey. *Economic Potential of Generative AI: The next Productivity Frontier*. McKinsey & Company; 2023. Accessed February 9, 2026. https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier
13. Schlossberg. Availity Launches AI Claims Denial Prediction Tool. *Availity*. April 2, 2024. https://www.availity.com/news/availity-launches-ai-claims-denial-prediction-tool/
14. Tozzi J. UnitedHealth's $3B AI push has bots calling doctors. June 19, 2026. http://www.modernhealthcare.com/insurance/mh-unitedhealth-ai-optum-sandeep-dadlani/
15. Waystar. Waystar Authorization Platform. Published online 2026. https://www.waystar.com/our-platform/financial-clearance/authorizations/
16. ACMA. Value-Based Care and Its Impact on Prior Authoriza… Prior Authorization Training. August 18, 2025. Accessed February 10, 2026. https://www.priorauthtraining.org/understanding-value-based-care-and-its-impact-on-prior-authorization/
17. Somajohassula DK. AI-Driven Decision Support Systems in Healthcare Claim Adjudication. *EJCSIT*. 2025;13(42):54-63. doi:10.37745/ejcsit.2013/vol13n425463
18. Anderson H. Using FHIR APIs Significantly Improved Prior Authorization Process. November 29, 2024. https://hl7news.hl7.org/2024/11/29/using-fhir-apis-significantly-improved-prior-authorization-process/
19. HCPLAN. *2022 APM Measurement Infographic*. Health Care Payment Learning & Action Network; 2022. Accessed February 24, 2026. https://hcp-lan.org/2022-infographic/

20. Leao DLL, Cremers HP, van Veghel D, Pavlova M, Groot W. The Impact of Value-Based Payment Models for Networks of Care and Transmural Care: A Systematic Literature Review. *Appl Health Econ Health Policy*. 2023;21(3):441-466. doi:10.1007/s40258-023-00790-z
21. KFF. *Health Care Costs and Affordability*. Kaiser Family Foundation; 2025. Accessed February 10, 2026. https://www.kff.org/health-costs/health-policy-101-health-care-costs-and-affordability/
22. Lenert LA, Lane S, Wehbe R. Could an artificial intelligence approach to prior authorization be more human? *JAMIA*. 2023;30(5):989-994. doi:10.1093/jamia/ocad016
23. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science*. 2019;366(6464):447-453. doi:10.1126/science.aax2342
24. Smith GL, Hannah S, Mountford J, Bisognano M. Model to reduce waste in healthcare and add value. *BMJ Open Qual*. 2022;11(1):e001655. doi:10.1136/bmjoq-2021-001655
25. Health Care Transformation Task Force. The Task Force Releases Principles for Prior Authorization Policies in Value-Based Care Arrangements. Health Care Transformation Task Force. September 6, 2023. https://hcttf.org/the-task-force-releases-principles-for-prior-authorization-policies-in-value-based-care-arrangements/
26. ACP. Toolkit: Addressing the Administrative Burden of Prior Authorization. February 5, 2024. Accessed February 8, 2026. https://www.acponline.org/advocacy/state-health-policy/toolkit-addressing-the-administrative-burden-of-prior-authorization
27. Karwar A. How to Responsibly Implement AI in Healthcare: Enhancing Prior Authorization Claims Assessments | CMSA. Case Management Society of America. June 7, 2024. https://cmsa.org/how-to-responsibly-implement-ai-in-healthcare-enhancing-prior-authorization-claims-assessments/
28. FDA. *Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) Action Plan*. U.S. Food & Drug Administration; 2021:1-8. https://www.fda.gov/media/145022/download
29. CMS. *CMS Interoperability and Prior Authorization Final Rule CMS-0057-F*. Centers for Medicare & Medicaid Services; 2024. Accessed February 14, 2026.

https://www.cms.gov/newsroom/fact-sheets/cms-interoperability-and-prior-authorization-final-rule-cms-0057-f
30. Turner A, Miller G, Lowry E. High U.S. Health Care Spending: Where Is It All Going? *The Commonwealth Fund*. Published online October 4, 2023. doi:10.26099/r6j5-6e66
31. Zuhair V, Babar A, Ali R, et al. Exploring the Impact of Artificial Intelligence on Global Health and Enhancing Healthcare in Developing Nations. *J Prim Care Community Health*. 2024;15. doi:10.1177/21501319241245847

Amanda Wong, Tanzina Khan, Melissa Florus, Milagros Ramos, Miguel Rodriguez, Yi Zhuang, and Tolga Aydinliyim, Zicklin School of Business, Baruch College, CUNY

Contact: Tolga.Aydinliyim@baruch.cuny.edu

Abstract

What is the message? Unplanned readmissions remain a persistent challenge for U.S. healthcare systems, generating substantial financial and operational problems in post-discharge care coordination. Traditional follow-up reminder programs typically apply uniform outreach across patients despite substantial heterogeneity in readmission risk, leading to inefficient allocation of limited care management resources. By studying the effectiveness of AI-assisted targeted intervention systems in reducing 30-day hospital readmissions among patients with diabetes, we demonstrate that the operational value of AI in healthcare lies not in perfect prediction but in data-driven prioritization and efficient allocation of scarce post-discharge resources.

What is the evidence? Using a large publicly available multi-hospital diabetes dataset, we integrate descriptive, predictive, and prescriptive analytics within a unified operational decision framework. We first estimate patient-level readmission risk through interpretable predictive models, then embed these risk scores into simulation-based decision models to assess alternative outreach policies. Our results show that follow-up outreach to the highest-risk 10–20% of patients in our sample captures approximately 25% to 40% of the

HMPI | HEALTH MANAGEMENT, POLICY & INNOVATION
BAHM BUSINESS SCHOOL ALLIANCE FOR HEALTH MANAGEMENT

total achievable reduction in readmissions. Beyond this range, marginal benefits decline sharply, indicating diminishing returns to uniform outreach to all patients.

Keywords: Artificial Intelligence, Hospital Operations

Timeline: Submitted June 1, 2026; accepted after review July 1, 2026.

Cite as: Amanda Wong, Tanzina Khan, Melissa Florus, Milagros Ramos, Miguel Rodriguez, Yi Zhuang, Tolga Aydinliyim. 2026. Operational Effectiveness of AI-Assisted Targeted Intervention System Design in Healthcare: Reducing 30-Day Readmission Rates For Diabetes Patients. Health Management, Policy and Innovation (www.HMPI.org). Volume 11, Issue 2.

Introduction, Background, Motivation

Healthcare systems in the United States continue to face mounting pressure to improve patient outcomes while operating under increasingly constrained financial and operational conditions. One persistent challenge is unplanned hospital readmissions, particularly among patients with chronic conditions such as diabetes. Readmissions within 30 days of discharge are associated with poorer outcomes, fragmented care, and substantial costs, and they trigger financial penalties under value-based reimbursement programs. Despite standardized discharge planning and follow-up procedures, readmission rates remain high, suggesting that existing approaches are not sufficiently tailored to patient-level risk and operational realities. In 2020, approximately 3.85 million 30-day all-cause readmissions occurred in the United States, with an average cost of $16,300 per readmission, representing the aggregate cost exceeding $62 billion annually [1]. Among Medicare beneficiaries alone, the readmission rate reached 17.0 per 100 index admissions [1].

Post-discharge interventions such as follow-up calls and patient reminders are resource-intensive. Staff time, coordination effort, and financial cost limit the feasibility of applying these interventions uniformly. As a result, healthcare organizations must decide whom to target, when to intervene, and how intensively to deploy limited outreach resources. Traditional reminder systems often adopt a one-size-fits-all approach, providing the same level of follow-up regardless of risk. This can lead to inefficiencies, with resources directed toward low-risk patients while higher-risk individuals may not receive sufficient or timely support.

Advances in data analytics and artificial intelligence (AI) provide an opportunity to redesign reminder systems in a more targeted and operationally efficient manner. AI-driven predictive models can estimate patients' relative risk of readmission at discharge, enabling health systems to prioritize outreach efforts toward those most likely to benefit. Importantly, the value of such models does not depend on near-perfect prediction. In readmission settings, outcomes are influenced by unobserved factors such as social determinants of health and post-discharge environments that are not captured in standard datasets. Even so, modestly informative risk rankings can guide decisions under capacity constraints.

In this essay, we examine the effectiveness of a targeted intervention approach compared to a more resource-intensive uniform approach, using an AI-assisted reminder system to reduce 30-day hospital readmissions among patients with diabetes. We integrate descriptive, predictive, and prescriptive analytics within a single operational framework: first characterizing patterns of readmission, then generating individualized risk scores, and finally embedding these scores into a prescriptive simulation-based model to evaluate alternative follow-up policies. Applying this framework to a large publicly available multi-hospital dataset, we find that targeting the top 10–20% of risk-ranked patients captures 25–40% of the total achievable reduction in 30-day readmissions while requiring substantially fewer interventions than universal outreach. Beyond this threshold, additional coverage yields diminishing marginal returns.

These findings suggest that the operational value of AI in healthcare lies not in perfect prediction, but in enabling disciplined prioritization and more efficient allocation of limited resources. The managerial implication is clear: hospitals can improve outcomes and protect financial performance by embedding interpretable risk ranking into post-discharge workflows rather than expanding blanket outreach programs. Importantly, existing risk assessment tools often focus on identifying high-risk patients, similar to sepsis alerts or traditional readmission prediction models. The novelty of this paper is not simply predicting readmission risk, but translating risk scores into an operational decision rule that helps leaders determine how much outreach to provide, which patients to prioritize, and where marginal returns begin to decline. This shifts AI from a prediction tool into a practical resource allocation framework for healthcare operations.

This paper proceeds in four parts. First, in Section 2, we review practical evidence and extant literature on readmissions, reminder systems, and AI-enabled risk stratification. Then, in Section 3, we present a case study, which uses sample diabetes readmission data to generate patient-level readmission risk scores and assesses the effectiveness of targeted follow-up intervention to reduce readmissions. In Section 4, we present a replication study using a large publicly available data set to derive a Pareto curve of readmission reduction versus intervention coverage under various model assumptions to confirm robustness of our main policy finding. Section 5 concludes with a discussion of managerial and policy trade-offs.

Literature Review

Readmissions as a Policy and Operational Challenge

Hospital readmissions represent a major financial and operational challenge for healthcare systems. Section 3025 of the Affordable Care Act of 2010 established the Hospital Readmissions Reduction Program (HRRP), with payment reductions for hospitals with excess 30-day readmissions taking effect on October 1, 2012 [2]. Under HRRP, Centers for Medicare and Medicaid Services reduces base operating diagnosis-related group payments for hospitals with higher than expected 30-day readmissions, which target six conditions: acute myocardial infarction, heart failure, pneumonia, chronic obstructive pulmonary disease, coronary artery bypass graft surgery, and elective knee or hip arthroplasty. Payment reductions are capped at 3% of base Medicare inpatient payments. Although diabetes is not a directly targeted HRRP condition, diabetic patients are disproportionately represented in HRRP-covered admission due to high rates of comorbid heart failure, AMI, and pneumonia. Since HRRP’s implementation, the Medicare Payment Advisory Commission (MedPAC) has documented sustained reductions in readmission rates for targeted conditions [3]. Unplanned 30-day readmissions impose substantial costs on hospitals and payers, with the average cost of readmission ($16,300) exceeding the average costs of the index admission ($14,500) by about 12.4%. [1] These costs contribute billions of dollars to the national healthcare spending each year and represent a significant component of preventable hospital expenditure under value-based reimbursement models. Beyond their financial burden, readmissions indicate care fragmentation, poor discharge planning, medication mismanagement, or inadequate follow-up during the post-discharge period [4, 5]. The literature on hospital readmissions emphasizes that reducing avoidable readmissions requires more than standardized discharge procedures or uniform post discharge outreach. As healthcare organizations operate under value based payment models, there is growing emphasis on using data to identify where risk is concentrated and where limited resources can be deployed most effectively to improve outcomes.

Reminder Systems as a Common Intervention

The post-discharge period is a particularly vulnerable phase of care, during which patients face an elevated risk of medication errors, care discontinuation, and unmet clinical needs [4]. In a foundational study of nearly 12 millions Medicare beneficiaries, [6] found that one in five patients was readmitted within 30 days of discharge, and that half of nonsurgical patients were hospitalized without having seen an outpatient physician for follow up. Inadequate medication reconciliation, fragmented communication between outpatient providers and patients, and gaps in discharge education all contribute to preventable admissions [4, 5]. Post-discharge reminder systems, particularly follow-up telephone calls, have been widely studied as mechanisms to reduce preventable hospital readmissions and improve care transitions. In a comprehensive systematic review and meta-analysis of randomized controlled trials, [7] examined a broad range of interventions to reduce readmission, including telephone follow-up and transitional care programs. Evidence from randomized trials suggests that post-discharge interventions can reduce 30-day readmission rates, though the magnitude and consistency of these effects vary substantially. Similarly, [8] evaluated a large-scale post-discharge telephonic outreach program involving over 30,000 patients enrolled in chronic disease management programs. Their findings showed that patients who received a follow-up call within 14 days of discharge were significantly less likely to be readmitted within 30 days than those who did not receive a call. The study also revealed important patterns to note: the median time to readmission was 11 days, with nearly one-third of readmissions occurring within the first 7 days and over half occurring within the first 14 days, underscoring the operational importance of timely post-discharge outreach.

Heterogeneity in Readmission Risk and Intervention Effectiveness

Despite promising evidence that post-discharge interventions can reduce readmissions, the magnitude and consistency of these effects vary substantially across studies. [7] reported marked heterogeneity in outcomes, with more intensive and tailored interventions generally producing stronger effects than simpler, uniform approaches. This variability suggests meaningful differences in baseline patient risk and responsiveness to follow-up support. When reminder systems are applied uniformly across heterogeneous patient populations, their impact may be diluted: low-risk patients may receive unnecessary outreach, while higher-risk individuals may require more sustained or time-sensitive intervention.

The broader literature on readmission risk factors reinforces this conclusion. Patients who experience readmission consistently exhibit greater clinical complexity, including longer hospital stays, higher numbers of diagnosed conditions, and greater prior healthcare utilization than those who are not readmitted [9, 10]. Several validated risk prediction tools exist for hospitalized diabetic patients, specifically the Diabetes Early Readmission Risk Indicator, which used 13 clinical and sociodemographic variables to estimate 30 day readmission risk and has been externally validated across multiple academic medical centers [11, 12]. However, these tools focus on identifying high risk patients rather than translating risk into structured decision rules under capacity constraints, leaving an open operational question of how outreach intensity and timing should be calibrated to predicted risk. These findings indicate substantial heterogeneity in post-discharge risk across patient populations. In addition to variation in patient characteristics, studies document uneven implementation of follow-up outreach across groups. Not all discharged patients can be reached in a timely manner due to incorrect contact information, delayed discharge notifications, and competing organizational priorities [8]. Operational constraints and incomplete clinical information at discharge further complicate efforts to identify and support high-risk individuals [9]. Together, these patterns suggest that one-size-fits-all follow-up strategies may dilute impact when applied across heterogeneous populations. Taken together, the literature indicates that the effectiveness of reminder systems depends not only on whether outreach occurs, but on how well intervention intensity aligns with patient-level risk. These findings motivate the development of risk-stratified approaches that allocate follow-up resources in a targeted and operationally efficient manner.

Operational Constraints and the Case for Targeted Decision Making Using AI

Despite evidence supporting the effectiveness of follow-up calls, both the literature and current practice reveal persistent operational challenges. [8] demonstrated that not all discharged patients could be reached in a timely manner due to incorrect contact information, delayed discharge notifications, or competing organizational priorities. Staffing limitations further constrain the number of calls that can be delivered, forcing healthcare organizations to implicitly prioritize which patients receive outreach and when. In many settings, these prioritization decisions rely on coarse heuristics or clinical judgment rather than systematic, data-driven approaches. At its core, the challenge is not whether follow-up calls are beneficial, but how limited outreach capacity should be allocated across heterogeneous patient populations. Hospitals operate under fixed staffing, budget, and time constraints, requiring trade-offs between breadth and intensity of intervention. In this setting, the central managerial question becomes how to deploy scarce post-discharge resources in a way that maximizes clinical impact and financial performance.

Recent advances in AI and predictive analytics offer a pathway to address these limitations. A growing body of research demonstrates that AI-enabled predictive analytics can improve risk stratification, support earlier intervention, and enhance patient outcomes by informing more precise clinical and operational decisions [13]. Rather than replacing existing reminder modalities, such systems augment them by estimating patients' relative likelihood of readmission at discharge, enabling targeted allocation of follow-up resources toward those most likely to benefit. Importantly, their operational value lies not in perfect prediction but in ranking patients by relative risk. While follow-up calls can reduce readmissions on average, their impact is diluted when applied uniformly without regard to heterogeneity in risk or capacity constraints [7, 8]. This perspective reframes reminder systems as adaptive decision-making tools that translate AI-generated insights into targeted, scalable post-discharge interventions.

Building on its role in clinical decision support, AI is increasingly influencing diagnostics assessment and border health system management practices. Nonetheless, ethical and organizational issues cannot be overlooked making responsible AI deployment challenging. For example, [14] evaluated the performance of AI chatbots in preliminary diagnosis of maxillofacial pathologies, and found at times it could be accurate, but a provider is very much needed to confirm all of the results. It underscored the promise yet the existing limitations AI currently has. Although chatbots may be able to assist with initial screening or patient education, this reinforces the need for clinical oversight. At the organizational level, [15] identified operational opportunities and ethical challenges that exist with AI implementation in nursing management, noting better workload allocations and staffing optimization, yet having big concerns with bias and accountability.

Case Study: Reducing 30-Day Readmission Rates for Diabetes Patients

This essay builds directly upon a team project completed by a team of Executive MBA students in Healthcare Administration as part of the Healthcare Analytics and Quality course. The project focused on designing a data-driven framework to design a follow-up reminder system to reduce 30-day hospital readmissions among patients with diabetes by targeting those identified as high-risk for readmissions.

The Details And The Findings Of The Underlying Project

The original case study explored how predictive modeling and simulation could inform operational decision-making under realistic capacity constraints. Rather than pursuing highly complex machine learning approaches, the project intentionally emphasized interpretability and practical implementation, reflecting real-world hospital discharge workflows. The central question was not simply whether readmissions could be predicted, but whether even modest predictive signals could meaningfully improve the allocation of limited post-discharge resources.

Using a subset of a publicly available dataset of 30,529 hospitalized diabetic patients, the team developed a forward stepwise logistic regression model to estimate individual readmission risk at discharge. The model identified five clinically plausible predictors of readmission risk, which we summarize and operationalize in Table 1. These factors are extended length of stay, greater comorbidity burden as measured by number of diagnoses, more complex diabetes medication regimens, prior inpatient utilization, and missing A1C documentation. Although predictive discrimination was modest (AUC = 0.554), the resulting risk scores provided meaningful relative ranking of patients. This reinforced an important operational insight: in readmission settings characterized by substantial unobserved heterogeneity, the value of prediction lies in risk stratification rather than precise individual-level accuracy.

Table 1. Operational Risk Screening Signals at Discharge

Risk Signal	Why It Matters Operationally	Availability at Discharge	Example Trigger Rule
Extended Length of Stay	Prolonged hospitalization often reflects greater illness severity and care complexity	Known at time of discharge	Top quartile of length of stay
High Number of Diagnoses	Greater comorbidity burden increases risk of post-discharge complications	Recorded in encounter summary	Above median number of diagnoses
Multiple Diabetes Medications	Complex medication regimens increase risk of non-adherence or dosing errors	Medication reconciliation complete at discharge	2 or more diabetes-related medications
Prior Inpatient Utilization	Recent hospital use indicates unstable health trajectory	Historical utilization data available in EHR	1 or more inpatient admission(s) in prior year
Missing A1C Documentation	May signal fragmented care or incomplete chronic disease monitoring	Lab data available at discharge	A1C not recorded during encounter

From an implementation perspective, the model results translate into a small set of operationally observable discharge signals that can support rapid risk screening. Patients with longer length of stay, higher comorbidity burden, more complex diabetes medication regimens, recent inpatient utilization, and missing A1C documentation consistently ranked higher on predicted readmission risk. Importantly, each of these signals is available at or before discharge, allowing risk-based outreach decisions to be embedded directly into existing discharge workflows. Rather than requiring complex algorithmic interpretation at the bedside, these variables can serve as practical triggers for targeted follow-up when outreach capacity is constrained.

The simulation component of the course project demonstrated how these risk scores could be embedded within prescriptive decision frameworks to evaluate alternative follow-up strategies. By modeling time-based readmission dynamics and capacity-constrained outreach policies, we found that interventions targeting the patients with the highest (i.e., top 20%) risk scores could reduce 30-day readmissions from 11.77 percent to 9.15 percent while lowering total combined costs under reasonable assumptions. These results directly inform the present submission by providing empirical evidence that AI-assisted reminder systems do not require high predictive accuracy to generate operational value. Instead, their impact emerges from integrating modest risk stratification with structured, time-sensitive intervention design.

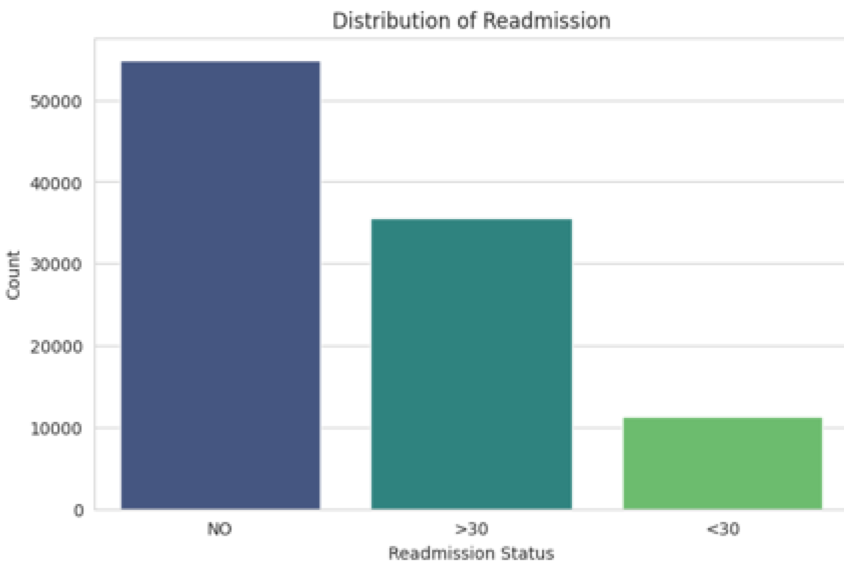
The Replication Study

To generalize the findings of the class project, we replicate the same analysis on the entire data set of 30,529 hospitalized diabetic patients using more advanced predictive models. We also challenge several simplifying assumptions that underlie the simulation phase of the original project and demonstrate the robustness of the main findings regarding the effectiveness of targeted intervention.

Data Set And Descriptive Statistics

The dataset used in this analysis is derived from the UCI "Diabetes 130-US Hospitals" database [1], which contains administrative and clinical information from inpatient encounters across 130 U.S. hospitals between 1999 and 2008. The unit of analysis is the individual inpatient encounter rather than a unique patient, meaning that patients with multiple admissions during the study period may appear more than once in the dataset. This structure reflects real-world hospital operations, where discharge decisions are made at the encounter level.

Figure 1. Outcome variable of interest in the data set



[1] The data set is publicly available at which is publicly available at https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008# .

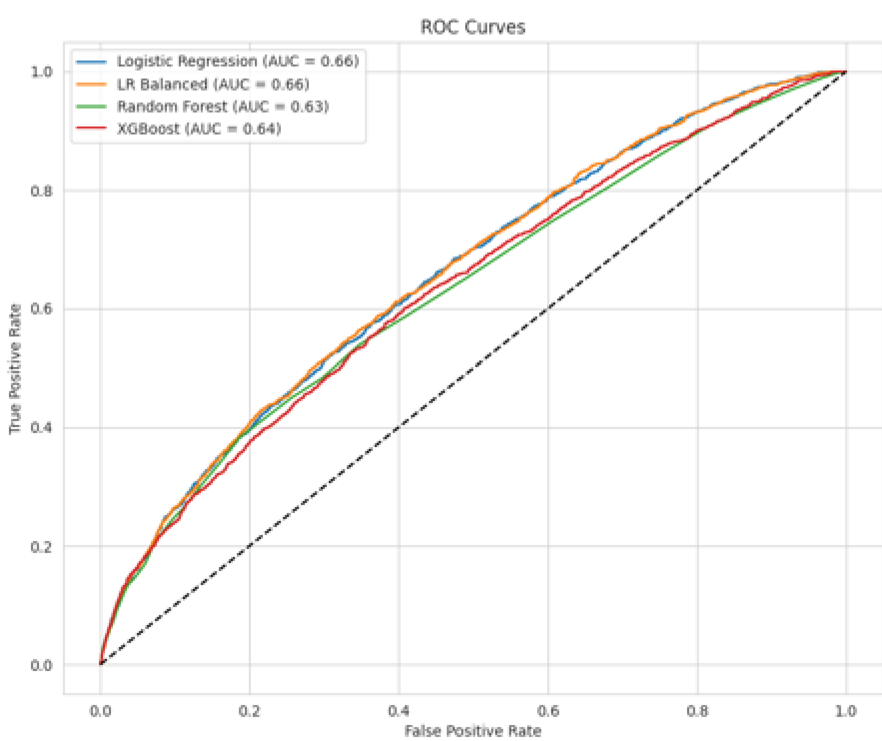
The dataset includes demographic variables (age category, gender, race), utilization measures (length of stay, number of diagnoses, number of procedures, prior inpatient visits), and indicators of diabetes management (medication status, insulin use, and A1C testing). The primary outcome of interest is 30-day hospital readmission. In the original dataset, readmission status is categorized as “readmitted within 30 days,” “readmitted after 30 days,” or “no readmission.” As illustrated in Figure 1, approximately 11.3% of patients were readmitted within 30 days of discharge, which indicates that while readmissions are not the majority outcome, they represent a meaningful operational concern.

Predictive Models

In the next phase of the analysis, we built predictive models to estimate each patient’s likelihood of being readmitted within 30 days of discharge. We compared several modeling approaches, including logistic regression, a balanced version of logistic regression, random forests, and gradient boosted models. Because readmissions are relatively rare, we paid particular attention to whether the models could meaningfully distinguish higher risk patients from lower risk patients, rather than focusing only on overall accuracy.

Figure 2 presents the receiver operating characteristic (ROC) curves comparing logistic regression, class-balanced logistic regression, random forest, and gradient boosted (XGBoost) models. The standard logistic regression and class-balanced logistic regression models achieved the highest performance, with an area under the curve (AUC) of approximately 0.66. Random forest and XGBoost models yielded slightly lower AUC values (0.63 and 0.64, respectively).

Figure 2. ROC Curves and the AUC values for all four considered prediction models



We ultimately selected the balanced logistic regression model as our primary predictive model. While its predictive discrimination was similar to more complex models, it offers two advantages that matter in practice: it performs well at identifying higher risk patients in an imbalanced setting, and it produces outputs that are easy to interpret as we summarize in Table 2.

Table 2. Statistically significant features of the balanced logistic regression model we use to compute risk scores

FACTORS INCREASING READMISSION RISK		
Feature	**Coefficient**	**Interpretation**
# prior inpatient visits	0.246	Frequent hospitalizations increase readmission risk
# prior ER visits	0.054	Frequent emergency visits increase readmission risk
# diagnoses at encounter	0.058	Many diagnoses indicate higher clinical complexity
Heart failure (ICD-9: 428)	0.270	Cardiac comorbidity increases readmission risk
Diabetes medication prescribed	0.148	Proxy for how chronic and severe the disease is
No medication change made	0.194	Adjustment may have been warranted but not acted on
Caucasian race	0.125	Likely proxy for care access disparities
FACTORS REDUCING READMISSION RISK		
Feature	**Coefficient**	**Interpretation**
No insulin prescribed	-0.399	Stable glycemic management at discharge
Insulin dose unchanged	-0.336	No adjustment needed; controlled at discharge
No glipizide prescribed	-0.381	Stable medication regimen
Glipizide dose unchanged	-0.257	No adjustment needed
Pneumonia (ICD-9: 486)	-0.310	Different readmission trajectory
Respiratory symptoms (ICD-9: 786)	-0.289	Different readmission trajectory

Although our predictive model does not achieve high individual-level predictive precision, it demonstrates meaningful capacity to rank patients by relative readmission risk. In operational settings where the goal is to prioritize limited outreach resources rather than generate perfect predictions, this level of discrimination is sufficient to support targeted intervention policies. Importantly, the comparable performance of simpler models reinforces the managerial value of interpretability, as discharge teams can more easily understand and implement risk-based decision rules derived from logistic regression.

Simulation Models To Prescribe Decisions

In this stage of the analysis, we used the predicted readmission probabilities from the balanced logistic regression model as patient level "risk scores." These scores allow us to rank patients by relative risk at discharge, which is the key input needed to decide who should be prioritized for follow-up outreach. Using these risk scores, we build simulation models to assess the percentage reduction in 30-day readmission rates at various intervention coverage levels.

The specific mechanics of the simulation model of the follow-up calls involve re-calibrating each patient type's 30-day predicted readmission probability, i.e., risk score, if they are included among those to receive a follow-up call on day 15 of the 30-day window. In such cases, that patient's 30-day readmission probability is revised as their (remaining) 15-day readmission probability, an idea we refer to as "resetting the clock." We simulated intervention policies on a sample of 10,000 discharged patients whose features are drawn from the data set we described in subsection 3.2.1, as one simulation run, and reported statistics averaged over 1,000 simulations in Figure 3. The vertical axis denotes the reduction in readmission rates (in percentage points) relative to zero coverage, i.e., when no patients receive follow-up calls. The horizontal axis indicates intervention coverage, i.e., percentage of patients with the highest predicted risk scores whose readmission clocks are reset on day 15 after discharge.

Figure 3. The trade-off between intervention coverage and expected reduction in 30-day readmissions

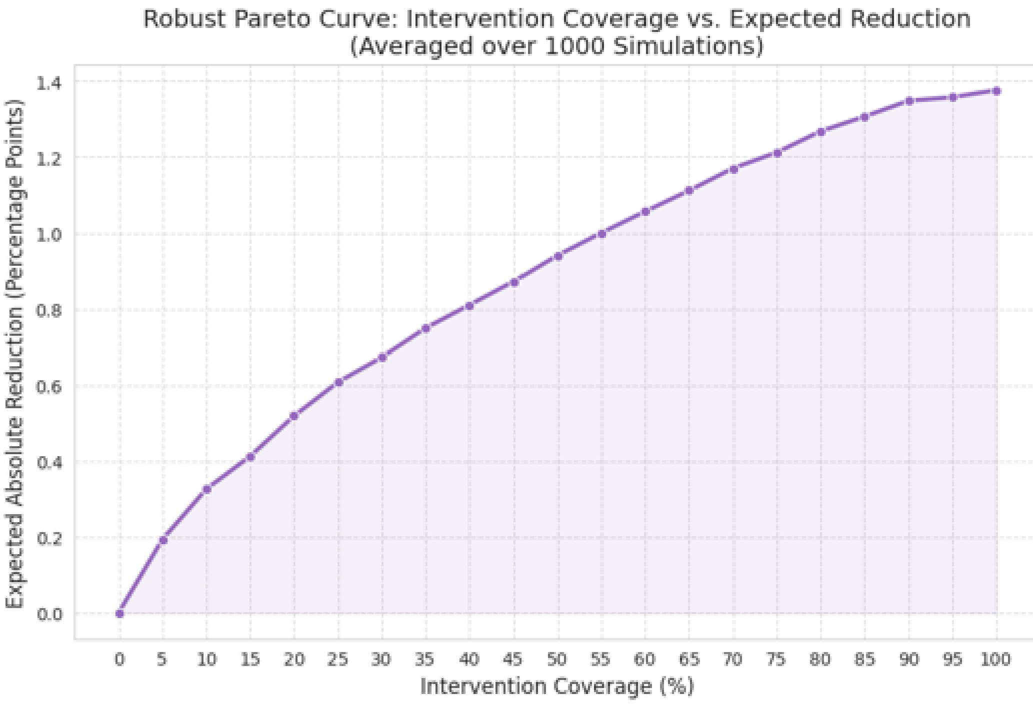


Figure 3 shows that if reminders are provided to 100% of discharged patients, i.e., uniform reminders, the model estimates a maximum absolute reduction of approximately 1.4 percentage points. However, the majority of this improvement can be achieved with substantially lower coverage. Targeting only the top 10% highest-risk patients captures roughly 25% of the total achievable reduction, yielding an estimated 0.35 percentage-point improvement. Expanding coverage to the top 20% of high-risk patients captures approximately 40% of the total potential reduction, corresponding to a 0.56 percentage-point improvement. Beyond this threshold, additional coverage yields progressively smaller marginal gains, indicating clear diminishing returns.

This pattern demonstrates that carefully targeted reminder strategies can approximate the benefits of uniform reminders at a fraction of the intervention volume, reinforcing the operational value of risk-based prioritization under capacity constraints. This is because the largest gains come from focusing on a relatively small group of high risk patients. Expanding the intervention to include lower-risk patients still improves outcomes, but with diminishing returns. This finding reinforces the practical value of using risk scores to guide limited resources. In other words, a targeted reminder strategy can capture much of the benefit of universal outreach at a fraction of the cost and effort.

Robustness Checks

In this subsection, we relax some of the underlying assumptions in the original replication study to demonstrate that our main finding, i.e., targeted data-driven intervention generates disproportional benefits in reducing 30-day readmissions, is robust. First, we note that we obtain the Pareto curve in Figure 3 by assuming that a follow-up intervention on day 15 after discharge would be successful with 25% probability. This implies a linear hazard function according to which failures, i.e., readmissions take place. The robustness checks adopt a more structurally flexible model, wherein each patient's readmission risk is characterized by a Weibull survival function calibrated to their individual predicted readmission probability. This model still captures the original replication study, i.e., a linear hazard function (Weibull k=2) with a full clock reset to day 0 when the follow-up intervention takes place. In what follows, we assess sensitivity in two robustness studies, which we label RC1 (Partial Clock Reset) and RC2 (Nonlinear Hazard).

RC1: Sensitivity to Intervention Effectiveness (Partial Clock Reset)

The original replication study assumes the intervention fully resets the readmission hazard clock to day 0, the most optimistic scenario. To assess the sensitivity to this assumption, we vary how far back the intervention on day 15 resets the clock, sweeping from day 0, 5, 10 and 15, representing a full reset, two partial resets, and no reset (i,e., the intervention has no effect on the hazard trajectory).

Figure 4. Robustness of the main finding with limited effectiveness of follow-up interventions on day 15

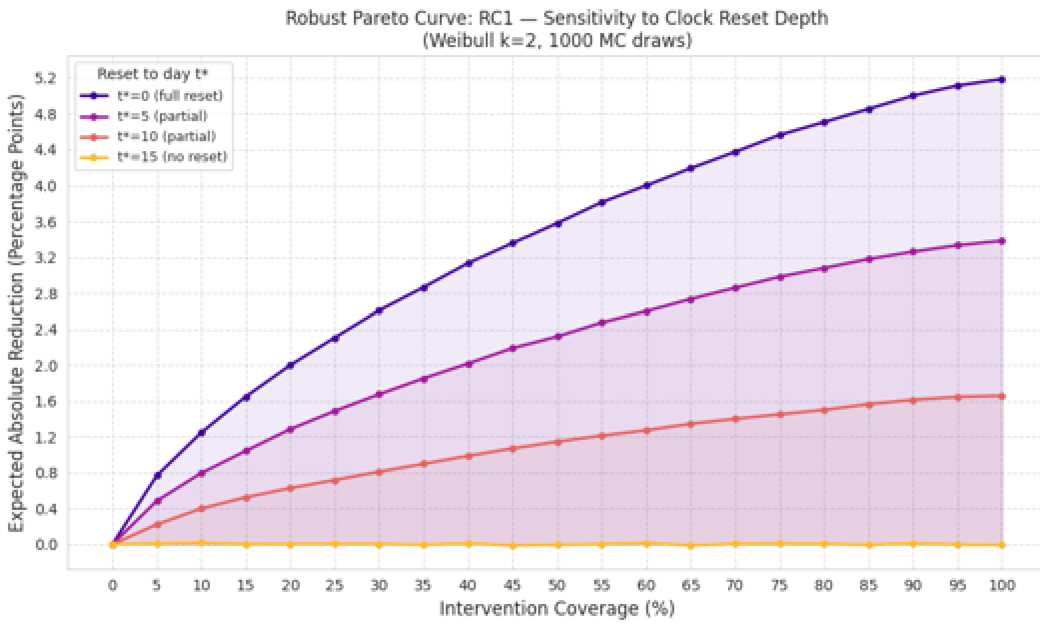
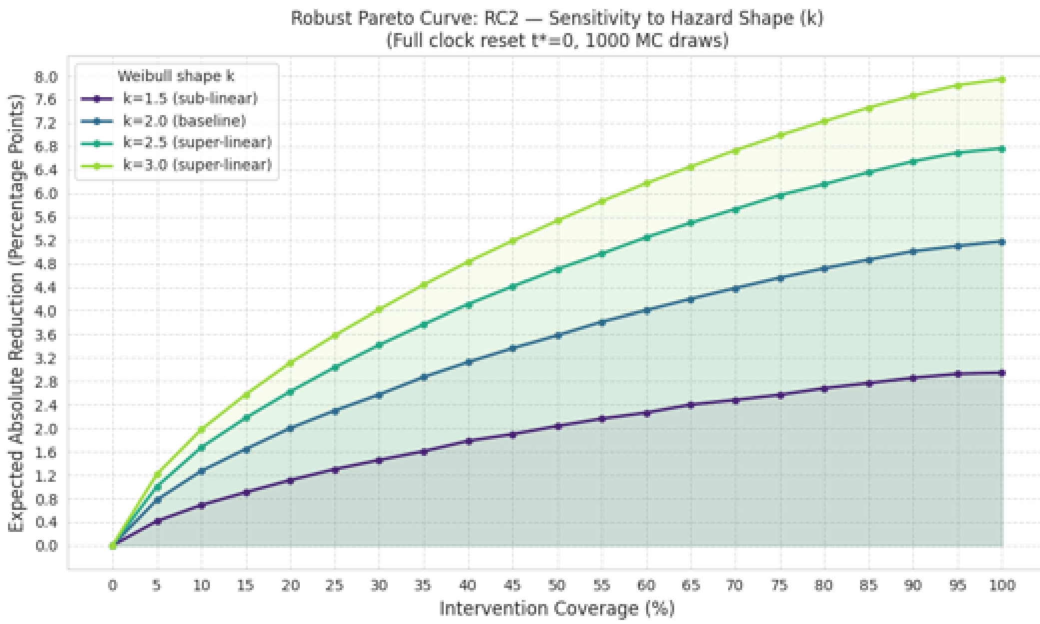


Figure 4 shows that across all reset depths, the Pareto curves exhibit consistent decreasing marginal returns to intervention coverage levels, confirming that the qualitative policy finding is robust. The absolute magnitude of reduction scales with reset depth: a full reset (to day 0) yields up to 5.2pp (percentage points) reduction at 100% coverage, while no reset (to day 15) yields zero. The partial reset cases (to day 5 and day 10) bracket a plausible range of realistic interventions. Importantly, across all three non-trivial cases (reset to day 0, 5 and 10), targeting the top 10-20% of highest-risk patients consistently captures approximately 25-50% of the maximum achievable reduction, confirming that the key policy recommendation is robust to the assumed depth of intervention effectiveness.

RC2: Sensitivity to Hazard Shape (Nonlinear Hazard)

The original replication study assumes a linearly increasing hazard function (Weibull shape k=2). To test whether this assumption drives the results, we vary the shape parameter k from 1.5 to 3.0, where k<2 implies a sub-linear hazard (risk decelerates over time, concentrating in day 1-15) and k>2 implies a super-linear hazard (risk accelerates, concentrating in days 16-30). We report comparative results in Figure 5.

Figure 5. Robustness of the main finding with non-linear hazard functions



Across all values of *k*, the Pareto curves retain the same concave shape, confirming the robustness of the qualitative finding. The magnitude of reduction varies meaningfully with k: super-linear hazards (*k*=2.5, 3.0) yield larger reductions to up to 8pp at 100% coverage for *k*=3.0, while the sub-linear case (*k*=1.5) yields the smallest benefit. This pattern is intuitive: when the hazard is super-linear, readmission risk concentrates in the later days of the 30-day window, a clock reset at day 15 intercepts the highest-risk period, making the intervention more effective. Conversely, when the hazard is sub-linear, most risk has already materialized before day 15, leaving less for the intervention to prevent. Critically, across all four curves, targeting the top 10–20% of highest-risk patients consistently captures approximately 25% to 40% of the maximum achievable reduction, confirming that the key policy recommendation is robust to the assumed shape of the readmission hazard.

Conclusion

This study integrates descriptive, predictive, and prescriptive analytics to evaluate the operational effectiveness of AI-assisted reminder systems in reducing 30-day hospital readmissions among patients with diabetes. Rather than treating prediction as an end in itself, we embed risk stratification within a structured decision framework that explicitly accounts for limited outreach capacity and real-world discharge workflows.

Across simulations, targeted intervention policies consistently outperform uniform reminder strategies when evaluated relative to resource intensity. Specifically, directing follow-up outreach to the top 10% to 20% highest-risk patients captures approximately 25% to 40% of the total achievable reduction in 30-day readmissions, while requiring substantially fewer interventions than universal outreach. Beyond this range, marginal gains diminish rapidly, indicating that expanding outreach to lower-risk patients yields progressively smaller returns per unit of effort.

These findings suggest that the operational value of AI in healthcare lies not in perfect prediction, but in enabling disciplined prioritization under constraints. For healthcare leaders, the actionable implication is clear: embed interpretable risk ranking into discharge workflows, define tiered outreach protocols based on risk thresholds, and continuously monitor performance, fairness, and model drift over time. By translating modest predictive signals into structured allocation rules, hospitals can improve patient outcomes and protect financial performance under value-based reimbursement without expanding blanket outreach programs.

References

1. Jiang, H. J., & Hensche, M. K. (2023). *Characteristics of 30-day all-cause hospital readmissions, 2016-2020* (HCUP Statistical Brief #304). Agency for Healthcare Research and Quality. https://hcup-us.ahrq.gov/reports/statbriefs/sb304-readmissions-2016-2020.jsp
2. Centers for Medicare & Medicaid Services. (2024). *Hospital Readmissions Reduction Program* (HRRP). U.S. Department of Health and Human Services. https://www.cms.gov/medicare/payment/prospective-payment-systems/acute-inpatient-pps/hospital-readmissions-reduction-program-hrrp
3. Medicare Payment Advisory Commission. (2018). Mandated report*: The effects of the Hospital Readmissions Reduction Program*. In Report to the Congress: Medicare and the health care delivery system (Chapter 1). https://www.medpac.gov/wp-content/uploads/import_data/scrape_files/docs/default-source/reports/jun18_ch1_medpacreport_rev_nov2019_v2_note_sec.pdf
4. Forster, A. J., Murff, H. J., Peterson, J. F., Gandhi, T. K., & Bates, D. W. (2003). The incidence and severity of adverse events affecting patients after discharge from the hospital. *Annals of Internal Medicine, 138*(3), 161–167.

https://doi.org/10.7326/0003-4819-138-3-200302040-00007

5. Kripalani, S., LeFevre, F., Phillips, C. O., Williams, M. V., Basaviah, P., & Baker, D. W. (2007). Deficits in communication and information transfer between hospital-based and primary care physicians: Implications for patient safety and continuity of care. *JAMA, 297*(8), 831–841. https://doi.org/10.1001/jama.297.8.831
6. Jencks, S. F., Williams, M. V., & Coleman, E. A. (2009). Rehospitalizations among patients in the Medicare fee-for-service program. *New England Journal of Medicine, 360*(14), 1418–1428. https://doi.org/10.1056/NEJMsa0803563
7. Leppin, A. L., Gionfriddo, M. R., Kessler, M., Brito, J. P., Mair, F. S., Gallacher, K., Wang, Z., Erwin, P. J., Sylvester, T., Boehmer, K., Ting, H. H., & Montori, V. M. (2014). Preventing 30-day hospital readmissions: A systematic review and meta-analysis of randomized trials. *JAMA Internal Medicine, 174*(7), 1095–1107. https://pmc.ncbi.nlm.nih.gov/articles/PMC4249925/
8. Harrison, P. L., Hara, P. A., Pope, J. E., Young, M. C., & Rula, E. Y. (2011). The impact of postdischarge telephonic follow-up on hospital readmissions. *Population Health Management, 14*(1), 27–32. https://pmc.ncbi.nlm.nih.gov/articles/PMC3128446/
9. Kansagara, D., Englander, H., Salanitro, A., Kagen, D., Theobald, C., Freeman, M., & Kripalani, S. (2011). Risk prediction models for hospital readmission: A systematic review. *JAMA, 306*(15), 1688–1698. https://pubmed.ncbi.nlm.nih.gov/22009101/
10. Zhang, S., Cardona, K., Liu, M., Davis, M., Hayward, R., Heisler, M., & Singh, K. (2020). Assessing the impact of social determinants of health on predictive models for potentially avoidable 30-day readmission or death. *PLOS ONE*, *15*(6), e0235064. – https://doi.org/10.1371/journal.pone.0235064
11. Rubin, D. J., Handorf, E. A., Golden, S. H., Nelson, D. B., McDonnell, M. E., & Zhao, H. (2016). Development and validation of a novel tool to predict hospital readmission risk among patients with diabetes. *Endocrine Practice, 22*(10), 1204–1215. https://doi.org/10.4158/E161391.OR
12. Rubin, D. J., Recco, D., Turchin, A., Zhao, H., & Golden, S. H. (2018). External validation of the Diabetes Early Re-admission Risk Indicator (DERRI™). *Endocrine Practice, 24*(6), 527–541. https://doi.org/10.4158/EP-2018-0035
13. Sattar, H., Dixon, D., Moros, N., Kesireddy, S. R., Ahsan, H., Lakkimsetti, M., Doshi, D., Fatima, M., Sadhu, K., & Hassan, M. J. (2024). *Unveiling the influence of AI predictive*

analytics on patient outcomes: A comprehensive narrative review. Cureus, 16(5), e59954. https://doi.org/10.7759/cureus.59954
14. Guler, R., & Yalcin, E. (2025). Performance of AI chatbots in preliminary diagnosis of maxillofacial pathologies. *Medical Science Monitor, 31*, e949076. – https://pmc.ncbi.nlm.nih.gov/articles/PMC12257980/
15. Katebi, A., et al. (2025). Artificial intelligence and nursing management: Opportunities, challenges, and ethical considerations—A scoping review. *Journal of Nursing Management*. Advance online publication. https://pubmed.ncbi.nlm.nih.gov/40771284/

[1] The data set is publicly available at which is publicly available at https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008# .