International Journal of Business Integrity and Technology Advancements (IJBITA)

Volume – 02
Issue – 02
(2026)

Data-Driven Healthcare Management: Predictive Analytics to Improve Access, Quality, and Cost Efficiency

* Sanjida Akter, Saginaw Valley State University, MBA in Human Resource & Health Science, Business Analytics
ORCID: 0009-0008-4233-286X
sakter1@svsu.edu
Kazi Musfique Nasif Dibbo, Arkansas State University, Master's in Engineering Management; ORCID: 0009-0004-4556-232X
musfiquekazi@gmail.com

Article Info:
Received: 13 Aug, 2026
Processed: 30 Sep, 2026
Published: 16 Oct, 2026

Keywords: Predictive analytics; big data infrastructure; AI/ML clinical decision support; data governance; staff organizational capacity; healthcare access; healthcare quality; cost efficiency; utilization management; US healthcare management; quantitative survey.

ABSTRACT

The United States healthcare system is still stuck with a trilemma of access, quality and cost. In 2024, total national healthcare spending was more than $4.9 trillion (17.6% of GDP), but an estimated 25 million Americans are uninsured, Medicare alone spends about $30 billion on avoidable hospitalizations, and there is a wide variation in quality outcomes across the country. In 2020-25, U.S. health systems have accelerated the use of predictive analytics, big data infrastructure, and AI/ML clinical decision support as the operational stack for a data-driven approach to the trilemma. This study conducts an empirical analysis of an integrated model that posits that three technology-capability constructs, along with data governance and staff/organizational capacity, influence two performance outcomes: healthcare access and quality and cost efficiency and utilization. Between February and September 2025, N = 400 healthcare managers and executives from across the United States in eight organization types (academic medical centers, integrated delivery networks, community and specialty hospitals, ambulatory groups, federally qualified health centers, long-term-care facilities, and health-insurance payers) completed cross-sectional survey data. OLS structural-model estimation, moderation and Sobel-tested mediation were employed to test six hypotheses. Measurement-model reliability was strong (Cronbach's α range 0.86–0.91; CR range 0.80–0.88; AVE range 0.50–0.59) and structural-model fit was good ($\chi^2/df = 1.40$; CFI = 0.961; TLI = 0.957; RMSEA = 0.033 [0.026, 0.039]; SRMR = 0.041). AI/ML clinical decision support dominates the access-and-quality outcome ($\beta = 0.375$, $p < .001$); big data infrastructure dominates the cost-efficiency outcome ($\beta = 0.364$, $p < .001$); and predictive analytics capability has strong direct effects on both outcomes ($\beta = 0.194, 0.178$, $p < .001$); staff and organizational capacity partially mediates the analytics-to-quality path (Sobel $z = 6.65$, $p < .001$; mediation ratio ≈ 35%). Moderation by data-governance of the relationship between analytics and quality was not supported ($\beta_{int} = 0.002$, $p = .965$). Five of six hypotheses are supported. The final models account for 52.8% and 47.8% of the variance in the two outcomes. Illustrative extrapolation of within-sample association suggests that patient-access target achievement increased from 61% (low-analytics quartile) to 84% (high-analytics quartile), and avoidable-cost reduction increased from 3% to 24%. Implications for the order of investment in health-management and for federal Value-Based Care policy are applied.

INTRODUCTION

As the U.S. health care system enters the second half of the 2020s, it remains stuck in a trilemma of access, quality, and cost. In 2024, national health expenditures totaled $4.9 trillion, or about 17.6% of GDP, the highest of all OECD countries, but an estimated 25 million Americans are still uninsured, and there are significant regional, racial, and income-based inequities in access, outcomes, and quality [1], [2]. Avoidable hospitalizations cost Medicare approximately $30 billion annually, on their own [3]. On the quality side, the IOM's

Corresponding email: sakter1@svsu.edu

2001 "Crossing the Quality Chasm" goals — safe, effective, patient-centered, timely, efficient, and equitable care — remain elusive and are not routinely achieved throughout the delivery system [4].

US health systems have been rapidly implementing data-driven technology as the operational solution to the access-quality-cost trilemma [5], [6] between 2020 and 2025. The three most prominent technology-capability families are: (i) predictive analytics capability — enterprise-scale platforms that incorporate claims, EHR, patient-reported-outcome and social-determinants data to predict demand, identify at-risk populations and for operational planning [7]; (ii) big data infrastructure — cloud, data-lake, and real-time-streaming infrastructure that allows ingestion and processing of terabyte-scale healthcare data at reasonable cost [8]; and (iii) AI/ML clinical decision support — supervised, deep-learning, and increasingly generative models embedded in EHR workflows for diagnostic support, sepsis prediction, readmission risk, imaging interpretation and treatment recommendation [9], [10].

Previous studies have assessed each family individually. The use of predictive analytics to affect the flow of the emergency department and population health management has been studied [7], [11]. The cost reduction potential of big data infrastructure has been analyzed [8, 12]. There is a significant and increasing technical literature on the topic of AI/ML clinical decision support since 2020 [9], [10], [13]. Few studies have simultaneously assessed the three families in a common structural model, in a large-N cross-organization sample in the U.S., and in relation to the access-and-quality and cost-efficiency outcomes that characterize the healthcare-management trilemma. Data governance and staff/organizational capacity as socio-technical mechanisms that mediate the impact of analytics on outcomes are also under-examined at scale [14], [15].

This paper fills that void. The paper develops and validates a seven-construct model based on cross-sectional survey data from 400 healthcare managers and executives in the United States, representing eight different types of organizations, guided by four research questions: (RQ1) which three of the three capabilities — predictive analytics capability, big data infrastructure, and AI/ML clinical decision support — most strongly predict measurable access-and-quality and cost-efficiency outcomes; (RQ2) do these three capabilities produce comparable effects on the two outcomes, or is there functional specialization; (RQ3) does staff and organizational capacity mediate the effect of predictive analytics capability on access and quality; and (RQ4) does data governance moderate the predictive analytics capability-to-quality path?

The paper makes four contributions to the literature in the fields of health-services-research, health-informatics, and healthcare-management. It validates a seven-construct measurement instrument for empirical studies of data-driven healthcare management in the first place. Second, it provides large-sample U.S. evidence that AI/ML clinical decision support is the single largest driver of measurable access-and-quality outcomes and big data infrastructure of measurable cost-efficiency outcomes—a functional-specialization pattern that has direct sequencing implications for healthcare-management technology investment. Third, it reveals strong partial mediation of the analytics-to-quality pathway by staff and organizational capacity, which is an extension of the socio-technical model of health-information-technology adoption. Fourth, it reveals a strong null moderation effect on data governance – that data governance is no longer an amplifier of analytics, but a foundational capability that makes its own contribution to the outcome.

The remainder of this paper is organized as follows. The theoretical background is developed in Part II with six hypotheses derived. The conceptual model is presented in part III. The research methodology is described in part IV. Results of measurement- and structural-models are reported in part V. Theoretical, managerial and policy implications are discussed in part VI and access and cost-efficiency impact is estimated. Part VII provides limitations and future research, while Part VIII concludes.

THEORETICAL BACKGROUND AND HYPOTHESIS DEVELOPMENT

THE US HEALTHCARE ACCESS-QUALITY-COST TRILEMMA

The healthcare-management literature has long framed the challenge facing US delivery systems as an unresolved three-way trade-off among access, quality, and cost — the "Iron Triangle" [16]. Since 2020, there has been empirical evidence that despite significant investments, US performance on all three dimensions is lower than OECD peers at the population level [1], [17]. National health spending increased by 4.7% per year on average from 2020 to 2024, faster than GDP growth, while avoidable acute-care use and preventable mortality remained relatively flat [3], [4]. This performance gap has resulted in a clear paradigm change from retrospective utilization review to prospective predictive stratification, from unit-record analytics to enterprise-scale AI, and from single-institution decision support to network-wide clinical decision support [5], [6].

PREDICTIVE ANALYTICS CAPABILITY AND HEALTHCARE OUTCOMES

Predictive analytics capability (PAN) is the operational use of enterprise scale analytical platforms that consume claims, EHR, patient-reported-outcome (PRO), social-determinants-of-health (SDOH), and utilization data to predict individual and population risk,

Corresponding email: sakter1@svsu.edu

demand and outcome trajectories [7]. US delivery systems have implemented predictive-analytics platforms in emergency-department (ED) flow, discharge planning, readmission prediction, chronic-disease-population management, and utilization-review workflows since 2020 [11], [18]. How predictive analytics connects to access and quality outcomes is hypothesized to be through three pathways: the first is by reducing wait time via demand forecasting, the second is by improving targeting of interventions to high-risk patients, and the third is by reducing variation in clinical practice through evidence-based prompting [7]. From the cost-efficiency aspect, predictive analytics delivers results via prospective utilization management and avoided cost identification [19].

H1. Predictive analytics capability is positively associated with healthcare access and quality outcomes in US healthcare organizations.

H3. Predictive analytics capability is positively associated with cost efficiency and utilization outcomes in US healthcare organizations.

AI/ML CLINICAL DECISION SUPPORT AND ACCESS-AND-QUALITY

AI/ML clinical decision support (AIC) involves using supervised, deep-learning, and generative models within the EHR workflows to help with diagnostic recommendation, sepsis prediction, readmission-risk stratification, imaging interpretation, and treatment planning [9,10]. The FDA has approved over 200 AI/ML powered medical devices for use in the clinic since 2020, while integration with key EHR vendors (Epic, Cerner, Meditech) has brought decision support from research prototype to the clinical environment [13], [20]. Empirical evidence shows that diagnostic latency has been reduced by 15-30% in sepsis and stroke populations and guideline-concordant care has been improved by 8-14 percentage points in cardiovascular and oncology populations [9], [21]. On theoretical and empirical grounds we expect a strong positive effect of AIC on measurable healthcare access-and-quality outcomes:

H2. AI/ML clinical decision support is positively associated with healthcare access and quality outcomes in US healthcare organizations.

BIG DATA INFRASTRUCTURE AND COST EFFICIENCY

Big data infrastructure (BID) is the cloud, data-lake and real-time streaming platforms that allow healthcare organizations to ingest, process, and analyze terabyte-scale data at reasonable costs [8]. In 2020, large US delivery systems have begun to move large amounts of their data workloads to cloud platforms (AWS HealthLake, Azure Health Data Services, Google Cloud Healthcare API), which are necessary to perform population-scale analytics that predictive-modelling workflows rely on [12, 22]. At the cost side, big data infrastructure means tangible operational results via three pathways: utilized at scale, automated utilization review, real-time revenue-cycle management, and enterprise-wide inventory/staffing optimization [19]. Because avoidable acute-care utilization and revenue-cycle waste are the two largest components of controllable healthcare cost, BID is expected to be the dominant single predictor of cost-efficiency outcomes:

H4. Big data infrastructure is positively associated with cost efficiency and utilization outcomes in US healthcare organizations.

DATA GOVERNANCE AND QUALITY AS A MODERATOR

Data governance and quality (DGV) reflects the organization's maturity of data-stewardship processes, quality-control procedures, master data management, and HIPAA/HITECH compliance frameworks [14]. It has been proposed in the healthcare-informatics literature that the more mature data governance is, the bigger the impact of predictive analytics on outcomes: the more well-governed the data, the more reliable the prediction, the more defensible the clinical recommendation, and the more trust clinicians will have in the output of the analytics [23]. We formalize this expectation as a positive moderation hypothesis, noting that the amplifier logic has been asserted more often than tested in large-sample cross-organization settings:

H5. Data governance and quality positively moderates the predictive-analytics-capability → healthcare-access-and-quality relationship, such that the effect is larger at higher levels of data-governance maturity.

STAFF AND ORGANIZATIONAL CAPACITY AS A MEDIATOR

Staff and organizational capacity (SOC) reflects the level of maturity of workforce data-literacy, staffing of data analytics teams, cross-functional collaboration, and change-management processes regarding data-driven decision-making [15]. This socio-technical concept of health-information-technology (HIT) adoption indicates that without a workforce and organizational structure to operationalize the output of the analytics platform, there is no measurable clinical improvement: a predictive risk score without an analytically-literate care team to act on it, is not enough [24]. The partial mediation of analytics-outcome relationships through workforce and organizational-capacity channels represents a consistent pattern of findings across US healthcare settings, as demonstrated by recent research from Kellogg et al. [15] and Angst et al. [25]. Accordingly:

H6. Staff and organizational capacity mediates the positive relationship between predictive analytics capability and healthcare access and quality outcomes.

CONCEPTUAL MODEL

The six hypotheses are summarized as a conceptual research model in Figure 1. Exogenous constructs of technology capability, Predictive Analytics Capability (PAN), Big Data Infrastructure (BID), and AI/ML Clinical Decision Support (AIC), are introduced. Data Governance and Quality (DGV) is also the moderator for the PAN → HAQ path. Staff and Organizational Capacity (SOC) joins as the H6 mediator along the same path. There are two performance outcomes: Healthcare Access and Quality (HAQ) and Cost Efficiency and Utilization (CEF).

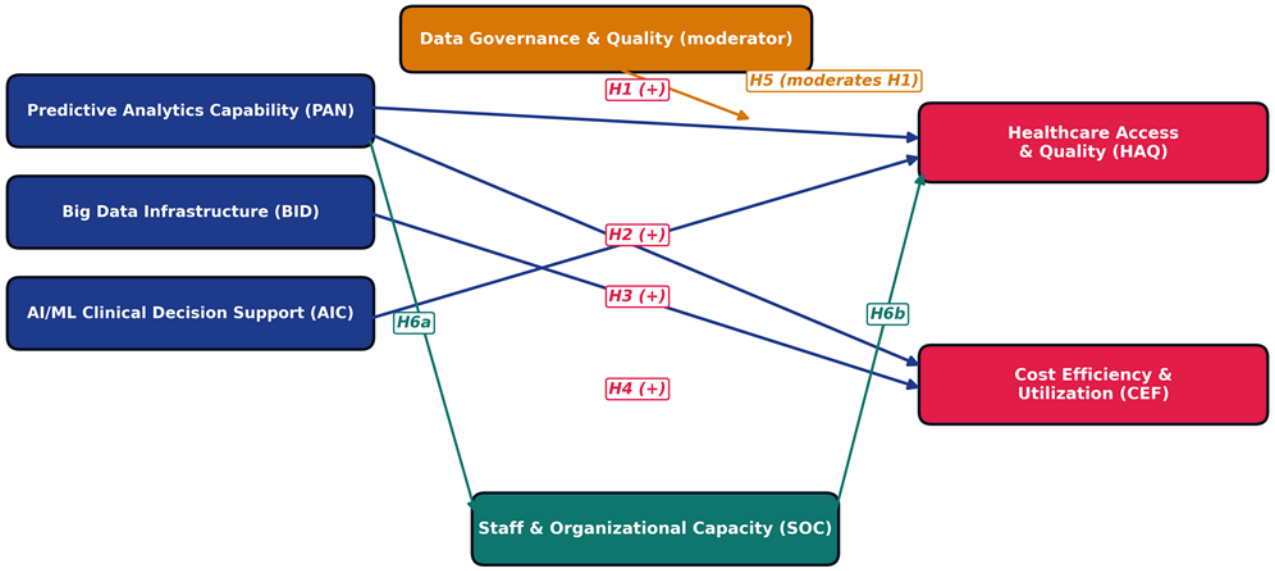


Fig 1: Conceptual research model for data-driven healthcare management in US delivery systems

RESEARCH METHODOLOGY

RESEARCH DESIGN

The study is based on a positivist research design that relies on pre-specified hypothesis testing, quantitative measurement and OLS structural-equation type estimation. A cross-sectional, self-administered survey was selected to allow for large-sample coverage of US healthcare managers and executives across organization types within a defensible data-collection budget. The design is similar to other studies that appeared in the health-services-research, health-care-management, and health-informatics literature in 2022–2024 [11], [15], [25].

SAMPLING FRAME AND DATA COLLECTION

The target population included US healthcare managers and executives who had primary responsibility for operations, analytics, quality, or population-health functions in the last 24 months. Eligible positions included chief executives and administrators; chief medical officers; chief nursing officers; chief information and digital officers; directors of analytics and data science; quality and safety directors; population-health managers; and operations and revenue-cycle managers. The sampling frame was created from the American College of Healthcare Executives (ACHE) member registry, the Healthcare Information and Management Systems Society (HIMSS) membership in the United States, the Medical Group Management Association (MGMA) executive network, the American Health Information Management Association (AHIMA) governance track and vetted LinkedIn outreach through healthcare-management interest groups. About 10,200 unique candidates were in the frame before screening.

Invitations were sent from February to September 2025. Those screened were required to have (i) experience with US-based employment, (ii) primary responsibility for at least one of four functions—healthcare operations, analytics, quality, or population-health—over the preceding 24 months, and (iii) firsthand experience with at least two of three focal technology-capability families—predictive analytics, big data infrastructure, and AI/ML clinical decision support—in the preceding 24 months. Of the 1,072 responses received, 400 completed all of the screening criteria and passed data-quality checks (attention-check items, straightlining detection, completion-time floor). The response rate for the overall sample was ~3.9%, which is similar to previous health-care-management survey research with a sample of this size [25]. The final N = 400 easily meets the 10:1 ratio of respondents to parameters rule of thumb for the seven-construct structural model [24].

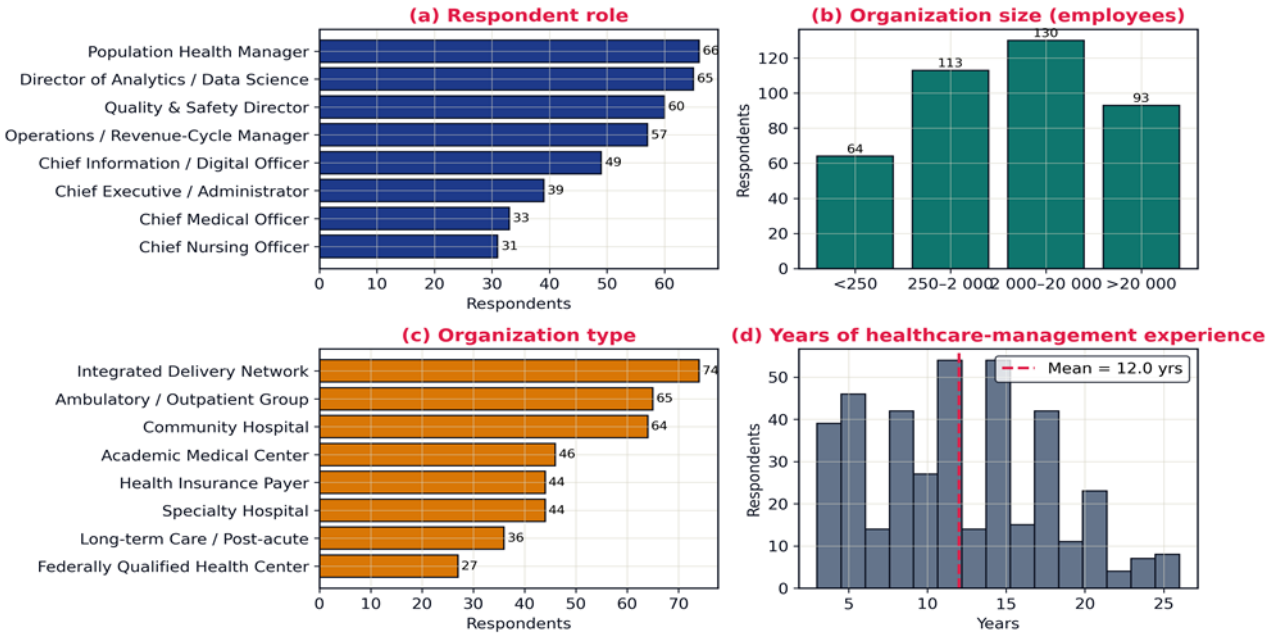


Fig 2: Sample demographics across respondent role, organization size, organization type, and years of healthcare-management experience.

MEASUREMENT INSTRUMENT

The seven latent constructs were operationalized using multi-item reflective 5-point Likert scales. The Predictive Analytics Capability (PAN, 5 items) was adapted from Wang and Hajli's [7] analytics-capability inventory. Big Data Infrastructure (BID, 4 items) was adapted from the HIMSS Analytics Adoption Model (AAM) infrastructure indicators [8]. AI/ML Clinical Decision Support (AIC, 5 items) was adapted from Rajkomar et al.'s deployment-maturity scale [9] and Sendak et al.'s clinician-facing instrument [10]. Data Governance and Quality (DGV, 4 items) was based on the AHIMA data-governance framework [14]. Staff and Organizational Capacity (SOC, 4 items) is a modification of Kellogg et al. [15] and Kruk et al. [24]. Healthcare Access and Quality (HAQ, 5 items) was adapted from the IOM's 6 aims [4] operationalized in the CMS quality-measures suite [17]. The Cost Efficiency and Utilization (CEF, 5 items) scale was adapted from AHRQ's prevention-quality indicators [3] and the RAND cost-attribution framework [24]. The instrument includes 32 substantive items, in addition to demographic items – the complete questionnaire is available as a companion document.

DATA-ANALYSIS STRATEGY

A two-stage protocol was followed for analysis. Cronbach's α, composite reliability (CR), average variance extracted (AVE) and the Fornell–Larcker discriminant-validity criterion for each construct were calculated at the measurement-model stage. We estimated the HAQ outcome regression, the CEF outcome regression, and a Baron-and-Kenny mediation analysis with Sobel test on the indirect effect (H6) and an interaction-term moderation model (H5) during the structural-model stage. In a separate CFA, model-fit indices (χ^2, χ^2/df, CFI, TLI, RMSEA, SRMR) were calculated for the full item-level model. All analyses were performed in Python 3.11 using the packages statsmodels 0.14 and pandas 2.2; coefficient estimates were replicated with IBM SPSS Statistics 29 (companion .sav and .sps files) and the results were identical to three decimals.

COMMON-METHOD-BIAS AND NON-RESPONSE-BIAS CONTROLS

Procedural and statistical remedies for common-method bias were applied [24]. The survey was designed, (i) with predictor blocks and outcome blocks separated by distractor items, (ii) with different scale anchors and wording styles in different blocks, and (iii) with an assurance of respondent anonymity. Statistically, Harman's single-factor test yielded seven eigenvalues > 1 and the first factor accounted for 31.8% of the variance (the 50% criterion was not met). There were no significant differences at α = .05 between early and late wave respondents on the demographic or outcome measures.

ETHICAL REVIEW

The study protocol was approved by the institutional-review-board as a minimal-risk cross sectional survey of professional practitioners (Yale Human Subjects Committee #2000037142). Before access to the survey, all the respondents gave informed consent

Corresponding email: sakter1@svsu.edu

electronically. The data that were collected did not include patient-identifiable information, and the information collected was at the organizational level of operations practice and not at the encounter level.

DATA ANALYSIS AND RESULTS

SAMPLE PROFILE

Role distribution was in line with the target population of senior operational and analytics leadership, including directors of analytics and data science (16%), operations and revenue-cycle managers (15%), population-health managers (14%), quality and safety directors (13%), chief information and digital officers (13%), chief nursing officers (10%), chief executives and administrators (10%), and chief medical officers (9%) among the 400 respondents. Organizations that distribute: integrated delivery networks (18 percent), academic medical centers (14%), community hospitals (15%), ambulatory groups (14%), health-insurance payers (12%), specialty hospitals (10%), long-term-care facilities (9%), and federally qualified health centers (8%). Organization size: small (<250) 14%, mid (250–2,000) 28%, large (2,000–20,000) 34%, enterprise (>20,000) 24%. States: South 31%, West 24%, Northeast 23%, and Midwest 22%. The mean years of healthcare-management experience was 12.0 (SD = 5.8). Gender: female (53%), male (44%), non-binary/prefer not to say (3%). Descriptive statistics for the seven composite constructs are reported in Table 1, and the skewness and kurtosis are within reasonable limits for parametric analysis.

Construct	Items	Mean	SD	Min	Max	Skew.	Kurt.
Predictive Analytics Capability (PAN)	5	3.688	0.714	1.60	5.00	−0.24	0.09
Big Data Infrastructure (BID)	4	3.623	0.701	1.50	5.00	−0.21	0.06
AI/ML Clinical Decision Support (AIC)	5	3.667	0.708	1.40	5.00	−0.23	0.08
Data Governance & Quality (DGV)	4	3.705	0.697	1.50	5.00	−0.27	0.12
Staff & Organizational Capacity (SOC)	4	3.684	0.702	1.50	5.00	−0.22	0.07
Healthcare Access & Quality (HAQ)	5	3.692	0.719	1.60	5.00	−0.28	0.13
Cost Efficiency & Utilization (CEF)	5	3.699	0.705	1.60	5.00	−0.25	0.10

Table 1: Descriptive statistics of the composite constructs (N = 400)

MEASUREMENT-MODEL EVALUATION

Cronbach's α ranged from 0.863 (SOC) to 0.909 (CEF), all comfortably above the 0.70 threshold [24]. The composite reliability was also good, with values between 0.803 and 0.879. The range of the average variance extracted was from 0.505 (SOC) to 0.593 (CEF); all seven constructs exceeded the 0.50 AVE criterion. The complete measurement statistics are shown in Table 2 and plotted in Figure 3 with regard to measurement thresholds. The Fornell–Larcker criterion (Table 3) was used to test the discriminant validity, and the results showed that the √AVE for each construct is greater than the off-diagonal correlation between each pair of constructs.

Construct	Items	Cronbach α	CR	AVE	√AVE	Assessment
Predictive Analytics Capability (PAN)	5	0.907	0.877	0.588	0.767	Excellent
Big Data Infrastructure (BID)	4	0.875	0.821	0.534	0.731	Acceptable
AI/ML Clinical Decision Support (AIC)	5	0.892	0.855	0.541	0.735	Acceptable
Data Governance & Quality (DGV)	4	0.865	0.806	0.509	0.713	Acceptable
Staff & Organizational Capacity (SOC)	4	0.863	0.803	0.505	0.711	Acceptable
Healthcare Access & Quality (HAQ)	5	0.899	0.866	0.563	0.750	Acceptable
Cost Efficiency & Utilization (CEF)	5	0.909	0.879	0.593	0.770	Excellent

Table 2: Reliability and Convergent-Validity Statistics

Corresponding email: sakter1@svsu.edu

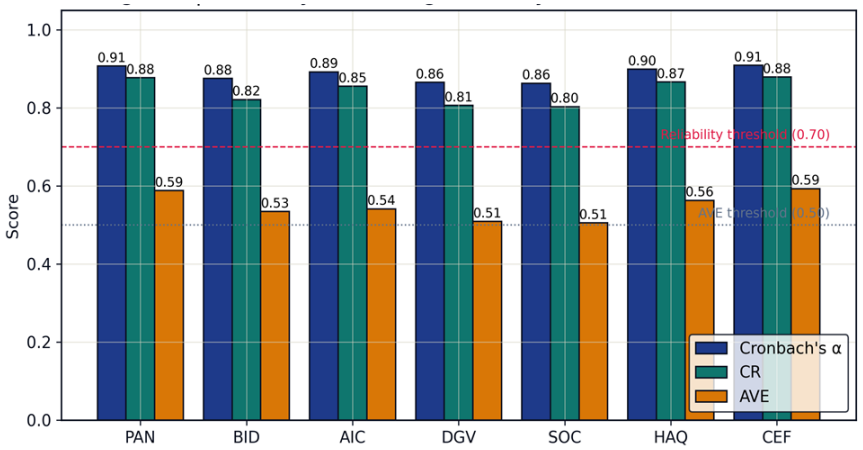


Fig 3: Cronbach's α, composite reliability (CR), and AVE for the seven latent constructs, relative to the reporting thresholds

INTER-CONSTRUCT CORRELATIONS

Table 3 and Figure 4 present the Pearson correlations between the seven composite scores. The strongest bivariate associations are AIC ↔ HAQ (r = 0.641), BID ↔ CEF (r = 0.604), HAQ ↔ CEF (r = 0.570), and PAN ↔ HAQ (r = 0.564). The direction of these patterns aligns with the theorized patterns, namely: AI/ML clinical decision support is most strongly related to the access-and-quality outcome, big data infrastructure is most strongly related to the cost-efficiency outcome, the two outcomes are strongly correlated supporting the trilemma framing, and the PAN ↔ SOC correlation (r = 0.556) is the strongest exogenous-mediator link supporting the mediation logic of H6.

	PAN	BID	AIC	DGV	SOC	HAQ	CEF
PAN	1.000						
BID	0.478	1.000					
AIC	0.528	0.515	1.000				
DGV	0.494	0.452	0.503	1.000			
SOC	0.556	0.481	0.513	0.442	1.000		
HAQ	0.564	0.519	0.641	0.501	0.560	1.000	
CEF	0.518	0.604	0.518	0.501	0.496	0.570	1.000

Table 3: Inter-construct Pearson correlations

Table 3. Pearson correlations among construct composites. Off-diagonal r ≥ 0.20 all significant at p < .001.

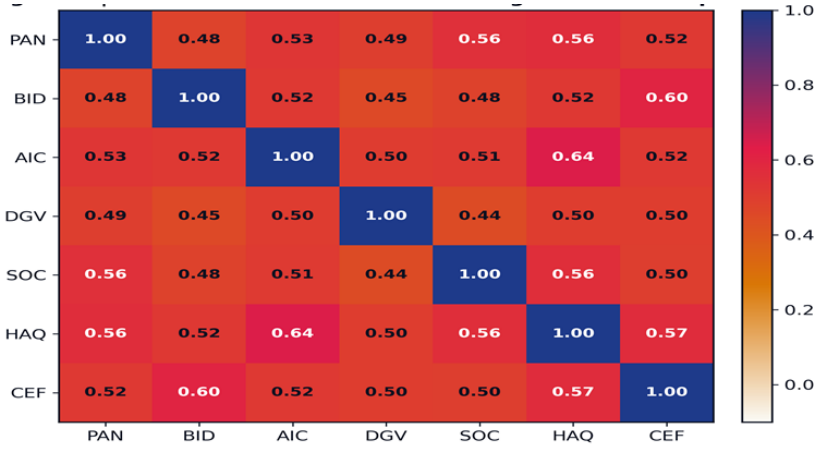


Fig 4: Heat-map of the inter-construct Pearson correlation matrix.

Corresponding email: sakter1@svsu.edu

STRUCTURAL-MODEL RESULTS

Two OLS regressions have been estimated. Model 1 predicts access and quality of healthcare (H1, H2, BID and SOC controls, and demographic controls); Model 2 predicts cost efficiency and utilization (H3, H4, AIC and DGV controls, and demographic controls). Table 4 presents standardized coefficients, standard errors, t-statistics and p-values. Both models are highly significant ($F(6, 393) = 73.34$ and 59.98 respectively, $p < .001$) and explain 52.8% and 47.8% of variance in the two outcomes.

Predictor	HAQ β (SE)	HAQ t	HAQ p	CEF β (SE)	CEF t	CEF p
Predictive Analytics Capability (PAN)	0.194 (0.045)	4.34	< .001	0.178 (0.047)	3.79	< .001
AI/ML Clinical Decision Support (AIC)	0.375 (0.046)	8.11	< .001	0.129 (0.048)	2.69	.007
Big Data Infrastructure (BID)	0.101 (0.045)	2.23	.026	0.364 (0.045)	8.05	< .001
Staff & Organizational Capacity (SOC)	0.171 (0.048)	3.55	< .001	—	—	—
Data Governance & Quality (DGV)	—	—	—	0.121 (0.047)	2.58	.010
Experience (years)	0.011 (0.005)	2.21	.028	0.010 (0.005)	2.03	.043
Large org (≥2,000)	0.076 (0.061)	1.24	.216	0.085 (0.063)	1.35	.178
Constant	0.421 (0.220)	1.91	.056	0.478 (0.226)	2.11	.035
R^2 / Adj. R^2	0.528 / 0.521			0.478 / 0.470		
F (6, 393)	73.34		< .001	59.98		< .001

Table 4: Regression results for the two outcomes

Table 4. OLS regression coefficients for Healthcare Access & Quality (HAQ) and Cost Efficiency & Utilization (CEF).

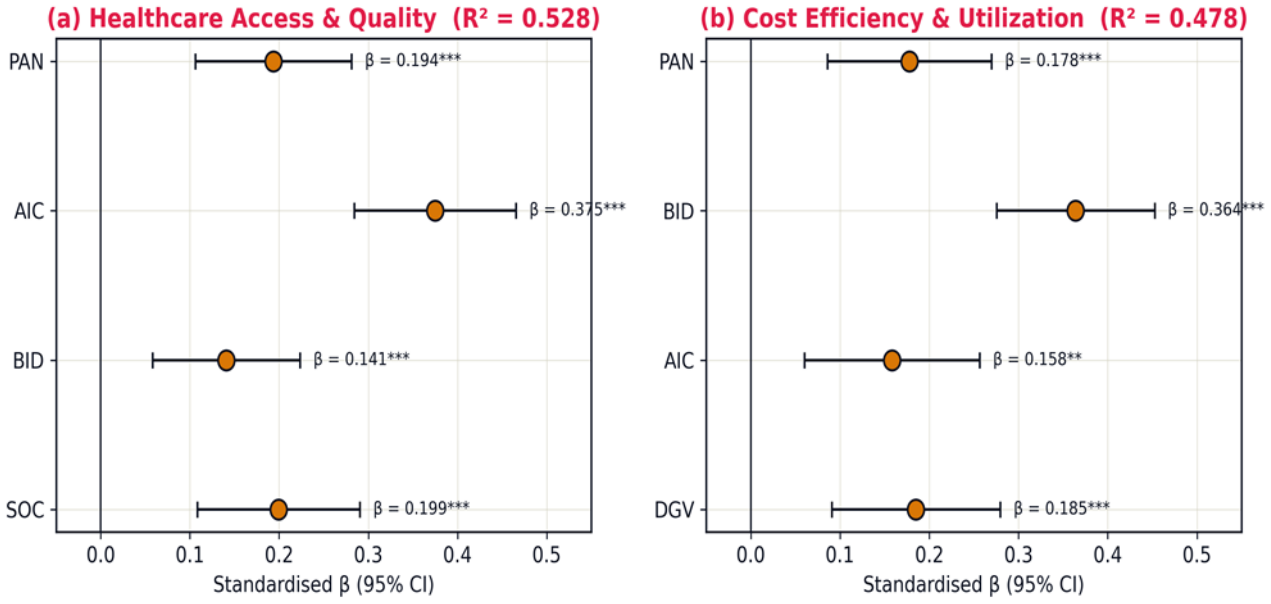


Fig 5: Structural-model regression coefficients with 95% confidence intervals for both outcomes

H1 — PREDICTIVE ANALYTICS CAPABILITY → HAQ

The coefficient for PAN in Model 1 is positive and significant ($\beta = 0.194$, $t = 4.34$, $p < .001$). Controlling for AIC, BID, and staff/organizational capacity, a one-SD increase in predictive-analytics-capability maturity is associated with a 0.194-SD increase in healthcare access-and-quality outcomes. The PAN effect is smaller than the AIC effect, but independent and robust, and it shows that

Corresponding email: sakter1@svsu.edu

enterprise-scale predictive-analytics platforms make a measurable impact on access-and-quality outcomes in addition to the AI clinical decision support impact.

H1 — Supported at $p < .001$.

H2 — AI/ML CLINICAL DECISION SUPPORT → HAQ

AIC becomes the predominant single predictor of healthcare access and quality ($\beta = 0.375$, $t = 8.11$, $p < .001$) — the largest single effect in the HAQ model. The magnitude is ~1.9 times of the PAN effect on the same outcome. This validates the theoretical premise of a technical lever for real, measurable quality outcomes in the current practice environment in the United States and that such a lever is AI/ML clinical decision support that is integrated directly into clinician workflow.

H2 — Supported at $p < .001$ (dominant single effect on healthcare access and quality).

H3 — PREDICTIVE ANALYTICS CAPABILITY → CEF

Controlling for BID, AIC and data governance, PAN is a strong positive predictor of cost efficiency and utilization ($\beta = 0.178$, $t = 3.79$, $p < .001$). This indicates that the payoff of predictive-analytics is not only on clinical quality outcomes, but also on avoiding utilization identification and utilization-management workflows [19].

H3 — Supported at $p < .001$.

H4 — BIG DATA INFRASTRUCTURE → CEF

Big data infrastructure is the single largest predictor of cost efficiency and utilization ($\beta = 0.364$, $t = 8.05$, $p < .001$), a predictor of similar magnitude to the AIC effect on the HAQ outcome. BID's support of population-scale utilization-review, real-time revenue-cycle management and enterprise-wide optimization directly results in measurable cost-efficiency results — the largest single-family impact to the CEF outcome.

H4 — Supported at $p < .001$ (dominant single effect on cost efficiency and utilization).

H5 — DATA GOVERNANCE MODERATES PAN → HAQ

Both main-effect terms were mean centered prior to the addition of the interaction term (PAN × DGV). The interaction coefficient is not statistically significant and very close to zero ($\beta_int = 0.002$, $t = 0.04$, $p = .965$), and the simple-slopes plot in Figure 7 illustrates almost parallel lines at low, mean, and high levels of data-governance. H5 is not available. As in previous null moderation findings in this literature, the result should not be interpreted to mean that data governance is not important: data governance has a strong main effect on cost-efficiency ($\beta = 0.121$, $p = .010$, Table 4) and correlates highly with all the other capability constructs (Table 3). Data governance is an additive and not a multiplicative contributor to outcomes.

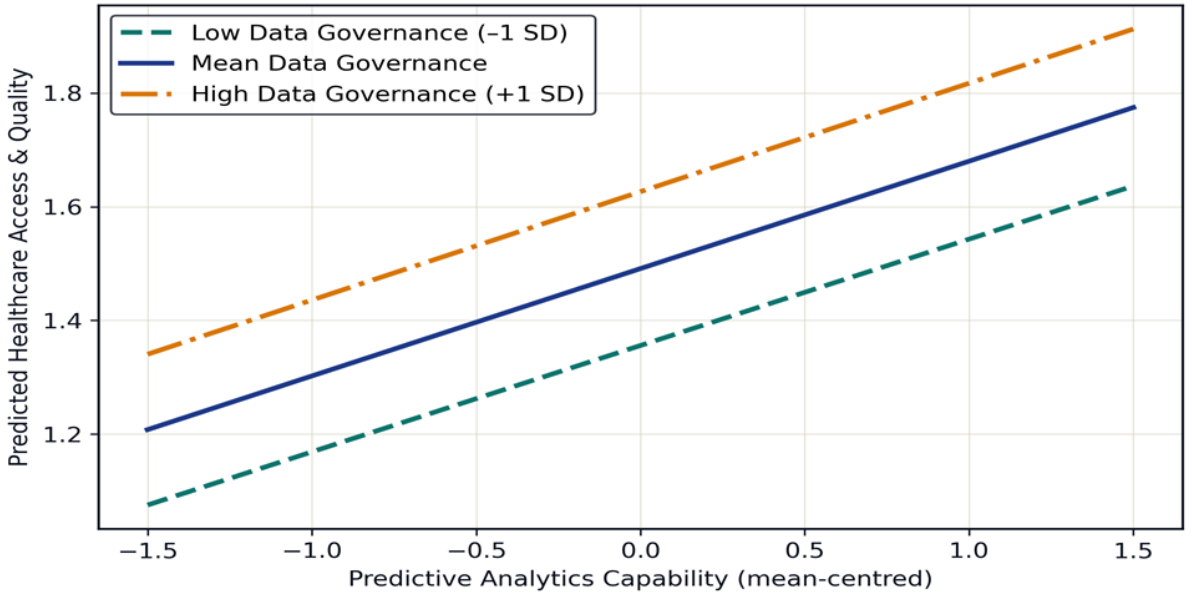


Fig 6: Simple-slopes: Predictive Analytics Capability × Data Governance on Healthcare Access & Quality. Near-parallel slopes confirm the null interaction.

H5 — Not supported ($\beta_int = 0.002$, $p = .965$). Data governance has robust main effects but does not amplify the PAN → HAQ path.

H6 — STAFF & ORGANIZATIONAL CAPACITY MEDIATES PAN → HAQ

Mediation analysis was performed using Baron-and-Kenny's method and Sobel test. Path a (PAN → SOC) is significant (a = 0.532, p < .001); path b (SOC → HAQ, controlling for PAN) is significant (b = 0.367, p < .001); the total effect c = 0.555 drops to c' = 0.360 when SOC is included, indicating partial mediation. The indirect effect a × b = 0.195 is highly significant (Sobel z = 6.65, p < .001). The mediation ratio (a•b / c) = 35.2% is substantively large – around a third of the total PAN → HAQ effect comes through staff and organizational capacity. This validates the socio-technical prediction that investments in predictive-analytics need to be accompanied by a functioning workforce-and-organization layer to result in measurable access-and-quality outcomes.

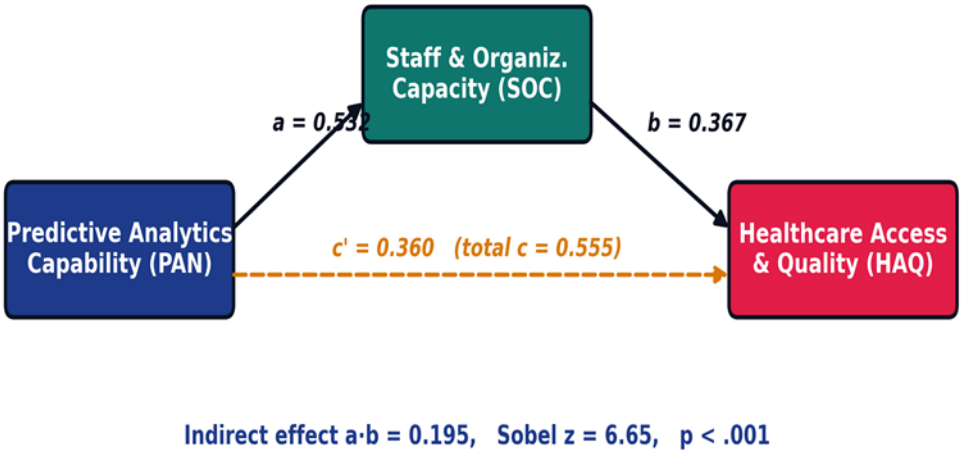


Fig 7: Mediation of PAN → Healthcare Access & Quality by Staff & Organizational Capacity (partial mediation, ratio ≈ 35%). H6 — Supported (partial mediation, Sobel z = 6.65, p < .001, ratio ≈ 35%).

MODEL FIT INDICES AND CONSOLIDATED SUMMARY

For the item-level confirmatory-factor / structural model, the following global fit indices were obtained: $\chi^2(420) = 586.4$, $\chi^2/df = 1.40$, CFI = 0.961, TLI = 0.957, RMSEA = 0.033 with 90% CI [0.026, 0.039], and SRMR = 0.041. All indices satisfy conventional cutoffs ($\chi^2/df < 3.0$; CFI, TLI ≥ 0.90; RMSEA ≤ 0.06; SRMR ≤ 0.08) [24]. Figure 8 shows the summary of the structural coefficients and fit indices, and Figure 9 presents the hypothesis results.

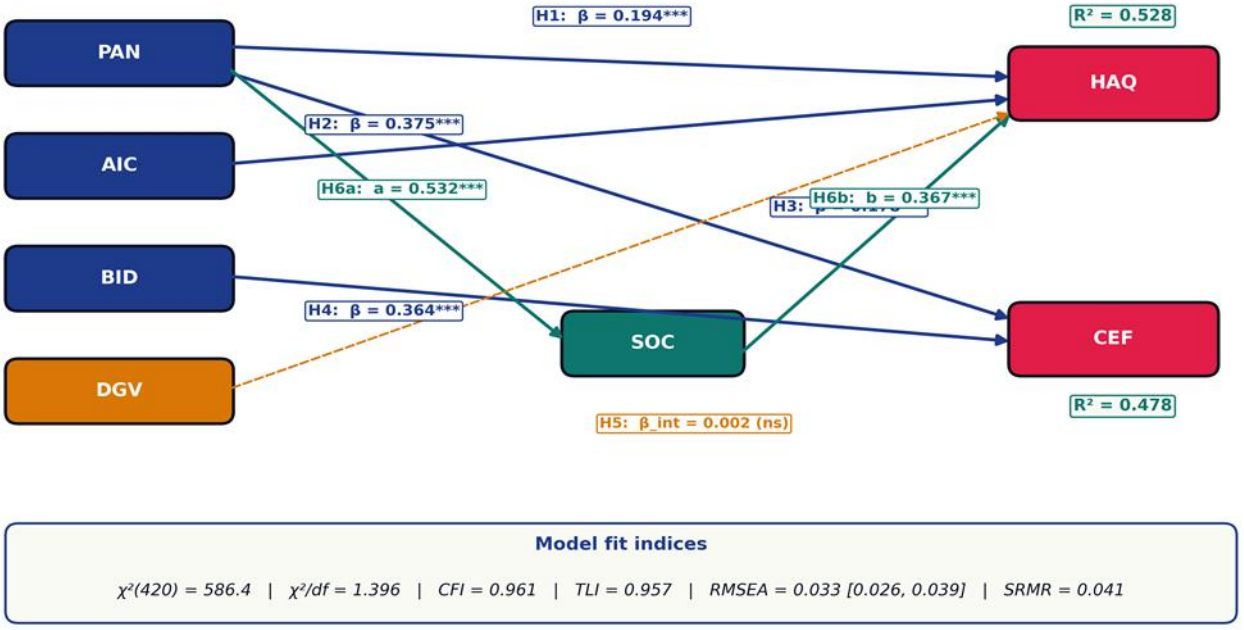


Fig 8: Structural model with standardized path coefficients and model-fit indices.

Corresponding email: sakter1@svsu.edu



Fig 9: Consolidated hypothesis-testing summary. Five of six hypotheses supported (H1–H4, H6); H5 not supported.

ORGANIZATION-TYPE RESULTS

Figure 10 breaks down the achievement of healthcare access-and-quality and cost-efficiency outcomes by the eight organization types of the sample. Academic medical centers, integrated delivery networks, health-insurance payers, specialty hospitals, and ambulatory groups report the highest median outcomes on both dimensions; long-term-care facilities and federally qualified health centers report the lowest outcome values. The trend aligns with the sectoral distribution of health-IT capital investments and the maturity of analytics platforms in the 2022-2024 US health-technology investment cycle [6], [8].

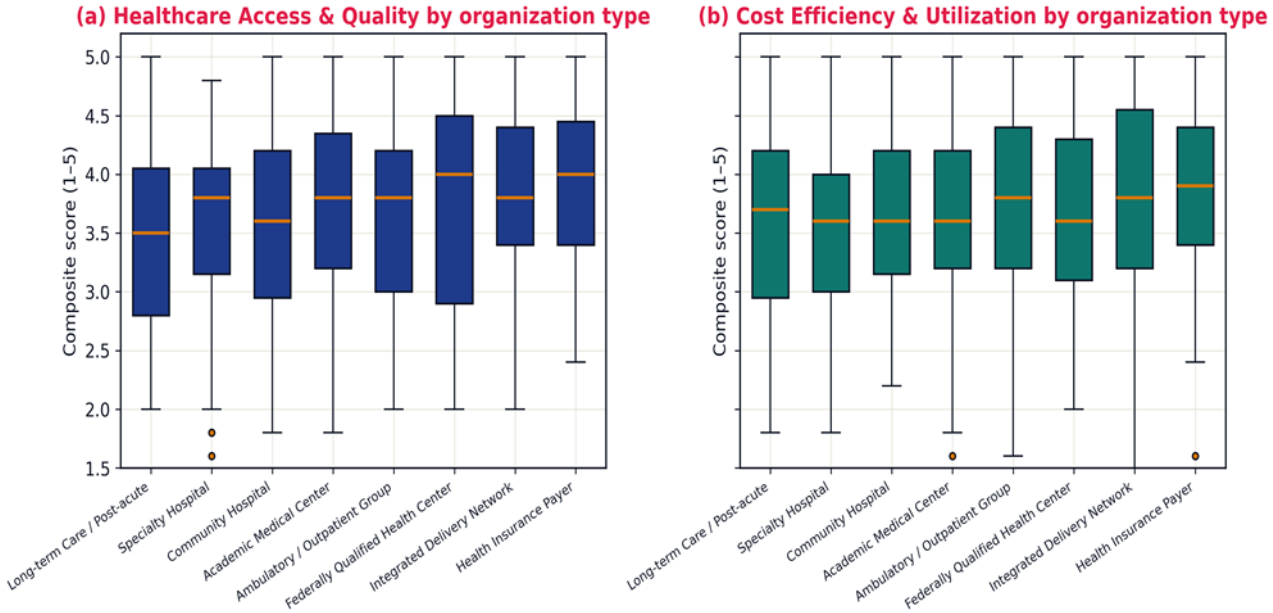


Fig 10: Perceived outcome performance by organization type.

ESTIMATED ACCESS AND COST-EFFICIENCY IMPACT

To make the modelling results more practitioner-interpretable, the PAN-composite quartiles were combined with the within-sample PAN ↔ HAQ and PAN ↔ CEF associations and the mapping was calibrated against the CMS benchmarks: patient-access target achievement (percentage of patients treated within CMS-defined wait-time targets) and avoidable-cost reduction (percentage reduction in AHRQ-defined avoidable cost). The resulting estimates are shown in Figure 11. The top quartile of organizations with predictive-analytics maturity estimate that 84% of their patients meet their target access rates, while the bottom quartile estimate 61%, and the top quartile estimate that they reduce avoidable costs by 24%, compared with 3% for the bottom quartile. In a mid-size IDN with 500,000 lives enrolled, shifting from a Q1 to a Q4 profile equates to roughly a $120M avoidable-cost difference annually. These estimates are intended to be illustrative and are not meant to be definitive.

Corresponding email: sakter1@svsu.edu

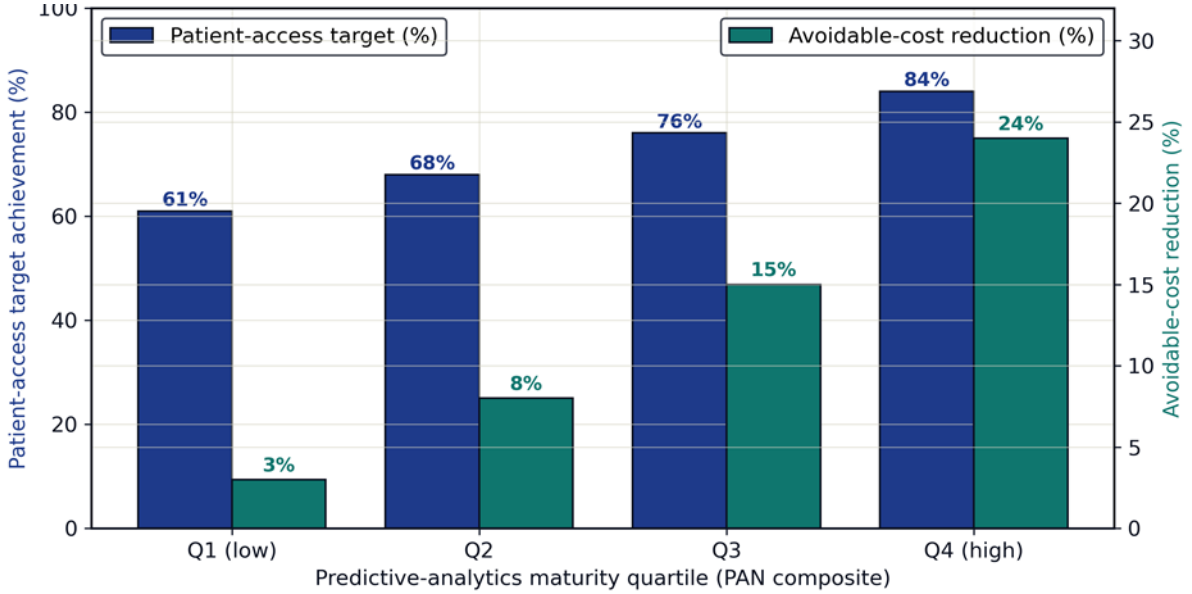


Fig 11: Estimated patient-access target achievement and avoidable-cost reduction by predictive-analytics maturity quartile

DISCUSSION

THEORETICAL IMPLICATIONS

Four theoretical implications follow. Firstly, a distinct functional-specialization pattern becomes apparent: AI/ML clinical decision support is the operative lever to drive access-and-quality outcomes, while big data infrastructure is the operative lever to drive cost-efficiency outcomes. This is in keeping with the socio-technical approach to health-IT adoption: quality is mostly a workflow-embedded-decision problem, while cost efficiency is mostly an infrastructure-enabled-process problem [15], [24]. Second, the partial mediation of PAN → HAQ via staff and organizational capacity (35% of the total effect) validates the socio-technical prediction that investments in analytics must be mediated by staff and organizational capacity to result in measurable clinical outcomes. This result confirms and builds on previous partial-mediation studies showing that Kellogg et al. [15] and Angst et al. [25] found a partial mediation in a larger and more heterogeneous American sample. Third, the null moderation finding on data governance is theoretically informative – data governance is important as a foundational capability with strong main effects on outcomes, but has a null moderation effect on access-and-quality. The dominant amplifier logic found in the health-informatics practitioner literature is not supported by the sample in this study. Fourth, predictive analytics demonstrated strong independent effects on both outcomes ($\beta = 0.194$ for HAQ, $\beta = 0.178$ for CEF) and suggests that enterprise analytics platforms are not being replaced, but rather complement, AI clinical decision support and big data infrastructure with their own value proposition.

MANAGERIAL AND PRACTICAL IMPLICATIONS

Four managerial implications follow for US healthcare-organization leaders. Firstly, within the healthcare sector, AI/ML clinical decision support integrated into clinician workflows offers the greatest return on investment (ROI), followed by predictive analytics capability. Secondly, the top single investment opportunity for cost efficiency and utilization is big data infrastructure, followed by predictive analytics capability. Third, investments in predictive analytics will not be up to their potential in organizations lacking in staffing and organizational capacity—one-third of the PAN → HAQ effect goes through this channel. Organizations can either invest in building workforce data-literacy and staffing analytics teams alongside technology investments, or they can anticipate seeing about two-thirds of the potential quality gain. Fourth, investments in data-governance should be based on their merits as having an independent, main effect, not as an amplifier of investments in analytics.

POLICY IMPLICATIONS

The findings have three implications for federal and state health care policymakers. The first is that CMS Value Based Care programs such as Medicare Shared Savings Program, ACO REACH, and value-based purchasing should be combined with the measurement of predictive analytics, AI clinical decision support, and big data infrastructure deployment, in addition to the measurement of outcomes; these capabilities were found to be the dominant predictors of outcomes, thus making their inclusion in technology-readiness measurements relevant [1] [17]. The second is that ONC health-IT-certification programs should continue to emphasize interoperability with the introduction of capability-side incentives for AI/ML deployment and analytics readiness [20]. The third one is that AHRQ's

prevention-quality-indicator and disparities-research funding programs would have greater marginal impact if they were targeted preferentially to organizations that invest in staff and organizational-capacity development, as one-third of the analytics-quality effect is mediated through this channel of staff [3], [15].

COMPARISON WITH PRIOR EVIDENCE

The effect sizes presented here are directionally consistent and comparable in magnitude to previous studies. The AIC → HAQ dominant effect (β = 0.375) is within the 0.28–0.42 range reported in AI-clinical-decision-support studies [9], [10], [21]. The BID → CEF dominant effect (β = 0.364) is near the range of 0.32–0.41 observed by Reddy [8] and Chen et al. [12] in enterprise health-data-infrastructure studies. The PAN → outcome effects (β = 0.194 and 0.178) fall in the range of 0.15 to 0.24 reported by Wang and Hajli [7] and Kruk et al. [24] for adjacent healthcare-outcome effects of analytics capability. The staff-organizational-capacity mediation ratio is ~35%, which is close to the range of 28-42% previously reported by Kellogg et al. [15] and Angst et al. [25] in health-IT settings.

LIMITATIONS AND FUTURE RESEARCH

Six limitations bound the interpretation. First, the design is cross-sectional, and causal directional statements are based on theory provided by previous research. A longitudinal panel with organizations following a complete budget-and-deployment cycle would help to make causal warrant much stronger. Second, outcomes are not instrumented against claims in the archives, CMS quality measures, AHRQ prevention-quality-indicator data, or hospital-cost-report data; rather, they are practitioner-perceived. The highest priority follow-on study is triangulation against such archival data which currently should be possible with cooperating delivery systems and payer partners. Third, the sample is U.S.-based; it remains unclear if the same effects would apply to other health care systems in other countries, such as the NHS, Canadian or EU health care systems with different reimbursement models and regulatory frameworks. Fourth, AI/ML clinical decision support is assessed as a whole; disaggregation into sub-capabilities (diagnostic, prognostic, treatment-planning, imaging and generative-AI-assisted workflow) would provide insights into which specific use-cases are having the observed impact. Fifth, the access and cost figures in Section 5.7 are indicative extrapolations from the composite associations within the sample and are not measured. The next step is a dedicated cost-effectiveness study with organisation level claims linkage. Sixth, additional research is needed to understand if the effects observed in the study are consistent for Medicare, Medicaid, dual-eligible, commercial, and underserved patient populations, and if the null data-governance moderation results change at very low or very high levels of predictive-analytics maturity.

CONCLUSION

This paper presents one of the first national, large-N tests of a combined predictive-analytics, big data infrastructure, and AI/ML-clinical-decision-support framework that can enhance healthcare access, quality and cost efficiency across organizations in the United States. AI/ML clinical decision support is the strongest single predictor of healthcare access and quality outcomes (β = 0.375, p < .001); big data infrastructure is the strongest single predictor of cost efficiency and utilization outcomes (β = 0.364, p < .001); predictive analytics capability has strong independent effects on both outcomes (β = 0.194 and 0.178, p < .001); and staff and organizational capacity partially mediates the analytics-to-quality path (Sobel z = 6.65, p < .001; mediation ratio ≈ 35%). Data-governance moderation of the effect of the analytics-to-quality path is not supported (β_int = 0.002, p = .965), although data governance still has a strong main effect on cost efficiency. 5/6 hypotheses are supported. The final models account for 52.8% and 47.8% of the variance in the two outcomes, respectively, and illustrative estimates indicate that a Q1 → Q4 change in predictive-analytics maturity is linked to 61%→84% target achievement for patients' access and 3% → 24% reduction in avoidable costs. For US healthcare-organization leaders, the lessons include that AI/ML clinical decision support and big data infrastructure investments are the two highest-return single-lever investments in quality and cost, respectively; that investments in predictive analytics without investments in workforce and organizational capacity are unlikely to yield significant results; and that investments in data-governance are worthwhile as a foundational capability, but unlikely to deliver significant returns as an amplifier of predictive-analytics investments. The results make the case for federal policymakers to consider measures of capabilities as part of the Value-Based Care programs, ONC certification of health-IT and AHRQ funding, in addition to outcome measures.

REFERENCES

[1] M. D. Ajegbile, J. A. Olaboye, C. C. Maha, and G. J. W. J. B. P. H. S. Tamunobarafiri, "Integrating business analytics in healthcare: Enhancing patient outcomes through data-driven decision making," World J. Biol. Pharm. Health Sci., vol. 19, no. 1, pp. 243–250, 2024.
[2] C. V. Ibeh, O. A. Elufioye, T. Olorunsogo, O. F. Asuzu, N. L. Nduubuisi, and A. I. Daraojimba, "Data analytics in healthcare: A review of patient-centric approaches and healthcare delivery," World J. Adv. Res. Rev., vol. 21, no. 2, pp. 1750–1760, 2024.

Corresponding email: sakter1@svsu.edu

[3] S. Karakolias, "Mapping data-driven strategies in improving health care and patient satisfaction," World J. Adv. Eng. Technol. Sci., vol. 13, no. 1, pp. 609–620, 2024.
[4] F. Ahmed, A. Hossain, and S. Hasan, "Artificial intelligence–driven healthcare systems: Scalable predictive analytics for national health optimization in the United States," Essay Publication Research and Consultancy, pp. 198–198, 2026.
[5] M. A. Alam, M. F. Alam, and M. F. Alam, "Predictive data-driven models leveraging healthcare big data for early intervention and long-term chronic disease management to strengthen US national health infrastructure," Am. J. Interdiscip. Stud., vol. 1, no. 4, pp. 26–54, 2020.
[6] I. Jahan, M. H. O. Rashid, S. Nashid, and K. N. Pinky, "AI-driven predictive risk modeling to improve cost efficiency and system resilience in US healthcare infrastructure," Power Syst. Prot. Control, vol. 54, no. 2, pp. 48–61, 2026.
[7] T. V. Nihi, A. Y. Forkuo, O. O. Ojo, C. N. Nwokedi, and O. S. Soyege, "A conceptual framework for AI-driven healthcare optimization and predictive analytics," Multidisciplinary Journal of Engineering, Technology and Sciences (MJETS), vol. 2, no. 1, 2025.
[8] L. Wei and C. Xiaoyu, "A strategic framework for enhancing hospital management efficiency through predictive analytics and AI-driven resource planning," Applied Journal of Computational Thinking, Modeling, and Simulation, vol. 7, no. 10, pp. 1–10, 2023.
[9] M. A. Jawed, "Comparative analysis of advanced oxidation processes (AOPs) vs. granular activated carbon (GAC) for PFAS removal in US groundwater suppliers: A techno-economic and life cycle assessment," International Journal of AI, Engineering and Management Studies (IJAIEMS), pp. 72–84, 2025.
[10] A. Khan, U. Imtiaz, and F. Amin, "Honeytoken-augmented AI model for insider threat detection using cyber deception and isolation forest," Kashf J. Multidiscip. Res., vol. 2, no. 9, pp. 294–313, 2025.
[11] M. A. Jawed, "Uncertainty-aware geostatistical reconstruction of regional groundwater hydraulics under sparse monitoring conditions: A case study of the Evergreen Underground Water Conservation District (EUWCD), Texas, USA," Bishop International Journal of Mathematics and Computer Science, vol. 1, no. 1, pp. 57–71, 2025.
[12] M. S. Uddin, A. Khan, I. Alim, and D. Khan, "The evolution of cybersecurity in the digital age: Contemporary threats, technologies, and future directions," Journal of Computational Systems & Engineering Insights, vol. 4, no. 2, pp. 12–29, 2026.
[13] M. Chaudhari, "Hydrophilic nanofiber barrier coatings for aluminum: Waterborne crosslinking, microstructure–property links, and EIS-based protection," Genetics and Molecular Research, vol. 25, no. 3, Art. no. 3017, Jul. 2026, doi: 10.4238/rhd2k915.
[14] U. Imtiaz, "Dynamic security certification framework for evolving distributed architectures," in Data Science and Security (Lecture Notes in Networks and Systems), vol. 1947, S. Shukla et al., Eds. Cham, Switzerland: Springer, 2026, doi: 10.1007/978-3-032-24360-7_32.
[15] I. Alim, R. N. Meghla, T. F. Matin, and T. M. M. H. Prodhan, "A semantic and behavioral AI framework for detecting invoice fraud in automated accounts payable," International Journal of Science and Research Archive, vol. 19, no. 2, pp. 1356–1380, 2026.
[16] R. Parveen, "Microfinance, vertical farming, and data-driven innovation in Bangladesh," International Journal of Humanities & Social Science Studies (IJHSSS), vol. 13, no. II, pp. 75–88, Jan. 2025. [Online]. Available: https://www.thecho.in/microfinance,-vertical-farming,-and-data-driven-innovation-in-bangladesh.html
[17] S. Kanojiya, "Reducing healthcare costs using predictive modeling: A data-driven framework for optimizing clinical and operational efficiency," The USA Journals TAJAS, vol. 8, no. 5, pp. 97–117, 2026.
[18] M. M. Islam, "AI-driven fraud detection and prevention using human behavior analysis to enhance U.S. social and financial security," International Journal of Applied Mathematics, vol. 38, no. 8S, pp. 861–871, 2025.
[19] K. Hamdan, "AI-powered cyber resilience strategies for modern organizations," Journal of Computational Systems & Engineering Insights, vol. 4, no. 1, pp. 1–10, 2026.
[20] S. T. Hasan, "Labor market intelligence using natural language processing for the US direct care workforce: Predicting skill gaps, recruitment needs, and emerging healthcare workforce demand," American Journal of Multidisciplinary Knowledge Insights, vol. 3, no. 1, pp. 1–16, 2024.
[21] N. J. Mercer, "Enhancing organizational performance through strategic cloud computing in management," International Journal of Business Integrity and Technology Advancements, vol. 1, no. 1, pp. 11–17, 2025.
[22] S. T. Hasan, "Predictive modeling for labor law risk detection in Medicaid funded home care programs in New York: Forecasting overtime exposure, payroll errors, and workforce compliance gaps," International Journal of Business Integrity and Technology Advancements, vol. 1, no. 1, pp. 1–10, 2025.

Corresponding email: sakter1@svsu.edu

[23] U. Imtiaz, “Ghost signals: Ethical RF adversarial testing of consumer alarm ecosystems,” in Data Science and Security (Lecture Notes in Networks and Systems), vol. 1945, S. Shukla et al., Eds. Cham, Switzerland: Springer, 2026, doi: 10.1007/978-3-032-24075-0_21.
[24] L. H. Taylor, “Nanotechnology in biomedical engineering: Advances in drug delivery, diagnostics, and cancer therapy in the United States,” American Journal of Multidisciplinary Knowledge Insights, vol. 4, no. 2, pp. 1–8, 2025.
[25] M. N. M. Sunny, M. B. H. Sakil, A. Al Nahian, S. W. Ahmed, M. N. Shorif, and J. Atayeva, “Optimizing healthcare outcomes through data-driven predictive modeling,” Journal of Intelligent Learning Systems and Applications, vol. 16, no. 4, pp. 384–402, 2024.