Journal of Medical Science, Biology, and Chemistry (JMSBC)

ISSN: 3079-2576 (Online)

Volume 2 Issue 2, (2025)

https://doi.org/10.69739/jmsbc.v2i2.1149

https://journals.stecab.com/jmsbc

Published by
Stecab Publishing

Review Article

From Models to Management: A Framework for Predictive Analytics in Health Systems

*[1]Chizube Obinna Chikezie, [2]Chibuzo Okechukwu Onah, [3]Chukwudi Anthony Okolue, [4]Mary-Jane Ezinne Ugbor, [5]Emmanuel Fagbenle, [6]Olukunle Akanbi

About Article

Article History

Submission: September 30, 2025
Acceptance : November 04, 2025
Publication : November 16, 2025

Keywords

Artificial Intelligence Healthcare, Decision Support Systems, Electronic Health Records, Health Services Research, Predictive Analytics in Healthcare

About Author

[1] School of Computing Engineering and Intelligent Systems, Ulster University, York Street, Belfast BT15 1ED, UK

[2] College of Business, Georgia State University, Atlanta, USA

[3] Owen Graduate School of Management, Vanderbilt University, Nashville, USA

[4] Faculty of Pharmaceutical Sciences, University of Port Harcourt, Port Harcourt, Nigeria

[5] Department of Decision Science, University of New Hampshire, Durham, USA

[6] Graduate School of Business & Leadership, National Louis University, Tampa, USA

Contact @ Chizube Obinna Chikezie
kezie.c.o@gmail.com

ABSTRACT

Health systems increasingly deploy predictive analytics to improve patient outcomes and operational performance, yet many projects stall at the interface between model output and managerial action. This review looks at real-world deployments connecting clinical prediction with market and operational levers, staffing, bed flow, outreach, and scheduling, and distills an integration framework spanning data architecture, model selection, real-time pipelines, governance, and evaluation. Three questions organize the review: which integration patterns improve outcomes, which technical and organizational conditions enable scale, and how transferable are U.S. findings to other health systems. Evidence emphasizes measurable effects on process and, in selected contexts, outcomes when models are embedded in event-driven workflows and governed with clear decision rights, calibration monitoring, and explainability support. Because much of the empirical literature originates in the United States, generalizability is assessed using compact international implementations (United Kingdom stroke AI, Singapore's C3 command center, and India's TB computer-aided CXR triage). The review argues that impact depends less on algorithmic novelty than on socio-technical integration: reliable data plumbing, execution discipline, and incentives aligned to net clinical and operational utility.

Citation Style:

Chikezie, C. O., Onah, C. O., Okolue, C. A., Ugbor, M.-J. E., Fagbenle, E., & Akanbi, O. (2025). From Models to Management: A Framework for Predictive Analytics in Health Systems. *Journal of Medical Science, Biology, and Chemistry, 2*(2), 252-261. https://doi.org/10.69739/jmsbc.v2i2.1149

1. INTRODUCTION

Predictive analytics has moved from pilot projects to routine decision support in many health systems, spanning clinical deterioration alerts, imaging triage, readmission risk, demand forecasting, and capacity optimization. Adoption is broad but uneven: analyses of U.S. hospital surveys indicate widespread use of AI-assisted predictive tools embedded in EHRs, while formal evaluation and bias assessment remain inconsistent (Nong *et al.*, 2025). Despite proliferation, the impact varies for a structural reason: analytics frequently stops at model development rather than management integration. Projects that succeed typically wire predictions into operational levers, such as bed assignment, escalation pathways, staffing rosters, outreach campaigns, or slot releases, via event-driven data flows and explicit decision rights. Conversely, deployments falter when calibration drifts, alerts lack actionability, or governance is unclear. The external validation of widely used proprietary models illustrates the stakes. For instance, Wong *et al.* (2021) reported poor discrimination and calibration of a national sepsis model, prompting calls for transparent evaluation and post-deployment monitoring (Wong *et al.*, 2021).

Equity and trust are very important. An influential investigation indicated that a commercial algorithm used for care management encoded racial bias because it optimized on cost rather than clinical need, systematically disadvantaging Black patients (Obermeyer *et al.*, 2019). The lesson is at the design level: pick targets and proxies that are good for the patient, then report how well each group is doing and make changes as needed. Emerging proposals for assurance laboratories and standardized model cards/factsheets aim to institutionalize transparent documentation, monitoring, and governance across sites (IBM, 2015; Olsen, 2024; Shah *et al.*, 2024).

Methodological clarity on clinical utility is also required. Discrimination metrics (e.g., AUROC) do not reveal whether acting on a score benefits patients or operations. Decision-curve analysis provides a threshold-aware net-benefit view and is increasingly recommended in clinical prediction reporting (Vickers *et al.*, 2019).

From a systems perspective, two integration gaps recur. First, the technical architecture requires robust interfaces, such as HL7® FHIR® Subscriptions and backport guides, to deliver near-real-time signals and change events to stream processors and feature stores that can perform idempotent, low-latency scoring (HL7 International, 2024). Second, organizational alignment: without defined owners, escalation pathways, and incentive-compatible KPIs, predictions cannot reliably alter throughput, safety, or cost-to-serve.

This review responds to those gaps with three research questions:

1. Which integration patterns reliably improve patient outcomes and operational performance?
2. Which technical (data, modeling, compute, latency) and organizational What technical (data, modeling, compute, latency) and organizational conditions, such as governance, incentives, and change management, enable scalability?
3. How transferable are U.S. derived lessons to other systems?

The scope and generalizability of the approach are crucial factors to consider. Much of the empirical base arises from the United States, where financing is dominated by commercial insurance and mixed-payer contracts; by contrast, many European systems feature tax-funded or social insurance universal coverage. Findings are therefore interpreted with attention to payment and incentive context, and international implementations are summarized to assess transferability (United Kingdom, Singapore, India) (OECD, 2023).

2. LITERATURE REVIEW

2.1. Model Performance and Calibration

Predictive analytics in health systems often ship with strong discrimination but weak calibration, which can mislead bedside decisions; methodologists argue for routine calibration diagnostics and recalibration during external validation (Van Calster *et al.*, 2019). In addition to AUC, decision-curve analysis should be used to measure clinical utility so that operational thresholds show real action rates (Vickers *et al.*, 2019). High-profile external validations underscore the gap between marketing claims and real-world performance; for example, the Epic Sepsis Model exhibited poor discrimination and calibration at a large academic center, challenging its widespread adoption (Wong *et al.*, 2021).

2.2. Ethics, Equity, and Trust

Trust and safety now rely on transparent documentation and bias monitoring. “Model Cards” outline the intended use, data lineage, subgroup metrics, and update frequency (Mitchell *et al.*, 2019), whereas “AI Factsheets” offer service-level provenance and assurance checks (Arnold *et al.*, 2019). Target–proxy mismatches can encode structural inequity at scale; a landmark study demonstrated that optimizing on costs rather than health needs systematically disadvantaged Black patients, motivating continuous subgroup dashboards and threshold audits (Obermeyer *et al.*, 2019).

2.3. Technical Architecture for Integration

Operational integration matters as much as algorithm choice. Event-driven delivery through FHIR® Subscriptions enables auditable, low-latency triggers that connect scores to concrete actions in clinical workflows (HL7 FHIR Subscriptions R5 Backport IG) (HL7 International, 2023, 2024). For multi-site learning without raw-data pooling, federated learning has emerged as a viable pattern provided sites perform local calibration and enforce secure aggregation (Rieke *et al.*, 2020).

Recent evidence supports a practical approach: prioritize calibration and net-benefit reporting, link models to event-driven workflows with clear decision rights, openly document limits and subgroup behavior, and use privacy-preserving training to increase data diversity while keeping local fit.

3. METHODOLOGY

A narrative literature review focused on real-world implementations of predictive analytics in health systems (2015–2025). Searches were executed across PubMed/MEDLINE, Scopus, and Web of Science. Keywords combined predictive modeling terms (“predictive analytics,” “machine learning,” “risk prediction,” “time-series,” “EHR,” “imaging,” “multimodal”) with operational/market terms (“patient flow,” “demand

forecasting," "hospital operations," "market intelligence," "patient engagement," "scheduling," "staffing"). Filters: English, humans, peer-reviewed or official organizational reports; exclusions: editorials, opinion pieces, vendor advertisements/whitepapers, and studies lacking explicit methods or outcomes. Titles/abstracts were screened for deployments with measurable clinical or operational effects (e.g., mortality, functional outcomes, time-to-treatment, length of stay, throughput, no-show reduction). Heterogeneity in designs and outcomes precluded meta-analysis; a narrative synthesis was used to derive common technical and organizational patterns. To assess generalizability beyond the U.S., a targeted search captured international implementations (UK NHS stroke AI, Singapore NUHS C3 command center, India TB CAD-CXR triage) with documented outcomes. Figures and tables consolidate the integration blueprint, real-time pipeline, and deployment evidence.

4. RESULTS AND DISCUSSION

4.1. Model selection & compute trade-offs

Tabular EHR prediction (readmission, no-show, utilization) often favors gradient-boosting trees for a strong accuracy-to-complexity ratio and CPU-level inference, while deep models rarely dominate after careful tuning. Recent healthcare-focused benchmarking and reviews report boosted trees outperforming or matching deep networks on large tabular cohorts, with simpler deployment and lower compute (Borisov *et al.*, 2024; Kowsar *et al.*, 2023). By contrast, time-series tasks (ICU deterioration, telemetry, streaming vitals) and imaging tasks (CXR/CT triage) benefit from sequence or convolutional architectures optimized for temporal or spatial structure; hybrid models that fuse structured EHR with signals and notes show promise but require careful alignment and more GPU resources (Patharkar *et al.*, 2024; Wang *et al.*, 2024).

Latency budgets shape architectural choices. Batch-scored, next-day risk lists (e.g., readmission outreach) tolerate heavier models and feature engineering. Real-time alarms (<1–5 min end-to-end) typically demand lightweight feature extraction, low-variance models, and efficient serving to maintain throughput and limit alert delays, especially under bursty event loads. Implementation playbooks from large systems underline the need to design according to the available data path and to prefer models that can be stably served where data actually land (Kawamoto *et al.*, 2023).

4.2. Preprocessing & multimodal alignment

Robust preprocessing addresses four recurring realities: (a) missingness (sporadic labs, sparse vitals), (b) timestamp irregularity (charting delays, order/result asynchrony), (c) label noise (billing vs clinical definitions), and (d) site heterogeneity (coding practices). Systematic reviews list these risks and suggest regular checks of data quality with clear logs and change tracking (Lewis *et al.*, 2023; Syed *et al.*, 2023). Temporal data benefit from bucketing and windowed features (trends, deltas, slope, volatility) or sequence models that consume raw trajectories when latency allows (Patharkar *et al.*, 2024).

For clinical text, assertion status and negation materially affect labels and features; hybrid pipelines that stack a NegEx-style layer with a transformer encoder improve robustness and portability across sites and languages (Argüello-González *et al.*, 2023; van Es *et al.*, 2023).

Multimodal fusion increases coverage but complicates alignment. Recent scoping and fusion studies in healthcare evaluate late-fusion (per-modality models with downstream combiner), intermediate-fusion (shared representation), and attention-based fusion with modality dropout to tolerate missing channels; these architectures can outperform single-modality baselines in prospective validations when alignment is correct (Ben-Miled *et al.*, 2025; Krones *et al.*, 2025).

Operational readiness requires feature registries with schema versioning and tests for data drift, as well as idempotent transformations to make replay possible for audits. When inference must happen close to the EHR, projects often pare features to those reliably populated in near-real-time feeds, deferring heavier feature engineering to nightly jobs (Kawamoto *et al.*, 2023).

4.3. Real-time/event-driven architecture

Event-driven delivery turns predictions into actions. Modern EHRs can emit FHIR® Subscriptions to push resource changes (e.g., Observation, Encounter) to downstream systems. The R5 Subscriptions Backport IG enables R4 servers to support topic-based events with standardized payloads, creating portable triggers for streaming pipelines (HL7 FHIR Subscriptions Backport) (HL7 International, 2023).

A reference pipeline contains: (1) an event bus (e.g., EHR → Subscriptions → gateway), (2) a stream processor that joins events with a feature store and enforces idempotency and back-pressure, (3) model-serving with tight SLAs and shadow-mode capability, (4) an alert policy that rate-limits, batches, or suppresses duplicates, and (5) delivery to workflow surfaces (inbasket, secure messaging, huddles) with closed-loop acknowledgment. Integration guidance from large health-system deployments emphasizes building to actual data paths, proving end-to-end latency with load tests, and versioning both features and policies so alerts remain auditable (Kawamoto *et al.*, 2023).

Batch vs. stream. Batch is simpler and cost-efficient for list-based interventions (e.g., next-day outreach, schedule optimization). Streams are preferred when time-to-action matters (deterioration, stroke code coordination, ED boarding thresholds). Resource planning follows the model family: CPU-bound gradient boosting often fits within existing app servers; GPU budgets are reserved for image/sequence inference or multimodal fusion. A pragmatic rule is to minimize dependency length between event and action; fewer joins yield lower tail latency and fewer failure modes.

4.4. Explainability & clinician trust

Trust grows when effects are observable and explanations are task-appropriate. SHAP or permutation-based global summaries, when paired with patient-level rationales, help clinicians anticipate alert "failure modes" in tabular EHR models; in image tasks, rigorous evaluation must accompany saliency to prevent over-trusting heatmaps. Modern reviews combine what works and what should be avoided (Alkhanbouli

et al., 2025; Sadeghi *et al.*, 2024).

Documentation also matters. Model Cards and AI Factsheets are maturing into procurement-grade artifacts that record intended use, data lineage, performance (including subgroup results), update cadence, and support/monitoring obligations (Arnold *et al.*, 2019; Mitchell *et al.*, 2019). The Coalition for Health AI (CHAI) is an example of a health-sector effort that aims to standardize pre-deployment testing, post-deployment monitoring, and buyer-facing transparency. It does this by providing a harmonized Blueprint and a programmatic vision for assurance labs and a model-card registry (Olsen, 2024; Shah *et al.*, 2024).

Critically, evaluation should connect explanations to decisions. Threshold-aware decision-curve analysis clarifies whether acting on predictions yields net benefit at operationally relevant action rates; combining decision curves with prospective calibration plots and subgroup dashboards provides a governance-ready evidence package (Collins *et al.*, 2024; Van Calster *et al.*, 2019; Vickers *et al.*, 2019).

4.5. Federated learning & privacy

Multi-institution generalization often falters because data cannot move, schemas differ, and governance slows cross-site pooling. Federated learning (FL) enables decentralized training and aggregation without raw-data sharing; reviews in clinical AI outline architectures, convergence considerations, and operational pitfalls (Rieke *et al.*, 2020). Complementary privacy-preserving techniques, differential privacy (DP) to bound leakage, homomorphic encryption (HE) or secure enclaves to protect gradients/updates, are maturing for health data; medical-imaging–focused overviews summarize attack surfaces and mitigations (Kaissis *et al.*, 2020). Recent work argues for explicit DP budgeting in medical models and shows operating regimes where privacy costs for performance are small enough to justify default use (Ziller *et al.*, 2024). Practical guidance in health care emphasizes combining FL with secure aggregation, formal data-use agreements, site-level calibration checks, and drift dashboards; surveys consolidate legal and security considerations for deployment (Pati *et al.*, 2024). For HE inference, computation remains non-trivial, but quantization and scheme choices (e.g., TFHE-style gates) are improving feasibility for select workloads (Selvakumar & Senthilkumar, 2025).

4.6. Evaluation beyond accuracy (calibration & decision-curves)

Deployment decisions hinge on well-calibrated risks at actionable thresholds. Calibration has been labeled the "Achilles heel" of prediction in medicine; guidance details how to measure and improve it (Van Calster *et al.*, 2019; Vickers *et al.*, 2019). A tutorial for clinical informatics clarifies connections among the Brier score, calibration-in-the-large (intercept), calibration slope, and reliability plots, and recommends refitting or Platt/Isotonic recalibration when local drift appears (Huang *et al.*, 2020). Threshold-aware decision-curve analysis (DCA) then quantifies net benefit versus "treat-all/none," aligning evaluation to operational action rates (Vickers *et al.*, 2019; Vickers & Holland, 2021). Recent deployment reports illustrate prospective calibration plots and equity dashboards (performance by age, sex, race/ethnicity, and deprivation), enabling oversight committees to adjust thresholds or retrain (Liou *et al.*, 2024). Minimum reporting for integrated rollouts should include (i) discrimination (AUROC/PR-AUC), (ii) calibration intercept/slope + reliability plots, (iii) DCA across the intended action-rate band, (iv) false-positive/alert rates per 100 patients/day, and (v) subgroup performance with pre-specified disparity bounds.

4.7. International implementations

United Kingdom (stroke AI). Multi-center experience with e-Stroke / Brainomix 360 reports improved pathway speed and treatment access; an observational study from England describes process-metric gains and better functional outcomes after deployment, while program summaries highlight shorter transfer times and higher independence at follow-up (Nagaratnam *et al.*, 2024). These results illustrate how imaging AI plus workflow standardization can shift door-in-door-out and treatment rates in a tax-funded system.

Singapore (NUHS C3/Endeavour AI). The National University Health System operates a command-center model that fuses NGEMR feeds with forecasting to project bed availability up to two weeks ahead, coordinating staffing and flow; academic and national-program sources detail the platform and EMR backbone that enable near-real-time operation (The Straits Times, 2022).

India (TB CAD-CXR triage). Following WHO endorsement of CAD for TB screening/triage, implementation studies in India show measurable yield increases when CAD supports CXR screening in active-case-finding programs; one mixed-methods deployment attributed ~15–16% additional TB case yield to AI-flagged images not deemed presumptive by human readers (Ridhi *et al.*, 2024; Vijayan *et al.*, 2023; WHO, 2025). Comparative evaluations across CAD products corroborate robust accuracy at practical thresholds (Codlin *et al.*, 2021).

Transferability note. These cases indicate portability of the integration framework, event-driven data, explicit alert/activation pathways, and calibration monitoring across a tax-funded NHS, a command-centered Asian academic cluster, and LMIC TB programs with donor procurement pathways; incentive structures differ, so operational levers (bed flow vs. screening yield) and KPIs must be localized.

4.8. Governance structures and regulatory frameworks

Formal AI governance sets decision rights, documentation standards, monitoring cadence, and rollback authority. The Coalition for Health AI has codified procurement-grade artifacts, including Applied Model Cards that specify intended use, data lineage, performance metrics by subgroup, and update frequency; CHAI is piloting assurance laboratories to operationalize pre-deployment testing and post-deployment monitoring (Coalition for Health AI, 2024b; Shah *et al.*, 2024).

Regulatory momentum is converging on algorithmic transparency. The U.S. ONC's HTI-1 final rule requires certified EHRs to expose metadata about decision support interventions, including data provenance and performance characteristics, establishing infrastructure for systematized oversight

(Department of Health and Human Services, 2024). National guidance on optimizing clinical decision support emphasizes aligning interventions to actual workflows and specifying responsible parties for maintenance, threshold adjustments, and updates (Tcheng *et al.*, 2017).

Implementation reports indicate that successful deployments bind authority (who can change thresholds or silence alerts), accountability (performance KPIs), and assurance (calibration and equity dashboards) within chartered oversight committees with explicit escalation and rollback procedures.

4.9. Adoption patterns and workflow integration

Empirical studies of clinical decision support document how alert fatigue rises with high workload, repetitive prompts, and poor timing, eroding acceptance rates and clinical impact (Ancker *et al.*, 2017). Design heuristics derived from the "five rights" of CDS right information, person, format, channel, and time remain practical for predictive tools, emphasizing concise content, appropriate recipients, and minimal disruption (Douthit *et al.*, 2020).

Two implementation patterns consistently improve adoption. First, shadow mode with silent scoring validates calibration and workload impact before go-live; second, champion networks normalize use and provide rapid feedback for threshold and wording adjustments (Tcheng *et al.*, 2017). Predicting alert acceptance using EHR telemetry and historical interactions can further improve signal-to-noise by selectively surfacing prompts likely to lead to action, a tactic shown to be feasible in multicenter studies (Baron *et al.*, 2021).

For imaging and high-stakes triage, closed-loop tasking, where alerts require acknowledgment and record time-to-action, supports accountability and measurable process gains, mirroring hospital command center practice for coordinating bed flow and escalation (Johnson *et al.*, 2024). Figure 1 illustrates the clinician-facing component of the integration framework, detailing how alerts flow through priority triage to clinical decision points and subsequent actions. The schematic highlights the bidirectional nature of the system, where clinical actions generate performance metrics that feed back to threshold adjustments and policy updates.

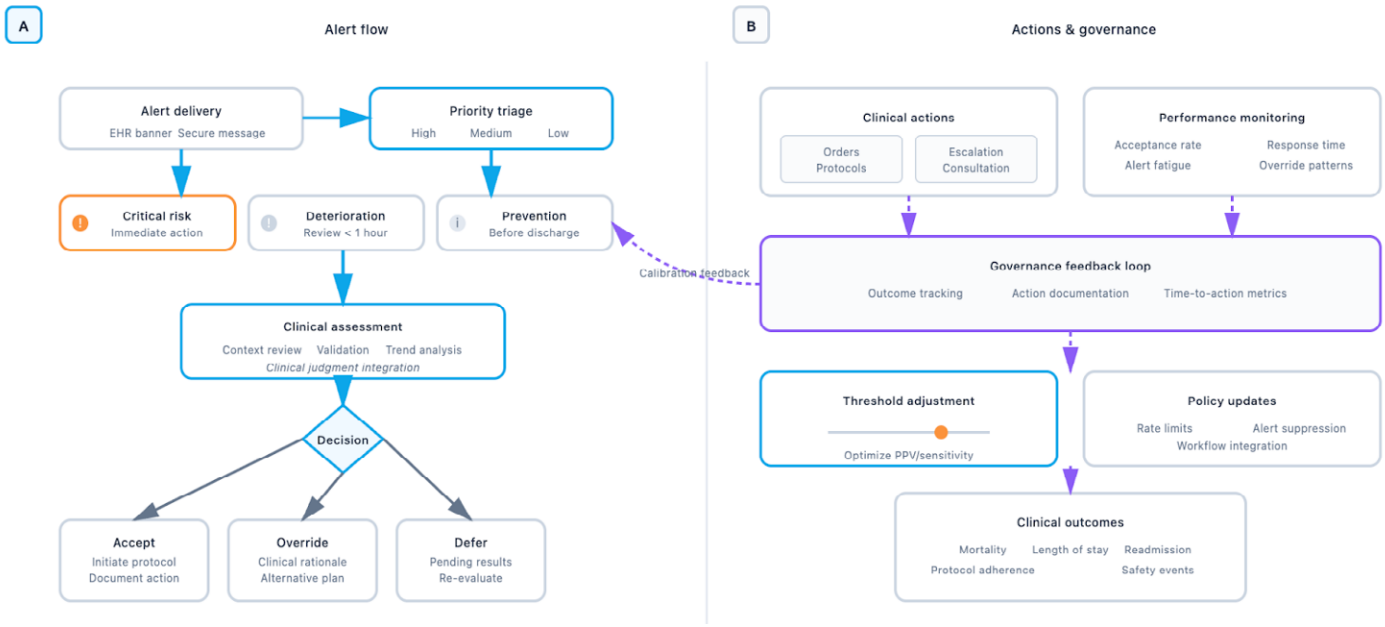

Figure 1. Real-time alert flow and governance feedback loop

4.10. Economic models and sustainability

Economic defensibility depends on alignment with payment structures and operational constraints. In U.S. settings, the Hospital Readmissions Reduction Program imposes penalties up to 3% of base Medicare payments for excess 30-day readmissions, creating a budgetary rationale for deploying readmission-reduction analytics integrated with transitional-care workflows (Jordan, 2021; U.S. Centers for Medicare & Medicaid Services, 2012). Early-warning programs paired with protocolized responses have reported lower mortality and costs when implemented with reliable data capture and nurse-driven activation, demonstrating measurable value when analytics connect to disciplined pathways (Jones *et al.*, 2015).

System-level command centers can smooth patient flow and staffing by forecasting demand and orchestrating actions; mixed-methods evaluations of NHS AI-enabled command centers detail organizational benefits improved situational awareness, faster escalation and adoption challenges, including tensions between centralized coordination and ward-level autonomy (Johnson *et al.*, 2024; Mebrahtu *et al.*, 2023).

Sustainability also depends on compute and support costs. CPU-servable tabular models typically fit within existing infrastructure, whereas GPU-bound imaging or multimodal pipelines require capacity planning and benefit from batching strategies and service-level objectives to cap tail latency. Benchmarking surveys of command centers and CDS programs recommend explicit operational KPIs, throughput, avoided boardings, and action rates that connect model performance

to operational value (Franklin *et al.*, 2023; Tcheng *et al.*, 2017). In universal-coverage systems, returns manifest as throughput and access improvements rather than revenue generation; NHS mixed-methods research demonstrates feasibility and adoption impacts but emphasizes the necessity for rigorous causal assessment to attribute outcomes amidst real-world confounders, including pandemics and concurrent initiatives (Johnson *et al.*, 2024).

Table 1. Integrated deployments methods, settings, outcomes

Use case	Inputs	Model family	Setting	Outcome	Validation	Latency/ compute	Explainability	Bias/notes
Hyperacute stroke AI (Nagaratnam *et al.*, 2024)	NCCT/CTA imaging + pathway data	CNN + rules	NHS regional networks	Faster transfers; higher treatment access; better mRS distribution	Observational, multi-site	Real-time; GPU for imaging	Saliency + protocol checks	Outcome registry linkage
Bed-flow forecasting (The Straits Times, 2022)	NGEMR events, notes, ADT	Gradient-boosting + time-series	NUHS command center	Bed availability forecasts (≤14 days); smoother staffing	Prospective ops metrics	Near-real-time; CPU-first	Global feature importances	EMR backbone critical
TB CXR CAD triage (WHO, 2025)	Digital CXR; ACF workflows	CNN CAD score	India ACF programs	+15–16% TB yield attributable to AI flags	Prospective program eval	On-device/ edge feasible	Threshold + heatmap	WHO-aligned QA required
Readmission prevention	EHR risk + care-path triggers	GBM/ logistic	U.S. integrated system	Lower 30-day readmissions; targeted outreach	Observational, system-wide	Batch daily; CPU	Model card + SHAP	Equity monitoring advised

4.11. Discussion

This review aimed to answer three questions: which integration patterns improve outcomes, which conditions enable scale, and how transferable are U.S. findings to other healthcare systems? The evidence reveals that impact follows execution discipline rather than algorithmic novelty. Predictive models improve care and operations when they function as managed programs, are calibrated, are embedded in event-driven workflows, are governed by explicit decision rights, and are evaluated for threshold-aware net benefit and equity, not as standalone algorithms.

4.11.1. Integration as the Core Mechanism

The findings demonstrate that technical sophistication without organizational integration produces shelfware. Models with strong discrimination but weak calibration mislead decisions; real-time alerts without clear ownership create noise; governance without enforcement permits drift. Conversely, even modest algorithms produce measurable gains when welded to disciplined pathways: stroke imaging AI accelerates treatment when tied to transfer protocols, bed-flow forecasting smooths staffing when command centers hold decision rights, and TB screening AI increases yield when workflow mandates flag review.

The integration framework that emerges from the evidence spans five interdependent layers: data architecture (feature stores, schema versioning, drift detection), model selection (balancing accuracy, complexity, and compute under latency constraints), real-time delivery (event-driven pipelines with FHIR Subscriptions and resilient stream processing), governance (decision rights, calibration monitoring, equity dashboards, model cards), and evaluation (calibration diagnostics, decision-curve analysis, subgroup reporting). Success requires coherence across all layers; weakness in any dimension undermines the system.

4.11.2. What Enables Scalability

The conditions necessary for scalable deployment can be grouped into technical and organizational requirements. Technically, systems need robust data plumbing, idempotent transformations, feature registries, event buses that tolerate bursty loads, and architecture matched to latency budgets: CPU-first gradient boosting for tabular batch jobs, GPU-backed sequence models for real-time streams. Organizationally, scale depends on aligned incentives (payment penalties that justify investment and throughput KPIs that drive action), explicit decision rights (who adjusts thresholds and who rolls back), and legitimacy built through shadow mode, champion networks, and closed-loop acknowledgment.

Equity and calibration monitoring are not optional extras; they are necessary for scalability. Drift degrades decision value silently; subgroup disparities erode trust and can encode structural harm at the population scale. Deployments that institutionalize prospective calibration checks, reliability plotting, and threshold-aware equity dashboards maintain validity and legitimacy over time. The regulatory and sector movements toward standardized documentation (Model Cards, assurance labs, and HTI-1 metadata requirements) create infrastructure for repeatable governance across sites.

4.11.3. Transferability Across Systems

The international implementations, NHS stroke AI, Singapore NUHS command center, and India TB CAD, indicate that the integration framework transfers across distinct financing and delivery models. What are the changes in the operational levers and KPIs? U.S. readmission analytics optimizes against HRRP penalties, NHS stroke pathways optimize for access

and functional outcomes, and Indian TB programs optimize screening yield in resource-limited settings. The socio-technical principles remain constant: calibrated models, event-driven delivery, clear pathways, and governance with authority.

However, transferability is not automatic. Local calibration is essential when populations, practice patterns, or data capture differ. Incentive structures must be redesigned: what works under fee-for-service may not align with capitated or tax-funded budgets. Command-center coordination that suits an integrated delivery network may clash with federated public hospitals lacking shared governance. The lesson is to localize levers and KPIs while standardizing the integration architecture.

4.11.4. Theoretical Contributions

From a theoretical standpoint, this review advances several propositions. First, decision value arises from integration, not discrimination alone: high AUROC without calibration and actionable thresholds does not improve outcomes. Decision-curve analysis operationalizes this concept by quantifying net benefit at clinically relevant action rates, shifting evaluation from model performance to decision impact.

Second, calibration is a dynamic property, not a static metric: models degrade as populations and practices evolve, so continuous monitoring with intercept/slope checks and reliability plots is essential for sustained utility. Third, assurance practices are converging toward standardized documentation and testing: Applied Model Cards, assurance labs, and regulatory metadata requirements create repeatable governance mechanisms that support procurement, deployment, and oversight across sites.

Fourth, privacy-preserving multi-site learning is feasible but requires local accountability: federated learning enables training on distributed data, but sites must retain calibration checks and subgroup reporting responsibilities to maintain local validity and equity. Finally, event-driven architecture aligns analytics with operational tempo: standards like FHIR Subscriptions enable low-latency triggers that connect predictions to actions, making models operational rather than informational.

4.11.5. Policy Implications

For health policy, the findings point to three priorities. First, transparency and auditability: regulations such as ONC HTI-1, which mandate public metadata for algorithmic decision support, enable systematic oversight. Sector coalitions building assurance labs and standardized "nutrition labels" complement regulation by providing buyer-facing evidence packages even outside formal compliance frameworks.

Second, payment incentives affect adoption: programs like HRRP impose material penalties that make predictive outreach investments worthwhile when evaluation shows a net benefit. Policy design should align financial incentives with measurable clinical and operational gains, not just deployment. In universal-coverage systems, policy value is based on throughput, access, and safety. Mixed-methods evaluations of command centers show that they can improve operations, but they also show that policy needs to deal with adoption barriers, workflow disruption, and autonomy tensions through change management and governance design.

Third, equity monitoring must be institutionalized: subgroup dashboards and threshold audits should be mandatory for deployment, not optional enhancements. Target-proxy mismatches can encode structural inequity at scale; continuous monitoring with pre-specified disparity bounds and corrective action protocols protects populations and sustains trust.

5. CONCLUSION

Predictive analytics improves care and operations when embedded as a managed program: calibrated models wired into event-driven workflows, governed by explicit decision rights, and evaluated for threshold-aware net benefit and equity. The proposed framework puts that principle into practice by ensuring data readiness and feature stewardship, selecting models appropriate for the modality while considering compute and latency constraints, delivering real-time results through standardized subscriptions and resilient stream processing, and establishing governance based on transparent documentation, prospective calibration, and assurance assets. International cases indicate transferability across distinct financing and delivery models when operational levers and KPIs are localized. The practical message is straightforward: impact follows execution discipline. Health systems, payers, and vendors that treat predictive analytics as socio-technical infrastructure, rather than isolated algorithms, are more likely to achieve durable improvements in patient outcomes, flow, and sustainability.

RECOMMENDATIONS

The evidence base would benefit from four research directions. First, prospective multi-site impact trials with pre-registered calibration and decision-curve endpoints should replace single-site retrospective reports. Where randomization is feasible, cluster or stepped-wedge designs can isolate effects of predictive interventions embedded in workflows, as demonstrated in emergency and primary care CDS evaluations.

Second, comparative architecture trials should test batch versus streaming pipelines and CPU-first versus GPU-dependent serving under varied latency and volume regimes, quantifying trade-offs in tail latency, alert rates, and action timeliness. This would provide evidence-based guidance for infrastructure investment.

Third, implementation science should study governance mechanics: how do decision rights, update cadences, and equity dashboards function in practice? Emerging applied model cards and assurance-lab artifacts can serve as standardized interventions, enabling rigorous evaluation of governance models across contexts.

Finally, privacy-preserving learning research should quantify the performance-privacy frontier in federated settings and produce operational playbooks for secure aggregation, differential privacy budgeting, and site-level calibration before global updates. This work would enable trustworthy multi-site learning at scale.

REFERENCES

Alkhanbouli, R., Matar Abdulla Almadhaani, H., Alhosani, F., & Simsekler, M. C. E. (2025). The role of explainable

artificial intelligence in disease prediction: A systematic literature review and future research directions. *BMC Medical Informatics and Decision Making, 25*, 110. https://doi.org/10.1186/s12911-025-02944-6

Ancker, J. S., Edwards, A., Nosal, S., Hauser, D., Mauer, E., Kaushal, R., & with the HITEC Investigators. (2017). Effects of workload, work complexity, and repeated alerts on alert fatigue in a clinical decision support system. *BMC Medical Informatics and Decision Making, 17*(1), 36. https://doi.org/10.1186/s12911-017-0430-8

Argüello-González, G., Aquino-Esperanza, J., Salvador, D., Bretón-Romero, R., Del Río-Bermudez, C., Tello, J., & Menke, S. (2023). Negation recognition in clinical natural language processing using a combination of the NegEx algorithm and a convolutional neural network. *BMC Medical Informatics and Decision Making, 23*(1), 216. https://doi.org/10.1186/s12911-023-02301-5

Arnold, M., Bellamy, R. K. E., Hind, M., Houde, S., Mehta, S., Mojsilovic, A., Nair, R., Ramamurthy, K. N., Olteanu, A., Piorkowski, D., Reimer, D., Richards, J., Tsay, J., & Varshney, K. R. (2019). FactSheets: Increasing trust in AI services through supplier's declarations of conformity. *IBM Journal of Research and Development, 63*(4/5), 6:1-6:13. https://doi.org/10.1147/JRD.2019.2942288

Baron, J. M., Huang, R., McEvoy, D., & Dighe, A. S. (2021). Use of machine learning to predict clinical decision support compliance, reduce alert burden, and evaluate duplicate laboratory test ordering alerts. JAMIA Open, 4(1). https://doi.org/10.1093/jamiaopen/ooab006

Ben-Miled, Z., Shebesh, J. A., Su, J., Dexter, P. R., Grout, R. W., & Boustani, M. A. (2025). Multi-Modal Fusion of Routine Care Electronic Health Records (EHR): A Scoping Review. *Information, 16*(1), 54. https://doi.org/10.3390/info16010054

Borisov, V., Leemann, T., Seßler, K., Haug, J., Pawelczyk, M., & Kasneci, G. (2024). Deep Neural Networks and Tabular Data: A Survey. *IEEE Transactions on Neural Networks and Learning Systems, 35*(6), 7499–7519. https://doi.org/10.1109/TNNLS.2022.3229161

Coalition for Health AI. (2024a, October 18). *CHAI Advances Assurance Lab Certification and 'Nutrition Label' for Health AI | CHAI.* Coalition for Health AI. https://www.chai.org/blog/chai-advances-assurance-lab-certification-and-nutrition-label-for-health-ai

Coalition for Health AI. (2024b, November 22). *The CHAI Applied Model Card Complete Documentation.* Coalition for Health AI. https://docs.google.com/document/d/10zXD1-8xbkH7WeplRY86-QEvJeGgzqHfvolhA07uYEs/edit?ouid=101975905259229057250&usp=docs_home&ths=true&usp=embed_facebook

Codlin, A. J., Dao, T. P., Vo, L. N. Q., Forse, R. J., Van Truong, V., Dang, H. M., Nguyen, L. H., Nguyen, H. B., Nguyen, N. V., Sidney-Annerstedt, K., Squire, B., Lönnroth, K., & Caws, M. (2021). Independent evaluation of 12 artificial intelligence solutions for the detection of tuberculosis. *Scientific Reports, 11*(1), 23895. https://doi.org/10.1038/s41598-021-03265-0

Collins, G. S., Moons, K. G. M., Dhiman, P., Riley, R. D., Beam, A. L., Calster, B. V., Ghassemi, M., Liu, X., Reitsma, J. B., Smeden, M. van, Boulesteix, A.-L., Camaradou, J. C., Celi, L. A., Denaxas, S., Denniston, A. K., Glocker, B., Golub, R. M., Harvey, H., Heinze, G., … Logullo, P. (2024). TRIPOD+AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ, 385*, e078378. https://doi.org/10.1136/bmj-2023-078378

Davis, S. E., Greevy, R. A., Lasko, T. A., Walsh, C. G., & Matheny, M. E. (2020). Detection of calibration drift in clinical prediction models to inform model updating. *Journal of Biomedical Informatics, 112*, 103611. https://doi.org/10.1016/j.jbi.2020.103611

Department of Health and Human Services. (2024, January 9). *Health Data, Technology, and Interoperability: Certification Program Updates, Algorithm Transparency, and Information Sharing.* Federal Register. https://www.federalregister.gov/documents/2024/01/09/2023-28857/health-data-technology-and-interoperability-certification-program-updates-algorithm-transparency-and

Douthit, B. J., Musser, R. C., Lytle, K. S., & Richesson, R. L. (2020). A Closer Look at the "Right" Format for Clinical Decision Support: Methods for Evaluating a Storyboard BestPractice Advisory. *Journal of Personalized Medicine, 10*(4), 142. https://doi.org/10.3390/jpm10040142

Franklin, B. J., Yenduri, R., Parekh, V. I., Fogerty, R. L., Scheulen, J. J., High, H., Handley, K., Crow, L., & Goralnick, E. (2023). Hospital Capacity Command Centers: A Benchmarking Survey on an Emerging Mechanism to Manage Patient Flow. *Joint Commission Journal on Quality and Patient Safety, 49*(4), 189–198. https://doi.org/10.1016/j.jcjq.2023.01.007

Gold, R., Larson, A. E., Sperl-Hillen, J. M., Boston, D., Sheppler, C. R., Heintzman, J., McMullen, C., Middendorf, M., Appana, D., Thirumalai, V., Romer, A., Bava, J., Davis, J. V., Yosuf, N., Hauschildt, J., Scott, K., Moore, S., & O'Connor, P. J. (2022). Effect of Clinical Decision Support at Community Health Centers on the Risk of Cardiovascular Disease: A Cluster Randomized Clinical Trial. *JAMA Network Open, 5*(2), e2146519. https://doi.org/10.1001/jamanetworkopen.2021.46519

HL7 International. (2023, January 11). *Subscriptions R5 Backport.* HL7 International. https://hl7.org/fhir/uv/subscriptions-backport/STU1.1/?utm_source=chatgpt.com

HL7 International. (2024, June 17). *Home—Subscriptions R5 Backport v1.2.0-ballot.* Health Level Seven Internation. https://build.fhir.org/ig/HL7/fhir-subscription-backport-ig/

Huang, Y., Li, W., Macheret, F., Gabriel, R. A., & Ohno-Machado, L. (2020). A tutorial on calibration measurements and

calibration models for clinical prediction models. *Journal of the American Medical Informatics Association, 27*(4), 621–633. https://doi.org/10.1093/jamia/ocz228

IBM. (2015, October 1). *Using AI Factsheets for AI Governance—Docs*. IBM Cloud Pak for Data; IBM Corporation. https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/factsheets-model-inventory.html

Jenkins, D. A., Martin, G. P., Sperrin, M., Riley, R. D., Debray, T. P. A., Collins, G. S., & Peek, N. (2021). Continual updating and monitoring of clinical prediction models: Time for dynamic prediction systems? *Diagnostic and Prognostic Research, 5*(1), 1. https://doi.org/10.1186/s41512-020-00090-3

Johnson, O. A., McCrorie, C., McInerney, C., Mebrahtu, T. F., Granger, J., Sheikh, N., Lawton, T., Habli, I., Randell, R., & Benn, J. (2024). Implementing an artificial intelligence command centre in the NHS: A mixed-methods study. *Health and Social Care Delivery Research*, 1–108. https://doi.org/10.3310/TATM3277

Jones, S. L., Ashton, C. M., Kiehne, L., Gigliotti, E., Bell-Gordon, C., Disbot, M., Masud, F., Shirkey, B. A., & Wray, N. P. (2015). Reductions in Sepsis Mortality and Costs After Design and Implementation of a Nurse-Based Early Recognition and Response Program. *Joint Commission Journal on Quality and Patient Safety / Joint Commission Resources, 41*(11), 483–491. https://doi.org/10.1016/s1553-7250(15)41063-3

Jordan, R. (2021, November 4). *10 Years of Hospital Readmissions Penalties*. KFF. https://www.kff.org/affordable-care-act/slide/10-years-of-hospital-readmissions-penalties/

Kaissis, G. A., Makowski, M. R., Rückert, D., & Braren, R. F. (2020). Secure, privacy-preserving and federated machine learning in medical imaging. *Nature Machine Intelligence, 2*(6), 305–311. https://doi.org/10.1038/s42256-020-0186-1

Kawamoto, K., Finkelstein, J., & Fiol, G. D. (2023). Implementing Machine Learning in the Electronic Health Record: Checklist of Essential Considerations. *Mayo Clinic Proceedings, 98*(3), 366–369. https://doi.org/10.1016/j.mayocp.2023.01.013

Kowsar, I., Rabbani, S. B., Akhter, K. F. B., & Samad, M. D. (2023). Deep Clustering of Electronic Health Records Tabular Data for Clinical Interpretation. ... ... *IEEE International Conference on Telecommunications and Photonics. IEEE International Conference on Telecommunications and Photonics*, 2023, 10.1109/ictp60248.2023.10490723. https://doi.org/10.1109/ictp60248.2023.10490723

Krones, F., Marikkar, U., Parsons, G., Szmul, A., & Mahdi, A. (2025). Review of multimodal machine learning approaches in healthcare. *Information Fusion, 114*, 102690. https://doi.org/10.1016/j.inffus.2024.102690

Lewis, A. E., Weiskopf, N., Abrams, Z. B., Foraker, R., Lai, A. M., Payne, P. R. O., & Gupta, A. (2023). Electronic health record data quality assessment and tools: A systematic review. *Journal of the American Medical Informatics Association, 30*(10), 1730–1740. https://doi.org/10.1093/jamia/ocad120

Mebrahtu, T. F., McInerney, C. D., Benn, J., McCrorie, C., Granger, J., Lawton, T., Sheikh, N., Habli, I., Randell, R., & Johnson, O. (2023). The impact of hospital command centre on patient flow and data quality: Findings from the UK National Health Service. *International Journal for Quality in Health Care, 35*(4). https://doi.org/10.1093/intqhc/mzad072

Melnick, E. R., Nath, B., Dziura, J. D., Casey, M. F., Jeffery, M. M., Paek, H., Soares, W. E., Hoppe, J. A., Rajeevan, H., Li, F., Skains, R. M., Walter, L. A., Patel, M. D., Chari, S. V., Platts-Mills, T. F., Hess, E. P., & D'Onofrio, G. (2022). User centered clinical decision support to implement initiation of buprenorphine for opioid use disorder in the emergency department: EMBED pragmatic cluster randomized controlled trial. *BMJ, 377*, e069271. https://doi.org/10.1136/bmj-2021-069271

Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model Cards for Model Reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229. https://doi.org/10.1145/3287560.3287596

Nagaratnam, K., Neuhaus, A., Briggs, J. H., Ford, G. A., Woodhead, Z. V. J., Maharjan, D., & Harston, G. (2024). Artificial intelligence-based decision support software to improve the efficacy of acute stroke pathway in the NHS: An observational study. *Frontiers in Neurology, 14*, 1329643. https://doi.org/10.3389/fneur.2023.1329643

Nong, P., Adler-Milstein, J., Apathy, N. C., Holmgren, A. J., & Everson, J. (2025). Current Use And Evaluation Of Artificial Intelligence And Predictive Models In US Hospitals: Article examines uses and evaluation of artificial intelligence and predictive models in US hospitals. *Health Affairs, 44*(1), 90–98. https://doi.org/10.1377/hlthaff.2024.00842

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science, 366*(6464), 447–453. https://doi.org/10.1126/science.aax2342

OECD. (2023). *Health at a Glance 2023: OECD Indicators*. OECD. https://doi.org/10.1787/7a7afb35-en

Olsen, E. (2024, October 18). *CHAI releases draft framework for AI quality assurance labs | Healthcare Dive*. Healthcare Dive. https://www.healthcaredive.com/news/coalition-health-ai-chai-quality-assurance-lab-nutrition-label-draft-frameworks/730225/

Patharkar, A., Cai, F., Al-Hindawi, F., & Wu, T. (2024). Predictive modeling of biomedical temporal data in healthcare applications: Review and future directions. *Frontiers in Physiology, 15*. https://doi.org/10.3389/fphys.2024.1386760

Pati, S., Kumar, S., Varma, A., Edwards, B., Lu, C., Qu, L., Wang, J. J., Lakshminarayanan, A., Wang, S., Sheller, M. J., Chang,

K., Singh, P., Rubin, D. L., Kalpathy-Cramer, J., & Bakas, S. (2024). Privacy preservation for federated learning in health care. *Patterns, 5*(7), 100974. https://doi.org/10.1016/j.patter.2024.100974

Ridhi, S., Robert, D., Soren, P., Kumar, M., Pawar, S., & Reddy, B. (2024). Comparing the Output of an Artificial Intelligence Algorithm in Detecting Radiological Signs of Pulmonary Tuberculosis in Digital Chest X-Rays and Their Smartphone-Captured Photos of X-Ray Films: Retrospective Study. *JMIR Formative Research, 8*(1), e55641. https://doi.org/10.2196/55641

Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. *Npj Digital Medicine, 3*(1), 119. https://doi.org/10.1038/s41746-020-00323-1

Sadeghi, Z., Alizadehsani, R., Cifci, M. A., Kausar, S., Rehman, R., Mahanta, P., Bora, P. K., Almasri, A., Alkhawaldeh, R. S., Hussain, S., Alatas, B., Shoeibi, A., Moosaei, H., Hladík, M., Nahavandi, S., & Pardalos, P. M. (2024). A review of Explainable Artificial Intelligence in healthcare. *Computers and Electrical Engineering, 118*, 109370. https://doi.org/10.1016/j.compeleceng.2024.109370

Selvakumar, S., & Senthilkumar, B. (2025). A privacy preserving machine learning framework for medical image analysis using quantized fully connected neural networks with TFHE based inference. *Scientific Reports, 15*(1), 27880. https://doi.org/10.1038/s41598-025-07622-1

Shah, N. H., Halamka, J. D., Saria, S., Pencina, M., Tazbaz, T., Tripathi, M., Callahan, A., Hildahl, H., & Anderson, B. (2024). A Nationwide Network of Health AI Assurance Laboratories. *JAMA, 331*(3), 245–249. https://doi.org/10.1001/jama.2023.26930

Syed, R., Eden, R., Makasi, T., Chukwudi, I., Mamudu, A., Kamalpour, M., Kapugama Geeganage, D., Sadeghianasl, S., Leemans, S. J. J., Goel, K., Andrews, R., Wynn, M. T., ter Hofstede, A., & Myers, T. (2023). Digital Health Data Quality Issues: Systematic Review. *Journal of Medical Internet Research, 25*, e42615. https://doi.org/10.2196/42615

Tcheng, J. E., Bakken, S., Bates, D. W., Bonner, H., Gandhi, T. K., Josephs, M., Kawamoto, K., Lomotan, E. A., Mackay, E., Middleton, B., Teich, J. M., Weingarten, S., Lopez, M. H., The Learning Health System Series, & National Academy of Medicine. (2017). *Optimizing Strategies for Clinical Decision Support: Summary of a Meeting Series* (p. 27122). National Academies Press. https://doi.org/10.17226/27122

The Straits Times. (2022, December 2). AI platform helps NUHS ease hospital bed crunch by predicting availability up to 2 weeks in advance. *The Straits Times*. https://www.straitstimes.com/singapore/health/ai-platform-helps-nuhs-ease-hospital-bed-crunch-by-predicting-availability-up-to-2-weeks-in-advance

U.S. Centers for Medicare & Medicaid Services. (2012). *Hosp. Readmission Reduction | CMS*. CMS.Gov. https://www.cms.gov/medicare/quality/value-based-programs/hospital-readmissions

Van Calster, B., McLernon, D. J., van Smeden, M., Wynants, L., Steyerberg, E. W., Bossuyt, P., Collins, G. S., Macaskill, P., McLernon, D. J., Moons, K. G. M., Steyerberg, E. W., Van Calster, B., van Smeden, M., Vickers, A. J., & On behalf of Topic Group 'Evaluating diagnostic tests and prediction models' of the STRATOS initiative. (2019). Calibration: The Achilles heel of predictive analytics. *BMC Medicine, 17*(1), 230. https://doi.org/10.1186/s12916-019-1466-7

van Es, B., Reteig, L. C., Tan, S. C., Schraagen, M., Hemker, M. M., Arends, S. R. S., Rios, M. A. R., & Haitjema, S. (2023). Negation detection in Dutch clinical texts: An evaluation of rule-based and machine learning methods. *BMC Bioinformatics, 24*(1), 10. https://doi.org/10.1186/s12859-022-05130-x

Vickers, A. J., & Holland, F. (2021). Decision curve analysis to evaluate the clinical benefit of prediction models. *The Spine Journal : Official Journal of the North American Spine Society, 21*(10), 1643–1648. https://doi.org/10.1016/j.spinee.2021.02.024

Vickers, A. J., van Calster, B., & Steyerberg, E. W. (2019). A simple, step-by-step guide to interpreting decision curve analysis. *Diagnostic and Prognostic Research, 3*(1), 18. https://doi.org/10.1186/s41512-019-0064-7

Vijayan, S., Jondhale, V., Pande, T., Khan, A., Brouwer, M., Hegde, A., Gandhi, R., Roddawar, V., Jichkar, S., Kadu, A., Bharaswadkar, S., Sharma, M., Vasquez, N. A., Richardson, L., Robert, D., & Pawar, S. (2023). Implementing a chest X-ray artificial intelligence tool to enhance tuberculosis screening in India: Lessons learned. *PLOS Digital Health, 2*(12), e0000404. https://doi.org/10.1371/journal.pdig.0000404

Wang, Y., Yin, C., & Zhang, P. (2024). Multimodal risk prediction with physiological signals, medical images and clinical notes. *Heliyon, 10*(5). https://doi.org/10.1016/j.heliyon.2024.e26772

WHO. (2025). Use of computer-aided detection so ware for tuberculosis screening. World Health Organization.

Wong, A., Otles, E., Donnelly, J. P., Krumm, A., McCullough, J., DeTroyer-Cooley, O., Pestrue, J., Phillips, M., Konye, J., Penoza, C., Ghous, M., & Singh, K. (2021). External Validation of a Widely Implemented Proprietary Sepsis Prediction Model in Hospitalized Patients. *JAMA Internal Medicine, 181*(8), 1065–1070. https://doi.org/10.1001/jamainternmed.2021.2626

Ziller, A., Mueller, T. T., Stieger, S., Feiner, L. F., Brandt, J., Braren, R., Rueckert, D., & Kaissis, G. (2024). Reconciling privacy and accuracy in AI for medical imaging. *Nature Machine Intelligence, 6*(7), 764–774. https://doi.org/10.1038/s42256-024-00858-y