

ORIGINAL RESEARCH

Open access

Federated Analytics Framework for Benchmarking Hospital Operational Performance Across Health Systems Using Secure Aggregation of Bed Occupancy, Discharge Delays, Staffing Ratios, and Service Demand Indicators

George Brown¹, Michael Taylor^{1*}, Sarah Wilson², Olivia Harris¹

Abstract

Hospitals lack an objective and privacy-preserving mechanism to compare operational performance against peer institutions. This limits shared learning around capacity, discharge flow, staffing, and service demand. Traditional benchmarking often depends on centralized data warehouses, voluntary reporting, or retrospective surveys. These approaches can create privacy, competitive, regulatory, and selection-bias concerns that discourage full participation. This article proposes a federated analytics framework for computing aggregate operational benchmarks without moving raw hospital data outside local institutional boundaries. The framework would support medians, percentiles, and risk-adjusted comparative indicators through secure aggregation. The framework combines a local data standardization engine, a secure multi-party computation aggregator, a differential privacy injector, and a participatory dashboard. Together, these components would allow each hospital to compare its position against anonymous peer distributions. The framework could enable hospitals to identify performance gaps in bed occupancy, discharge delays, staffing ratios, and service demand while preserving confidentiality. It would be expected to encourage more honest participation because institutional data sovereignty remains intact. A federated analytics approach offers a practical pathway for collaborative operations improvement across health systems. It aligns benchmarking, privacy protection, and organizational learning within a single governance-aware framework.

Keywords Hospital operations, Federated analytics, Secure aggregation, Differential privacy, Benchmarking, Bed occupancy

*Correspondence:

Michael Taylor
michael.taylor@outlook.com

¹ Department of Healthcare AI Systems and Analytics, Faculty of Medicine, University of Auckland, Auckland, New Zealand

² Department of Digital Clinical Informatics, Faculty of Engineering, University of Otago, Dunedin, New Zealand

Introduction

Hospitals face persistent pressure to improve operational performance while balancing patient flow, workforce availability, and service demand. Objective benchmarking, as suggested by hospital command-center and bed-capacity research [1, 2], can help health systems understand whether bottlenecks reflect local inefficiency, structural constraints, or broader peer-group pressures. Bed occupancy, discharge delays, staffing ratios, and service demand indicators are especially important because they connect capacity, throughput, labor deployment, and access to care. Research linking discharge

delay, emergency department spillover, and nurse staffing to operational strain [3-5] reinforces the need for comparable measures across institutions.

Current benchmarking practices often rely on voluntary surveys, claims-based summaries, centralized registries, or internal data warehouses that require institutions to disclose sensitive performance information. Such approaches can introduce selection bias because, as delayed-discharge reviews indicate [6, 7], hospitals may differ substantially in reporting capacity, data maturity, and willingness to expose operational constraints. Centralized aggregation may also create competitive concerns when shared metrics reveal capacity limits, workforce shortages, or inefficient processes. Staffing-ratio studies [8, 9] further show that comparisons can be misleading when institutional context and measurement definitions are not aligned.

Federated analytics offers a privacy-preserving alternative by enabling hospitals to compute shared summaries without transferring raw operational records to a central repository. Federated learning in digital health [10, 11] and secure medical analytics [12, 13] have shown that institutions can collaborate analytically while retaining local data control. Secure aggregation, differential privacy, and decentralized computation provide building blocks that could be adapted from clinical model development to operational benchmarking. Although these methods have been applied primarily to prediction, imaging, and clinical analytics [14, 15], their logic is well suited to collaborative computation of operational indicators.

This article proposes a federated analytics framework for benchmarking hospital operational performance across health systems using secure aggregation of bed occupancy, discharge delays, staffing ratios, and service demand indicators. The framework follows federated machine learning principles while shifting the target from predictive model training to risk-adjusted operational benchmarking. It is conceptual and does not report experiments, datasets, model performance, or empirical results. By combining secure aggregation, hybrid privacy-preserving computation, and healthcare-specific federated design, the framework could support fair comparison without undermining institutional confidentiality.

Background

Hospital operational performance indicators

Hospital operational performance indicators describe how effectively institutions convert capacity, staffing, and service resources into timely care. Bed occupancy can be interpreted through patient-flow and capacity-management work [1, 2], while discharge delay captures the interval between clinical readiness, disposition decisions, and actual exit from the hospital. Staffing ratios, as examined in nurse staffing and mortality research [5, 8], represent the alignment between workforce inputs and patient care burden. Service demand indicators, including emergency visits, imaging activity, and surgical caseloads, describe operational pressure across access points and resource-intensive services.

Cross-institutional benchmarking and barriers

Cross-institutional benchmarking aims to help hospitals learn from peer performance, but it is constrained by reluctance to share sensitive operational data and by inconsistent metric definitions. Measures such as discharge delay can vary because, as studies of delayed discharge show [3, 6], hospitals may use different timestamps, readiness criteria, or post-acute-care assumptions. Staffing benchmarks are also context-dependent, since nurse-to-patient ratio research [9, 16] indicates that workforce measures must be interpreted alongside acuity, care model, and hospital environment. These barriers make privacy-preserving standardization essential before hospitals can use peer comparisons for improvement.

Federated learning and analytics in healthcare

Federated learning enables multiple institutions to participate in a shared analytic task while keeping data local, and federated analytics extends this logic to collaborative summary computation and benchmarking. In healthcare, federated

approaches have been articulated as a future direction for digital health [10] and demonstrated as a method for multi-institutional collaboration without sharing patient data [11]. Medical imaging and healthcare informatics studies [12, 13] show that distributed computation can reduce the need for central data pooling while preserving analytic collaboration. A hospital operations framework could adapt these principles from clinical model development [17, 18] to the computation of aggregate operational benchmarks.

Secure multi-party computation and aggregation protocols

Secure multi-party computation and secure aggregation protocols allow participants to contribute encrypted or masked values so that only the aggregate result is recoverable. Practical secure aggregation methods [19] provide a foundation for computing operational sums, averages, and distributional summaries without revealing any single hospital's inputs. Hybrid privacy-preserving federated learning [20] and secure end-to-end health analytics [21] further illustrate how privacy mechanisms can be embedded within distributed workflows. In a benchmarking setting, these methods would be expected to reduce the trust burden placed on any central aggregator.

Risk adjustment for fair comparisons

Risk adjustment is essential because hospitals differ in case mix, teaching status, transfer patterns, payer composition, regional demand, and available post-acute resources. Without adjustment, discharge-delay research [6, 7] suggests that hospitals serving more complex patients or constrained discharge environments could appear inefficient even when their internal processes are appropriate. Staffing studies [8, 9] similarly indicate that workforce performance cannot be evaluated without considering patient acuity, care intensity, and organizational setting. A federated benchmarking framework should therefore compute expected-versus-observed indicators or peer-stratified comparisons before presenting results to participants.

Framework Overview

High-level architecture

The proposed architecture consists of local hospital gateways, a secure aggregation layer, and a participatory benchmarking interface. Each hospital would deploy a gateway behind its firewall to standardize operational data, compute local summary representations, and participate in federated computation rounds. A central or decentralized aggregation service, consistent with privacy-preserving healthcare analytics principles [10, 12], would receive only encrypted, masked, or privacy-protected updates rather than raw records. This design adapts the institutional collaboration model demonstrated in federated medicine [11, 14] to operational metrics rather than clinical prediction alone.

Figure 1 presents the proposed federated analytics workflow through which hospitals retain raw operational data locally while generating secure, anonymous, risk-adjusted benchmarks for governed operational improvement.

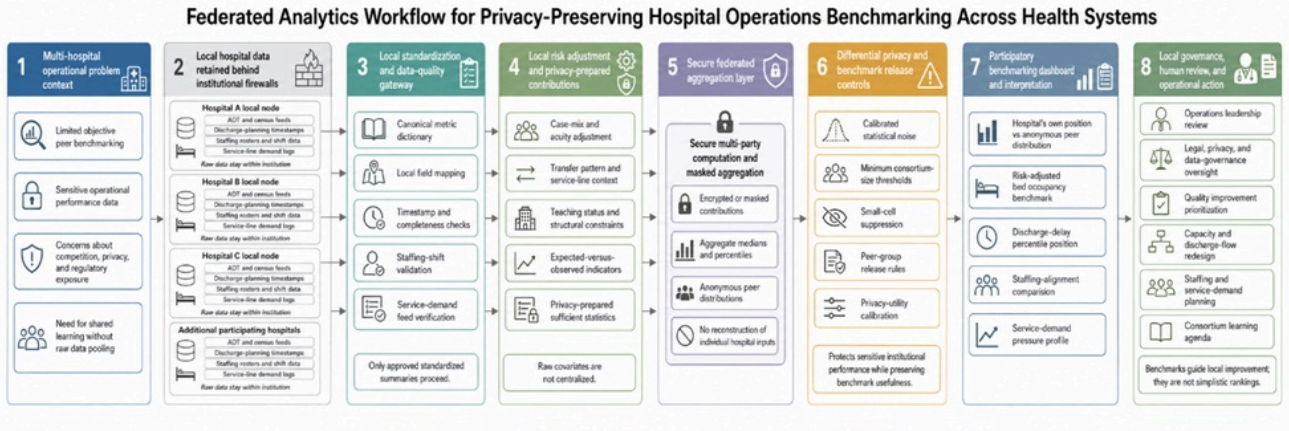


Figure 1. Federated Analytics Workflow for Privacy-Preserving Hospital Operations Benchmarking Across Health Systems

Core metrics and data streams

The framework focuses on four operational domains: bed occupancy, discharge delays, staffing ratios, and service demand. Bed occupancy would be represented conceptually as census relative to staffed capacity, a framing aligned with bed-capacity and patient-flow studies [1, 2]. Discharge delay would be derived from local discharge-planning timestamps, while staffing ratios would reflect nursing labor relative to care burden, as supported by discharge and staffing research [3, 5]. Service demand would be mapped from emergency, imaging, surgical, and related service-line systems to represent the pressure placed on hospital operating units.

Design principles

The framework is guided by privacy-first participation, risk-adjusted fairness, low-effort adoption, incremental deployment, and institutional transparency. Privacy-first participation follows the logic of secure, privacy-preserving, and federated healthcare analytics [12, 13], ensuring that raw operational data remain within hospital control. Risk-adjusted fairness requires that hospitals are compared only after contextual factors are considered, while low-effort adoption would be supported by local mapping libraries and containerized deployment. Transparency to institutional review boards, legal counsel, and executive sponsors would be necessary because decentralized and confidential learning approaches [22, 23] depend on organizational trust.

Operational Metrics Standardization and Risk Adjustment

Standardized metric definitions and local mapping

The framework would provide a canonical operational metric dictionary specifying definitions for bed occupancy, discharge delay, staffing ratio, and service demand indicators. Local mapping libraries would translate hospital-specific fields from admission-discharge-transfer systems, staffing rosters, scheduling tools, and service logs into these canonical measures. This approach would reduce variation caused by local naming conventions, timestamp practices, and departmental reporting structures. Prior studies of patient flow [2], discharge delay [3, 6], and nurse staffing [5] demonstrate why operational indicators require precise definitions before cross-site comparison can be meaningful.

Risk-adjustment models for hospital comparison

Risk-adjustment models would be embedded at the local node or within the secure aggregation workflow to account for institutional context before benchmark construction. Candidate adjustment variables could include case-mix severity, transfer patterns, teaching status, service-line composition, and structural constraints affecting discharge or staffing. The

framework would not require institutions to disclose raw covariates centrally; instead, local nodes could compute adjusted contributions or privacy-protected sufficient statistics. This design responds to evidence from delayed-discharge and staffing research [6-9] that hospital performance is shaped by case complexity and organizational context as well as operational execution.

Data quality and completeness checks at the node

Local gateways would perform data quality and completeness checks before any federated computation round begins. These checks would validate time ranges, identify missing staffing shifts, detect inconsistent timestamp sequences, and flag service-demand feeds that are incomplete or delayed. Data-quality outputs would remain visible to the local institution while only approved privacy-protected contributions would proceed to aggregation. Such safeguards are necessary because studies of command-center design [1], discharge delay [3, 7], and staffing measurement [16] indicate that operational benchmarking can be distorted by incomplete or inconsistently defined source data.

Federated Node Architecture and Secure Aggregation

Local gateway installation and data privacy

Each participating hospital would install a containerized local gateway within its firewall to transform operational data into standardized and privacy-protected analytic contributions. The gateway would connect to local systems through institution-approved interfaces, compute predefined summary statistics or share-like representations, and prevent row-level records from leaving the hospital environment. This architecture preserves institutional data sovereignty while enabling collaboration across hospitals that may otherwise be unable or unwilling to pool sensitive operational data. Federated medicine studies [11, 13] and privacy-preserving healthcare analytics frameworks [14, 23] support the feasibility of local computation as the basis for collaborative analysis.

Table 1 summarizes the federated architecture and data-governance responsibilities required to convert local hospital operational data into privacy-preserving cross-site benchmark outputs.

Table 1. Federated Architecture and Data Governance Matrix for Hospital Operations Benchmarking

Framework layer	Operational role in the manuscript	Local data or analytic object	Privacy-preserving mechanism	Governance requirement	Practical value for participating hospitals
Participating hospital node	Serves as the local institutional environment where operational data remain under hospital control	ADT feeds, census data, discharge timestamps, staffing rosters, service demand logs	Raw data remain behind the hospital firewall	Institutional approval for data connection, local ownership, and permitted metric use	Encourages participation by preserving data sovereignty and reducing fear of exposing sensitive operational performance
Local data standardization engine	Converts heterogeneous hospital fields into shared operational definitions	Bed occupancy, discharge delay, staffing ratio, emergency demand, imaging demand, surgical demand	Local transformation before any external contribution	Shared metric dictionary, version control, local mapping documentation	Makes cross-site comparison more meaningful by reducing definitional inconsistency

Data-quality and completeness checks	Screens local inputs before benchmark computation	Missing timestamps, implausible discharge sequences, incomplete staffing shifts, delayed service feeds	Faulty data are flagged locally before aggregation	Local remediation process and audit trail	Prevents misleading benchmarks caused by incomplete or inconsistent source data
Local risk-adjustment module	Accounts for hospital context before peer comparison	Case mix, acuity, transfer burden, teaching status, service-line composition, structural constraints	Local computation of adjusted values or sufficient statistics	Predefined adjustment variables and transparent model specification	Supports fairer comparison by distinguishing operational performance from contextual burden
Privacy-prepared local contribution	Produces the analytic input sent to the federation	Standardized summaries, sufficient statistics, masked values, adjusted indicators	No row-level operational data leave the hospital	Approval of contribution schema and release boundaries	Allows hospitals to contribute to benchmarking without exposing raw operational records
Secure aggregation layer	Computes consortium-level summaries from protected local contributions	Pooled medians, percentiles, interquartile ranges, expected-versus-observed indicators	Secure multi-party computation, encrypted updates, masking	Protocol audit, server trust assumptions, participant compliance rules	Enables shared benchmark computation while preventing recovery of any single hospital's input
Differential privacy layer	Reduces disclosure risk in released benchmark outputs	Aggregate distributions and peer-comparison summaries	Calibrated noise, suppression thresholds, release controls	Privacy-budget governance and minimum cell-size policies	Balances confidentiality with the need for interpretable operational benchmarks
Participatory dashboard	Returns actionable benchmark insights to each hospital	Hospital-specific position relative to anonymous peer distributions	Anonymous peer display and restricted release of comparative outputs	Access control, interpretation guidance, permissible-use agreements	Converts privacy-preserving analytics into local operational learning
Consortium governance structure	Maintains trust, fairness, and sustained participation	Metric definitions, privacy parameters, audit rules, dashboard use policies	Governance complements technical privacy controls	Participant agreements, legal review, independent oversight, periodic revision	Prevents misuse of benchmarks as simplistic rankings and supports collaborative improvement

Secure aggregation protocol

The secure aggregation protocol would mask each hospital's local contribution before transmission so that the aggregation layer could recover only the approved consortium-level result. Techniques such as secure multi-party sums, encrypted updates, and pairwise masking follow the practical secure aggregation logic described by Bonawitz and colleagues [19]. Even if an aggregation server were curious or compromised, it should not be able to reconstruct any single hospital's raw occupancy, delay, staffing, or demand profile from the masked messages. End-to-end privacy-preserving health analytics [21] and hybrid federated privacy approaches [20] provide conceptual support for this secure computation layer.

Differential privacy injection

Differential privacy would add calibrated statistical noise to local or aggregate outputs so that benchmark reports do not reveal sensitive information about any participating hospital or small subgroup. In an operational benchmarking context, the privacy mechanism should preserve useful peer comparison while reducing the risk that auxiliary information could expose an institution's internal performance. Federated oncology analytics [24] and privacy-preserving health care analytics implementations [23] illustrate how formal privacy concepts can be integrated into multi-institutional health data workflows. The privacy parameters, aggregation thresholds, and release policies would require governance review because excessive protection could reduce interpretability, while insufficient protection could undermine trust.

Privacy-Preserving Benchmark Computation

Aggregate metric computation and percentile estimation

The benchmark computation layer would transform each hospital's local operational summaries into consortium-level distributions without exposing identifiable institutional inputs. Secure aggregation methods [19] could support pooled computation of aggregate occupancy, discharge-delay, staffing, and demand indicators, while federated healthcare analytics principles [10, 13] would guide how local nodes contribute standardized summaries rather than raw records. For distributional benchmarking, the framework could use privacy-preserving approximations of medians, interquartile ranges, and percentile positions so that each hospital can interpret its relative standing. Because differential privacy may affect distributional precision, privacy-preserving health analytics research [24, 25] suggests that benchmark utility should be evaluated conceptually alongside disclosure risk.

Returning comparative reports to each site

After aggregation, each participating hospital would receive only its own risk-adjusted position relative to an anonymous peer distribution. Federated medicine studies [11, 14] support this return-of-insight model because collaborative learning can occur without revealing site-level records or identifiable peer performance. The report would be expected to describe areas of operational opportunity using contextualized language, such as relative discharge-flow burden, staffing alignment, or demand pressure, rather than exposing another hospital's internal metrics. Privacy-preserving healthcare implementation work [23] also indicates that reports should be understandable to operational leaders, legal teams, and data governance committees.

Detecting and preventing strategic manipulation

A federated benchmarking system would need safeguards against strategic manipulation, incomplete reporting, or inconsistent local interpretation of metric definitions. Local validation rules could flag implausible timestamp sequences, abrupt definitional shifts, or inconsistent relationships among bed occupancy, discharge activity, staffing deployment, and service demand before contributions enter the aggregation round. Command-center and patient-flow research [1, 2] supports the need for operational consistency checks, while delayed-discharge studies [3, 7] show that small definitional

differences can materially alter interpretation. The audit layer should therefore protect privacy while still allowing consortium governance to identify suspicious patterns and require local remediation.

Participatory Dashboard and Actionable Insights

Visualizing anonymous peer comparisons

The participatory dashboard would present each hospital's operational profile against anonymous, risk-adjusted peer distributions rather than named competitors. Hospital command-center design principles [1] suggest that operational visualizations should emphasize actionable flow signals, while bed-capacity research [2] indicates that capacity metrics become more useful when displayed in relation to demand and staffing context. The dashboard could show whether a hospital's occupancy pressure, discharge-delay burden, staffing alignment, or service demand pattern differs from the consortium distribution after risk adjustment. To preserve trust, the interface would avoid revealing peer identities, small-cell comparisons, or any visualization that could permit re-identification.

From benchmarks to operational improvement

The framework would connect benchmark interpretation to improvement opportunities by translating anonymous peer patterns into operational hypotheses. Delayed-discharge reviews [6] indicate that interventions such as structured discharge planning, cross-sector coordination, and dedicated transition processes may be relevant when delay patterns persist, while staffing studies [5, 9, 26] suggest that workforce alignment should be interpreted alongside patient acuity and care intensity. Instead of presenting benchmarks as league tables, the dashboard would be expected to support reflective questions about whether capacity, staffing, discharge processes, or service demand management require local redesign. This approach would preserve confidentiality while helping hospitals learn from aggregate patterns of better operational alignment.

Governance, Trust, and Multi-Institutional Collaboration

Governance structure and participant agreements

A federated benchmarking consortium would require a governance structure that defines participant eligibility, metric definitions, privacy parameters, audit rights, and permissible uses of benchmark outputs. Federated healthcare frameworks [10, 13] emphasize that technical privacy mechanisms alone are insufficient without institutional agreements, and secure aggregation research [19] similarly depends on assumptions about participant behavior and protocol compliance. Governance agreements would need to specify how local mappings are validated, how risk-adjustment models are revised, and how benchmark reports may be used internally or externally. A minimal central coordinating body or distributed consensus model could maintain protocol integrity while avoiding unnecessary centralization of sensitive operational data.

Building trust through transparency and pilot studies

Trust would be built through transparent code, third-party review, staged implementation, and clear explanations of how privacy-preserving computation works. Decentralized confidential learning research [22] and privacy-preserving healthcare analytics implementations [23] suggest that participants are more likely to engage when the computational workflow is auditable and locally understandable. Early deployment could use synthetic or historical operational data to test data mappings, governance procedures, and dashboard interpretation before live benchmarking is introduced. Such staged participation would allow hospitals to evaluate privacy claims, operational relevance, and institutional acceptability before committing to sustained consortium activity.

Evaluation Strategy

Privacy and security validation

The privacy and security evaluation should assess whether the aggregation protocol prevents reconstruction of individual hospital inputs under realistic adversarial assumptions. Practical secure aggregation methods [19], secure medical analytics [12], and end-to-end privacy-preserving health computation [21] provide conceptual foundations for formal protocol review, penetration testing, and independent audit. The evaluation would examine whether masked contributions, differential privacy injection, release thresholds, and access controls jointly protect participating institutions. Because federated analytics may involve both technical and organizational risk, validation should include legal, operational, and information-security review rather than relying solely on mathematical guarantees.

Benchmark accuracy and risk-adjustment validity

Benchmark validity should be evaluated by comparing whether federated computation preserves the conceptual meaning of operational indicators after standardization, risk adjustment, and privacy protection. Federated learning surveys [27, 28] indicate that distributed analytics can introduce heterogeneity-related challenges, while hospital operations studies [3, 4, 7] show that local context strongly shapes interpretation of delay, flow, and demand metrics. The framework should therefore be assessed for whether it produces fair expected-versus-observed comparisons without requiring hospitals to disclose raw covariates centrally. Privacy noise and secure aggregation would be expected to introduce some utility trade-off, so the evaluation should focus on whether benchmark interpretation remains operationally useful rather than on artificial performance claims.

Participatory experience and impact on practice

A pilot evaluation should examine whether participating hospitals understand, trust, and act on the dashboard outputs. Federated clinical collaboration studies [11, 18, 29] show that multi-institutional analytics can support shared learning when participants perceive the workflow as credible and privacy-preserving. The evaluation could use interviews, governance feedback, and workflow observations to assess whether operational leaders find risk-adjusted peer comparisons useful for local improvement planning. Studies of delayed discharge, staffing ratios, and alternative staffing models [6, 9, 26] suggest that impact should be judged by whether the framework helps institutions identify plausible improvement pathways rather than by reporting numerical performance gains.

Table 2 translates the proposed benchmarking framework into operational indicators, validation requirements, failure modes, safeguards, and practical implementation actions for participating hospitals.

Table 2. Cross-Site Benchmarking, Implementation Use, Failure Modes, and Readiness Framework

Benchmarking domain	Core operational indicator	Practical decision relevance	Required validation before deployment	Key failure mode	Governance or mitigation strategy	Implementation action supported
Bed occupancy	Census relative to staffed capacity	Identifies capacity strain and whether occupancy pressure exceeds anonymous peer patterns	Confirm consistent staffed-bed definitions, census time windows, and service-line inclusion rules	Hospitals count staffed capacity differently, producing misleading comparisons	Shared metric dictionary, local mapping review, and periodic definition audits	Capacity planning, command-center escalation thresholds, bed-management redesign

Discharge delay	Time between clinical readiness, discharge order, disposition readiness, and actual hospital exit	Reveals discharge-flow bottlenecks and post-acute transition pressure	Validate timestamp sequence, readiness criteria, excluded patient groups, and discharge-status definitions	Local timestamp practices differ across hospitals or units	Standard timestamp hierarchy, local data-quality flags, and peer-group stratification	Discharge planning redesign, transition-of-care coordination, escalation of delayed placements
Staffing ratio	Workforce availability relative to patient volume, acuity, or care burden	Supports staffing alignment and interpretation of workforce strain	Validate shift-level staffing feeds, nurse role categories, agency staff inclusion, and acuity adjustment	Simple nurse-to-patient ratios ignore skill mix, care intensity, and unit context	Risk adjustment for acuity and service mix; transparent reporting of denominator definitions	Staffing model review, shift planning, workforce allocation, float-pool strategy
Service demand	Emergency visits, imaging activity, surgical caseloads, consult volume, or other demand indicators	Shows external and internal demand pressure affecting capacity and throughput	Validate scheduled vs urgent activity, cancellations, overflow activity, and service-line boundaries	Demand indicators are not comparable across hospitals with different service portfolios	Peer grouping by hospital type and service-line composition	Demand forecasting, service-line capacity planning, resource prioritization
Risk-adjusted peer percentile	Hospital position relative to anonymous peer distribution after contextual adjustment	Helps leaders distinguish local process gaps from structural or case-mix burden	Validate adjustment variables, expected-versus-observed logic, and fairness across hospital types	Inadequate adjustment unfairly labels complex hospitals as poor performers	Governance review of adjustment model and periodic recalibration	Fairer benchmarking, executive reporting, targeted improvement selection
Privacy-preserving benchmark release	Medians, percentiles, interquartile ranges, and anonymous comparative indicators	Enables shared learning without exposing identifiable peer performance	Test privacy leakage risk, re-identification risk, and utility loss from noise	Small peer groups or rare service patterns allow inference about a specific hospital	Minimum consortium size, small-cell suppression, differential privacy calibration	Safe release of operational benchmarks to participants

Participatory dashboard interpretation	Local benchmark report showing anonymous peer context	Converts analytic output into operational insight	Test usability with operations leaders, quality teams, and data-governance stakeholders	Dashboard outputs are interpreted as league tables rather than improvement signals	Interpretive guidance, governance-approved language, and human review before action	Quality improvement planning, operational review meetings, capacity and staffing discussions
Strategic manipulation detection	Consistency checks across occupancy, discharge, staffing, and demand measures	Protects benchmark integrity and discourages selective reporting	Validate anomaly-detection rules, abrupt-definition-change flags, and data-submission completeness	Hospitals underreport unfavorable indicators or alter local definitions	Audit rights, local remediation procedures, consortium compliance policies	Trustworthy participation and durable multi-site collaboration
Pilot implementation readiness	Feasibility, trust, usability, privacy validation, and operational relevance	Determines whether the framework can move from concept to real-world consortium testing	Conduct synthetic-data testing, historical-data pilots, security review, and workflow observation	Technical privacy works but operational users do not trust or use the output	Staged rollout, transparent code review, legal/privacy review, and stakeholder training	Multi-health-system pilot, governance refinement, dashboard iteration, future empirical evaluation

Limitations

Metric heterogeneity and local interpretation

Even with a canonical metric dictionary, hospitals may interpret operational concepts differently because local systems, workflows, and documentation practices vary. Discharge-delay studies [3, 6, 7] show that readiness for discharge, actual exit, and post-acute constraints can be measured in different ways, while staffing research [5, 8, 16] indicates that nurse-to-patient ratios may not fully capture care intensity or skill mix. Service demand indicators may also differ depending on whether hospitals count scheduled, urgent, cancelled, or overflow activity in the same way. The framework would therefore reduce but not eliminate heterogeneity, and every benchmark should be interpreted as a structured comparison rather than a definitive ranking.

Privacy-utility trade-off and minimum consortium size

The privacy-utility trade-off is a central limitation because stronger privacy protections can make benchmarks less precise or less interpretable. Differentially private federated analytics [24] and privacy-preserving healthcare implementations [23, 25] suggest that release policies must balance confidentiality with operational usefulness. Small or highly specialized consortia may face greater re-identification risk, especially when rare service-line patterns or distinctive staffing models are included. The framework should therefore include governance rules for aggregation thresholds, peer-group construction, and suppression of outputs that cannot be released safely.

Conclusion

A federated analytics framework for hospital operational benchmarking would allow health systems to compare performance without centralizing sensitive institutional data. By using local gateways, standardized operational definitions, secure aggregation, and privacy-preserving reporting, the framework would create a structure for collaborative learning while respecting institutional data sovereignty.

The main strength of this approach is that it aligns privacy protection with operational usefulness. Hospitals could receive risk-adjusted, actionable comparisons on bed occupancy, discharge delays, staffing ratios, and service demand without exposing raw records or identifiable peer performance.

Important challenges remain, including metric standardization, governance design, privacy-utility calibration, and sustained participation across diverse institutions. The framework would also require careful communication so that benchmarks are interpreted as decision-support tools rather than simplistic rankings.

A multi-health-system pilot would be a practical next step to assess feasibility, trust, usability, and governance readiness. Such a pilot could help determine whether federated analytics can become a durable infrastructure for collaborative hospital operations improvement.

Acknowledgements

None

Conflict of interest

None

Financial support

None

Ethics statement

None

Received: 10 Apr 2025

Revised: 11 May 2025

Accepted: 30 Jun 2025

Published online: 20 July 2025

Rights and permissions

Open Access The author(s) retain copyright. This article is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. It may be shared and adapted for non-commercial purposes with appropriate attribution, an indication of changes, and distribution of adaptations under the same license. Third-party material may be subject to separate terms identified in its credit line. View the license at <https://creativecommons.org/licenses/by-nc-sa/4.0/>.

References

Kane EM, Scheulen JJ, Püttgen A, Martinez D, Levin S, Bush BA, et al. Use of systems engineering to design a hospital command center. *Jt Comm J Qual Patient Saf.* 2019;45(5):370-9.

Hu Y, Dong J, Perry O, Cyrus RM, Gravenor S, Schmidt MJ. Use of a novel patient-flow model to optimize hospital bed capacity for medical patients. *Jt Comm J Qual Patient Saf.* 2021;47(6):354-63.

van den Ende E, Schouten B, Pladet L, Merten H, van Galen L, Marinova M, et al. Leaving the hospital on time: hospital bed utilization and reasons for discharge delay in the Netherlands. *Int J Qual Health Care.* 2023;35(2):mzad022.

Friebel R, Juarez RM. Spill over effects of inpatient bed capacity on accident and emergency performance in England. *Health Policy.* 2020;124(11):1182-91.

Griffiths P, Maruotti A, Recio Saucedo A, Redfern OC, Ball JE, Briggs J, et al. Nurse staffing, nursing assistants and hospital mortality: retrospective longitudinal cohort study. *BMJ Qual Saf.* 2019;28(8):609-17.

Cadel L, Guilcher SJT, Kokorelias KM, Sutherland J, Glasby J, Kiran T, et al. Initiatives for improving delayed discharge from a hospital setting: a scoping review. *BMJ Open.* 2021;11(2):e044291.

Vinci A, Furia G, Cammalleri V, Colamesta V, Chierchini P, Corrado O, et al. Burden of delayed discharge on acute hospital medical wards: a retrospective ecological study in Rome, Italy. *PLoS One.* 2024;19(1):e0294785.

Zaranko B, Sanford NJ, Kelly E, Rafferty AM, Bird J, Mercuri L, et al. Nurse staffing and inpatient mortality in the English National Health Service: a retrospective longitudinal study. *BMJ Qual Saf.* 2023;32(5):254-63.

McHugh MD, Aiken LH, Sloane DM, Windsor C, Douglas C, Yates P. Effects of nurse-to-patient ratio legislation on nurse staffing and patient mortality, readmissions, and length of stay: a prospective study in a panel of hospitals. *Lancet.* 2021;397(10288):1905-13.

Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, et al. The future of digital health with federated learning. *NPJ Digit Med.* 2020;3(1):119.

Guo P, Wang P, Zhou J, Jiang S, Patel VM. Multi-institutional collaborations for improving deep learning-based magnetic resonance image reconstruction using federated learning. In: *Proc IEEE/CVF Conf Comput Vis Pattern Recognit*; 2021. p. 2423-32.

Kaissis GA, Makowski MR, Rückert D, Braren RF. Secure, privacy-preserving and federated machine learning in medical imaging. *Nat Mach Intell.* 2020;2(6):305-11.

Xu J, Glicksberg BS, Su C, Walker P, Bian J, Wang F. Federated learning for healthcare informatics. *J Healthc Inform Res.* 2021;5(1):1-19.

Pati S, Kumar S, Varma A, Edwards B, Lu C, Qu L, et al. Privacy preservation for federated learning in health care. *Patterns.* 2024;5(7):100987.

Yang Q, Liu Y, Chen T, Tong Y. Federated machine learning: concept and applications. *ACM Trans Intell Syst Technol.* 2019;10(2):1-19.

Lasater KB, Sloane DM, McHugh MD, Cimiotti JP, Riman KA, Martin B, et al. Evaluation of hospital nurse-to-patient staffing ratios and sepsis bundles on patient outcomes. *Am J Infect Control.* 2021;49(7):868-73.

Chen Y, Qin X, Wang J, Yu C, Gao W. FedHealth: a federated transfer learning framework for wearable healthcare. *IEEE Intell Syst.* 2020;35(4):83-93.

Dou Q, So TY, Jiang M, Liu Q, Vardhanabhuti V, Kaissis G, et al. Federated deep learning for detecting COVID-19 lung abnormalities in CT: a privacy-preserving multinational validation study. *NPJ Digit Med.* 2021;4(1):60.

Bonawitz K, Ivanov V, Kreuter B, Marcedone A, McMahan HB, Patel S, et al. Practical secure aggregation for privacy-preserving machine learning. In: *Proc ACM SIGSAC Conf Comput Commun Secur*; 2017. p. 1175-91.

Truex S, Baracaldo N, Anwar A, Steinke T, Ludwig H, Zhang R, et al. A hybrid approach to privacy-preserving federated learning. In: *Proc ACM Workshop Artif Intell Secur*; 2019. p. 1-11.

Kaissis G, Ziller A, Passerat-Palmbach J, Ryffel T, Usynin D, Trask A, et al. End-to-end privacy preserving deep learning on multi-institutional medical imaging. *Nat Mach Intell*. 2021;3(6):473-84.

Warnat-Herresthal S, Schultze H, Shastri KL, Manamohan S, Mukherjee S, Garg V, et al. Swarm learning for decentralized and confidential clinical machine learning. *Nature*. 2021;594(7862):265-70.

Choudhury A, Volmer L, Martin F, Fijten R, Wee L, Dekker A, et al. Advancing privacy-preserving health care analytics and implementation of the personal health train: federated deep learning study. *JMIR AI*. 2025;4(1):e60847.

Ryffel T, Créquit P, Baillet M, Paumier J, Marfoq Y, Girardot O, et al. Federated analysis with differential privacy in oncology research: longitudinal observational study across hospital data warehouses. *JMIR Med Inform*. 2025;13(1):e59685.

Fang C, Dziedzic A, Zhang L, Oliva L, Verma A, Razak F, et al. Decentralised, collaborative, and privacy-preserving machine learning for multi-hospital data. *EBioMedicine*. 2024;101:104995.

Lasater KB, Muir KJ, Sloane DM, McHugh MD, Aiken LH. Alternative models of nurse staffing may be dangerous in high-stakes hospital care. *Med Care*. 2024;62(7):434-40.

Li T, Sahu AK, Talwalkar A, Smith V. Federated learning: challenges, methods, and future directions. *IEEE Signal Process Mag*. 2020;37(3):50-60.

Guan H, Yap PT, Bozoki A, Liu M. Federated learning for medical image analysis: a survey. *Pattern Recognit*. 2024;151:110424.

Dayan I, Roth HR, Zhong A, Harouni A, Gentili A, Abidin AZ, et al. Federated learning for predicting clinical outcomes in patients with COVID-19. *Nat Med*. 2021;27(10):1735-43.