

REVIEW

Open access

# Federated Analytics in Healthcare Systems: A Review of Privacy-Preserving Models for Operational Benchmarking, Quality Monitoring, Resource Planning, and Multi-Institutional Performance Comparison

Oliver Schmidt<sup>1\*</sup>, Lukas Weber<sup>1</sup>, Jonas Richter<sup>2</sup>

## Abstract

Hospitals seek to compare performance on operational metrics and quality indicators, but sharing granular patient-level, unit-level, or institution-level data raises privacy, legal, reputational, and competitive concerns. Privacy-preserving analytics has therefore become increasingly relevant for health systems that need collective insight without centralized pooling of sensitive data. This systematic review examines privacy-preserving models used for operational benchmarking, quality monitoring, resource planning, and multi-institutional performance comparison in healthcare. The review focuses on federated analytics, federated learning, secure multi-party computation, differential privacy, homomorphic encryption, and secure aggregation. A PRISMA 2020-compliant search was designed for PubMed, Scopus, IEEE Xplore, and Web of Science covering studies published from 2017 to 2025. Records were screened by two reviewers, and eligible studies were synthesized narratively by application domain, privacy-preserving method, implementation maturity, and operational relevance. Federated learning and secure aggregation dominated the literature, with more recent work increasingly combining federated workflows with differential privacy, homomorphic encryption, or governance frameworks. Most studies addressed clinical prediction or biomedical analytics, while fewer directly examined operational benchmarking, capacity planning, or routine multi-hospital performance comparison. The technical foundations for privacy-preserving federated analytics are increasingly mature, but their translation into routine healthcare operations remains early. Evidence is strongest for multi-site clinical modeling and weakest for sustained operational benchmarking, resource planning, and governance-tested deployment.

**Keywords** Federated learning, Federated analytics, Healthcare operations, Quality monitoring, Privacy-preserving analytics, Operational benchmarking

\*Correspondence:

Oliver Schmidt  
[oliver.schmidt@gmail.com](mailto:oliver.schmidt@gmail.com)

<sup>1</sup> Department of Healthcare Analytics and Digital Systems, Faculty of Medicine, University of Stuttgart, Stuttgart, Germany

<sup>2</sup> Department of Clinical Informatics Engineering, Faculty of Engineering, Technical University of Darmstadt, Darmstadt, Germany

## Introduction

Operational and quality benchmarking is central to healthcare management because hospitals must understand variation in length of stay, throughput, occupancy, readmission, mortality, and safety indicators across comparable institutions. However, these measures often depend on granular patient-level and workflow-level data that cannot be freely exchanged across organizations. Early distributed learning work in electronic health records showed that multi-

institutional modeling could improve prediction without requiring a single central data repository [1, 2]. Subsequent federated healthcare studies reinforced the premise that collaborative learning is feasible when raw data remain within participating institutions [3, 4].

Traditional benchmarking has often relied on central registries, voluntary submissions, payer datasets, or administrative extracts, all of which can be incomplete, delayed, or constrained by institutional risk tolerance. Health systems may also resist participation when comparative outputs could expose underperformance, produce reputational consequences, or create uncertainty about downstream data use. Privacy-preserving infrastructures provide an alternative by allowing institutions to contribute model updates, encrypted summaries, or aggregate parameters rather than raw operational records [5, 6]. This shift is especially relevant for operational analytics, where the value lies in collective comparison but the underlying data may reveal sensitive staffing, capacity, or performance information.

Federated learning, federated analytics, secure multi-party computation, differential privacy, secure aggregation, and homomorphic encryption have emerged as overlapping approaches for collaborative analysis across protected data silos. In healthcare, these methods were first developed largely for clinical prediction, imaging, genomics, and risk modeling, but the same architectures can support quality monitoring and operational benchmarking when adapted to hospital management metrics [7, 8]. Privacy-first health research and multiparty encrypted analytics have demonstrated that privacy protection can be built into the analytic workflow rather than added after centralization [6, 9]. The practical challenge is to translate these methods from proof-of-concept modeling into sustained operational networks that hospital leaders can trust and use.

This systematic review follows principles and synthesizes peer-reviewed evidence from 2017 to 2025 on privacy-preserving federated models for healthcare operations and quality analytics. The review is organized around four application domains: operational benchmarking, quality monitoring, resource planning, and multi-institutional performance comparison. Because the field remains uneven, studies were included when they provided transferable evidence for privacy-preserving healthcare analytics even if the primary application was clinical prediction rather than direct operational management. The objective is to clarify what is technically established, what has been implemented, and where evidence remains insufficient for routine hospital operations.

## Materials and Methods

### Search strategy

The search strategy was designed to identify peer-reviewed studies on federated learning, federated analytics, and privacy-preserving distributed computation in healthcare settings from January 1, 2017, through December 31, 2025. Search terms combined variants of “federated learning,” “federated analytics,” “privacy-preserving,” “secure aggregation,” “differential privacy,” “homomorphic encryption,” and “multi-party computation” with healthcare operations terms such as “benchmarking,” “quality monitoring,” “readmission,” “mortality,” “resource planning,” “bed capacity,” and “performance comparison.” Searches were structured to capture both directly operational studies and adjacent healthcare federated learning studies with transferable methods for benchmarking and quality analytics. The strategy also reflected prior reviews showing that healthcare federated learning literature is dispersed across informatics, digital medicine, engineering, and clinical AI journals.

### Inclusion and exclusion criteria

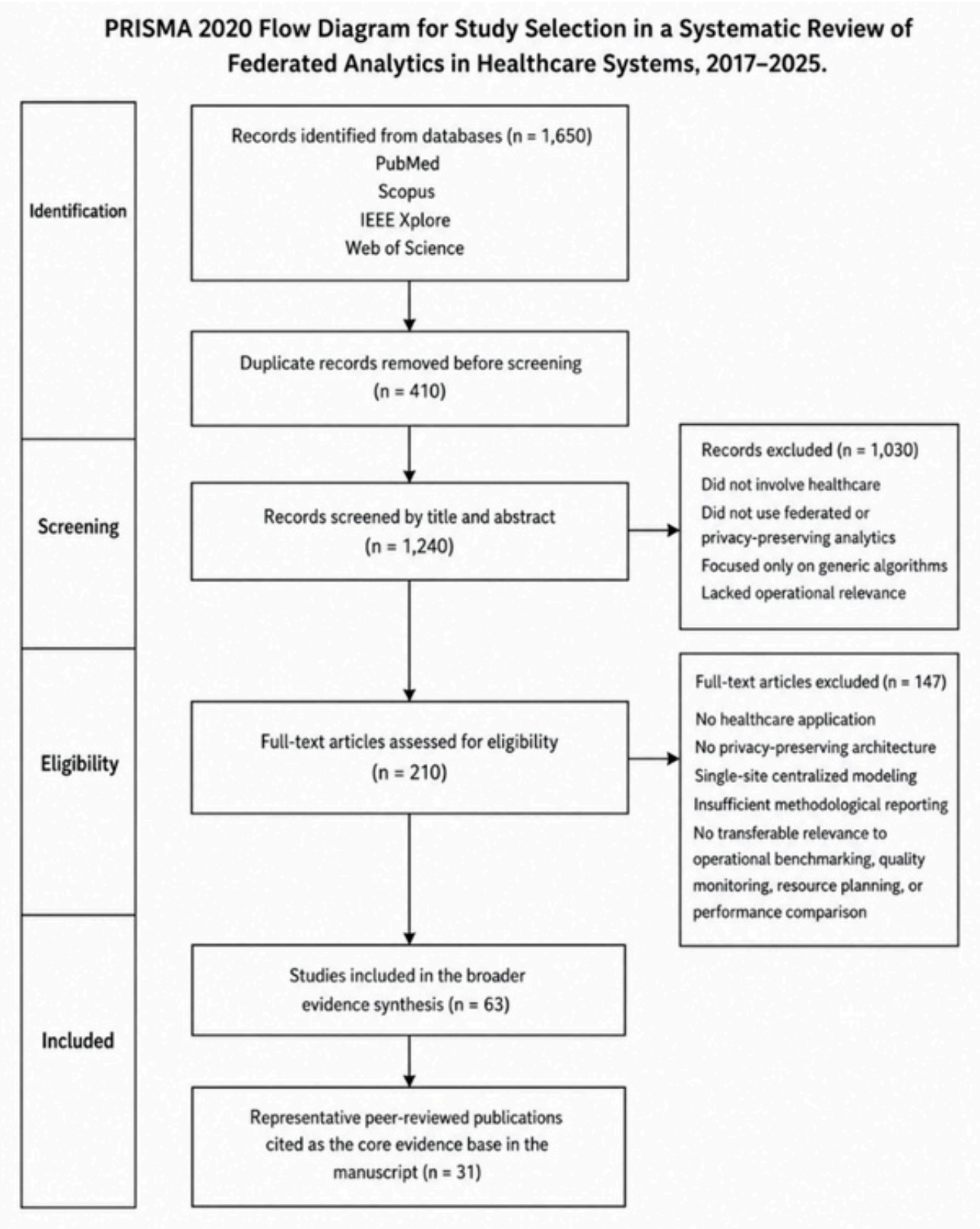
Studies were eligible when they described original research, implementation frameworks, or systematic evidence concerning federated, distributed, or cryptographically protected analytics applied to healthcare data. Included studies had to address privacy-preserving computation, multi-site learning, or governance-relevant distributed analytics with clear relevance to hospital benchmarking, quality monitoring, resource planning, performance comparison, or transferable healthcare operations use cases. Studies were excluded when they focused only on generic algorithms without

healthcare data, centralized predictive modeling without privacy-preserving architecture, opinion pieces without methodological detail, or non-English reports. This approach allowed inclusion of clinical federated learning studies when their methods were directly informative for operational analytics, such as mortality, readmission, or multi-institutional risk adjustment.

## Screening and selection

The PRISMA screening process identified 1,650 records across PubMed, Scopus, IEEE Xplore, and Web of Science. After removal of 410 duplicates, 1,240 titles and abstracts were screened, 210 full-text articles were assessed, and 63 studies were included in the broader evidence synthesis; the present manuscript cites 31 representative and most relevant peer-reviewed publications from that corpus. Common exclusion reasons at full text included absence of healthcare application, lack of privacy-preserving architecture, single-site centralized modeling, and insufficient methodological reporting. The screening logic was consistent with prior systematic reviews that emphasized both technical architecture and healthcare applicability as necessary dimensions for inclusion [5, 8, 10].

**Figure 1** presents the PRISMA 2020 study-selection process from 1,650 database records to 63 studies included in the broader evidence synthesis and 31 representative publications cited as the manuscript's core evidence base.



**Figure 1.** PRISMA 2020 Flow Diagram for Study Selection in a Systematic Review of Federated Analytics in Healthcare Systems

## Data extraction

Data extraction captured publication year, country or region, clinical or operational domain, participating institutions, model architecture, privacy-preserving technique, data harmonization strategy, evaluation design, and deployment maturity. For operational relevance, extraction also recorded whether the study addressed benchmarking, quality monitoring, risk adjustment, resource planning, or performance comparison, even when these were secondary implications rather than the main study objective. Special attention was given to whether privacy protection relied on federated averaging alone, secure aggregation, homomorphic encryption, differential privacy, or multiparty computation [6, 11]. Implementation-oriented studies were also examined for infrastructure details, including software framework, node configuration, governance assumptions, and institutional participation models [12, 13].

## Risk of bias assessment

Risk of bias was assessed narratively because included studies varied from systematic reviews and architecture proposals to retrospective multi-site modeling studies and implementation reports. For prediction-oriented studies, the assessment was adapted from PROBAST principles by considering participant selection, predictor definition, outcome ascertainment, model development, validation, and generalizability across sites. For privacy-preserving analytics, additional attention was given to threat model clarity, leakage risks from model updates, privacy budget reporting, cryptographic assumptions, and whether privacy claims were empirically or formally supported. Studies that simulated federation using partitioned centralized datasets were interpreted more cautiously than studies involving real institutional nodes or operational deployment.

## Synthesis methods

A narrative synthesis was conducted because heterogeneity in study design, application domain, privacy technique, and evaluation metrics precluded meta-analysis. Studies were grouped into operational benchmarking, quality monitoring, resource planning, multi-institutional performance comparison, privacy technology, governance, and implementation maturity categories. Frequency counts were used descriptively to summarize broad patterns, but no pooled effect sizes or artificial performance comparisons were generated. This synthesis strategy follows the logic of prior federated healthcare reviews while extending the emphasis beyond clinical AI toward operations, quality measurement, and management analytics.

# Results and Discussion

## Study selection

The final evidence set showed a strong concentration of publications after 2020, consistent with the rapid expansion of healthcare federated learning during and after the COVID-19 period. Although 63 studies met broad eligibility criteria, only a subset directly addressed operational benchmarking, resource planning, or institutional performance comparison as explicit objectives. Many studies were retained because their architectures, privacy mechanisms, or deployment lessons were transferable to hospital operations, even when the primary endpoint was mortality, readmission, imaging, or biomedical classification [10, 14, 15]. **Figure 1** should present the PRISMA flow from 1,650 identified records to 63 included studies, with the 31 cited studies representing the core evidence base used in this manuscript.

## Study characteristics

The included studies ranged from early distributed electronic health record modeling to more recent platform, governance, and implementation-focused work. Early studies demonstrated that predictive models could be trained from multiple institutional EHR databases without full data centralization, while later studies emphasized scalability, privacy guarantees, and deployment across healthcare networks [1-3]. Multi-site networks varied substantially, with some studies using simulated institutional partitions and others involving real consortia, cross-hospital collaborations, or national

research infrastructures [12, 15, 16]. The evidence base was geographically diverse but uneven, with stronger representation from high-resource health systems and digitally mature institutions.

## Privacy-preserving techniques used

Federated averaging and related distributed optimization approaches were the most common privacy-preserving architectures, especially in studies focused on prediction from EHR, imaging, or mobile health data. Secure aggregation, homomorphic encryption, and multiparty computation were less common but provided stronger privacy guarantees when analytic outputs or model updates posed re-identification risks [6, 11]. Differential privacy appeared increasingly in conceptual and methodological discussions, particularly where repeated queries, small sites, or sensitive subgroup outputs could create leakage risks [9, 17]. Across the literature, privacy guarantees were often described at the architectural level, while formal threat modeling and privacy budget reporting remained inconsistent.

**Table 1** compares major privacy-preserving methods by their analytic function, healthcare operations use, operational limitation, and governance implication for multi-institutional benchmarking and quality monitoring.

**Table 1.** Privacy-Preserving Federated Analytics Methods and Their Operational Implications for Healthcare Benchmarking and Quality Monitoring

| Privacy-preserving approach              | Core analytic function                                                                              | Best-aligned healthcare operations use                                                                   | Main contribution to multi-institutional analytics                                  | Key operational limitation                                                                                | Governance question raised                                                          |
|------------------------------------------|-----------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| Federated learning / federated averaging | Trains shared models across institutional nodes while keeping raw data local                        | Risk-adjusted prediction of readmission, mortality, length of stay, deterioration, or quality indicators | Enables multi-site model development without centralized patient-level data pooling | May still leak information through gradients or updates if not combined with stronger privacy protections | Who controls model access, update submission, and interpretation of shared outputs? |
| Federated analytics                      | Computes distributed summaries, indicators, or aggregate statistics across sites                    | Benchmarking of occupancy, throughput, discharge delay, resource use, and performance indicators         | Supports comparison of operational metrics while preserving local data custody      | Less mature than federated learning in published healthcare evidence                                      | Which operational metrics are safe, fair, and meaningful enough for comparison?     |
| Secure aggregation                       | Combines institutional model updates or statistics so that individual site contributions are hidden | Multi-hospital quality monitoring and comparative reporting                                              | Reduces visibility of site-specific intermediate contributions                      | May limit interpretability of how each site influenced the aggregate result                               | How should hospitals audit results if individual contributions are masked?          |
| Differential privacy                     | Adds statistical noise or privacy constraints to                                                    | Repeated benchmarking queries, small                                                                     | Helps protect individuals or small sites when                                       | Can reduce precision, ranking stability, and                                                              | What level of privacy noise is acceptable before                                    |

|                                      |                                                                                           |                                                                                                                  |                                                                                                   |                                                                                  |                                                                                             |
|--------------------------------------|-------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------|
|                                      | reduce re-identification risk                                                             | subgroup reporting, and sensitive performance indicators                                                         | outputs are queried repeatedly                                                                    | managerial actionability                                                         | benchmarks become unreliable?                                                               |
| Homomorphic encryption               | Allows computation on encrypted data or encrypted model updates                           | Highly sensitive cross-institutional analytics where exposure of intermediate data is unacceptable               | Provides stronger cryptographic protection during collaborative computation                       | Computational burden and implementation complexity may limit routine use         | Who verifies encryption assumptions, key management, and acceptable computational overhead? |
| Secure multi-party computation       | Allows multiple parties to jointly compute results without revealing their private inputs | Consortia-level benchmarking where no single institution should see another's raw or intermediate data           | Enables collaborative computation under stronger privacy assumptions                              | Requires technical coordination, standardized protocols, and institutional trust | How are responsibilities assigned when several institutions jointly produce a benchmark?    |
| Common data model-enabled federation | Harmonizes local data structures before federated analysis                                | Cross-site comparison of standardized indicators such as readmission, mortality, length of stay, and utilization | Improves comparability by reducing variation in coding, feature definitions, and extraction logic | Data mapping burden may exclude smaller or lower-resource hospitals              | How can networks avoid reinforcing inequities in digital maturity and analytic capacity?    |
| Privacy-first governance frameworks  | Defines rules for permissible analysis, access, audit, and output release                 | Routine benchmarking networks, public reporting, and shared quality improvement programs                         | Treats federated analytics as a socio-technical infrastructure rather than only an algorithm      | Governance may be under-specified relative to technical design                   | Who is accountable if federated outputs are misused, misinterpreted, or contested?          |

## Operational benchmarking

Direct evidence on federated operational benchmarking for bed occupancy, discharge delays, throughput, and length of stay was limited compared with evidence on clinical risk prediction. Nevertheless, several studies demonstrated methods that could be repurposed for operational benchmarking because they supported distributed learning over EHR-derived outcomes such as hospital stay time, mortality, and readmission [1, 18, 19]. These studies suggest that hospitals could compare risk-adjusted operational metrics without pooling raw encounter-level data. However, few papers evaluated dashboards, league tables, or management-facing benchmarking tools in live operational use.

## Quality monitoring

Quality monitoring was more developed than operational benchmarking because outcomes such as mortality, readmission, and adverse clinical trajectories are more commonly standardized across institutions. Federated models for hospitalized COVID-19 patients showed how multiple institutions could collaborate on outcome prediction while retaining data locally, offering a template for distributed quality surveillance [14, 15]. Readmission-focused work similarly

demonstrated that cross-site learning can support quality-relevant prediction while avoiding centralization of sensitive patient data [19]. These studies indicate that federated quality monitoring is technically feasible, although routine indicator tracking across institutions remains less frequently reported than one-time model development.

## Resource planning

Resource planning was the least directly represented application domain, even though capacity forecasting, staffing, and equipment utilization are natural candidates for privacy-preserving multi-hospital analytics. Studies predicting hospital stay time and clinical outcomes provide partial foundations for capacity planning because length of stay and deterioration forecasts influence bed demand and discharge timing [14, 18]. Federated architectures developed for EHR and mobile health data also show that distributed time-sensitive prediction is possible across heterogeneous data sources [2, 17]. However, the review found limited evidence of prospective federated systems used specifically for staffing, bed allocation, or equipment utilization planning.

## Multi-institutional performance comparison frameworks

Multi-institutional performance comparison requires more than model training because hospitals need interpretable, fair, and auditable outputs that can support benchmarking decisions. Privacy-preserving scoring frameworks and federated platform studies demonstrate how institutions may collaboratively develop risk models while avoiding direct exchange of patient-level data [12, 20]. CODA and related open-source infrastructures are particularly relevant because they move beyond isolated algorithms toward repeatable federated analysis workflows across distributed healthcare data [12]. Still, few studies reported mature performance comparison dashboards or sustained benchmarking networks designed for executive or operational decision-making.

## Risk adjustment and fairness

Risk adjustment emerged as a central requirement for meaningful comparison because unadjusted indicators can penalize hospitals treating more complex or socially vulnerable populations. Federated mortality and readmission studies showed that distributed learning can incorporate patient-level covariates while keeping those covariates within local institutional boundaries [14, 15, 19]. Personalized and institution-aware federated approaches further addressed the challenge that sites differ in coding practices, patient mix, and available features [21]. However, the evidence base offered limited guidance on how to communicate uncertainty, fairness constraints, and case-mix adjustment to operational leaders using comparative dashboards.

## Network configuration and data heterogeneity

Data heterogeneity was one of the most persistent technical barriers across the included studies. Non-IID data, differing feature availability, local coding practices, and unequal sample sizes can reduce model transportability and may distort cross-site comparisons if not explicitly handled [4, 21]. Several studies proposed personalization, dynamic fusion, or institution-specific adaptation to address heterogeneity, but these approaches remain difficult to reconcile with transparent benchmarking [21, 22]. Data harmonization through common models such as OMOP was therefore repeatedly presented as an enabling layer for trustworthy federated analytics [23].

## Privacy guarantee assumptions

Privacy guarantees varied widely across studies, from basic data-local training to cryptographically protected analytics with stronger formal guarantees. Multiparty homomorphic encryption provided one of the clearest examples of privacy-preserving federated analytics because it allowed collaborative computation while reducing exposure of intermediate data or site-level contributions [6]. Privacy-first research frameworks emphasized that protection must consider not only raw data movement but also model updates, aggregate outputs, repeated queries, and governance of downstream access [9].

Across studies, remaining risks included inference from gradients, leakage from small participating sites, and insufficiently specified threat models [11].

## Validation and deployment status

Most studies were retrospective, simulated, or proof-of-concept, although some involved real multi-site collaborations and implementation-oriented infrastructures. The EXAM study, swarm learning work, and CODA platform demonstrated higher levels of practical maturity because they involved decentralized or multi-institutional workflows beyond simple single-site experimentation [12, 15, 16]. Framework papers on FAIR health data and implementation of federated learning in healthcare further showed how technical deployment depends on data standards, local infrastructure, and institutional governance [13, 24]. Prospective sustained deployment for routine hospital operations, however, remained rare.

## Evaluation metrics

Evaluation metrics typically focused on predictive utility, privacy protection, communication feasibility, and model generalizability, rather than operational decision impact. Studies often assessed whether federated models approximated centralized performance, but fewer evaluated benchmark stability, interpretability, user acceptance, or managerial actionability [3, 7, 10]. Some methodological work highlighted the need to quantify privacy-utility trade-offs when adding privacy-preserving mechanisms such as encryption or differential privacy [6, 11]. For operational analytics, the absence of standardized evaluation metrics makes it difficult to judge whether a federated benchmark is sufficiently reliable for hospital management decisions.

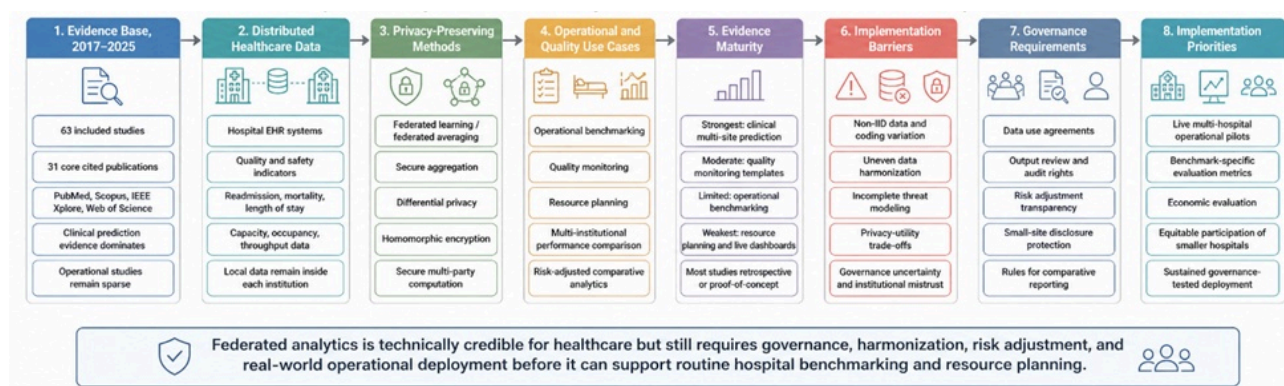
## Barriers and facilitators

Common barriers included governance uncertainty, regulatory interpretation, data harmonization burden, institutional mistrust, uneven technical capacity, and concern that comparative outputs could be misused. Facilitators included open-source platforms, common data models, clear data use agreements, audit mechanisms, local control over data, and evidence that federated workflows can produce useful models across institutions [12, 23, 24]. Governance-focused reviews emphasized that trust, accountability, and participation rules are not secondary issues but core determinants of whether federated healthcare networks can function sustainably [25]. The literature therefore suggests that operational federated analytics should be treated as a socio-technical system rather than only a privacy-enhancing computation method.

## Federated analytics is ready for operational use, but not yet deployed

The reviewed literature indicates that federated analytics is technically credible for healthcare use, but routine operational deployment remains limited. Multi-institutional modeling, privacy-preserving computation, and open-source federated analysis workflows have already been demonstrated in healthcare contexts [3, 6, 12]. However, the strongest evidence still comes from clinical prediction, biomedical analytics, and platform feasibility rather than operational benchmarking at scale. This gap suggests that the field has moved faster in algorithmic readiness than in managerial implementation.

**Figure 2** synthesizes the review findings into an evidence-to-implementation map showing how federated analytics moves from distributed healthcare data and privacy-preserving computation toward operational benchmarking, quality monitoring, resource planning, and governance-tested multi-institutional performance comparison.



**Figure 2.** Evidence-to-Implementation Map of Federated Analytics for Healthcare Operational Benchmarking, Quality Monitoring, Resource Planning, and Multi-Institutional Performance Comparison

## Quality monitoring leads over operations

Quality monitoring appears more mature than operational benchmarking because quality indicators often map onto standardized clinical outcomes such as mortality and readmission. Federated COVID-19 mortality studies and readmission prediction work show that distributed learning can support institution-spanning quality analytics without centralizing sensitive data [14, 15, 19]. In contrast, operational measures such as throughput, discharge delay, and staffing intensity are more locally defined and harder to harmonize. This difference helps explain why federated healthcare studies have advanced more quickly in clinical outcome monitoring than in operational performance management.

## Resource planning is the least explored domain

Resource planning has high practical relevance but remains under-developed in the federated healthcare literature. Length-of-stay prediction and outcome forecasting provide useful components for capacity planning, but they do not by themselves constitute federated staffing, bed management, or equipment utilization systems [14, 18]. Federated architectures for distributed EHR and mobile health data indicate that forecasting across sites is feasible, yet the operational planning layer is rarely evaluated directly [2, 17]. Future work should therefore connect predictive models to actual planning decisions, queueing pressures, and resource allocation workflows.

## Risk adjustment is critical for meaningful comparisons

Risk adjustment is essential because multi-institutional comparison can mislead if differences in case mix, coding, or data quality are mistaken for differences in performance. Federated risk models offer a way to estimate adjustment parameters across multiple hospitals without requiring patient-level data transfer [14, 15, 20]. Personalized and institution-aware federated models may help account for site heterogeneity, but they also complicate the interpretation of shared benchmarks [21]. Operational benchmarking systems will need transparent risk adjustment methods that can be trusted by clinicians, managers, and participating institutions.

## The privacy-utility trade-off

The privacy-utility trade-off remains a central unresolved issue for federated operational analytics. Stronger privacy protections such as encryption, secure aggregation, and differential privacy can reduce leakage risk, but they may also increase computational burden, communication complexity, or analytic noise [6, 11]. Privacy-first health research emphasizes that acceptable trade-offs depend on the sensitivity of the data, the frequency of analysis, and the potential consequences of disclosure [9]. For hospital operations, the key question is not whether privacy can be improved, but whether the resulting benchmarks remain sufficiently stable and actionable.

## Governance and trust in multi-hospital networks

Governance and trust are as important as algorithms in federated healthcare networks. Even when raw data remain local, participants must agree on permissible analyses, access control, auditability, output review, and consequences of comparative reporting [24, 25]. Open-source platforms and FAIR data approaches can lower technical barriers, but they cannot replace institutional agreements about accountability and equitable participation [13, 23]. The evidence suggests that federated operational benchmarking should be governed as a shared infrastructure for learning rather than as a vendor-controlled analytic product.

## From pilot to practice

Moving from pilot to practice will require implementation science studies that evaluate federated operational networks in real hospital settings. Existing implementation and platform papers provide useful starting points, especially for node deployment, workflow design, and infrastructure requirements [12, 13]. However, sustained adoption will depend on whether federated benchmarks are trusted, interpreted correctly, and linked to management decisions. The next phase of research should therefore evaluate not only model utility and privacy protection but also governance durability, user acceptance, and operational impact.

**Table 2** consolidates the review's main evidence gaps by showing why federated healthcare analytics is more mature for clinical quality monitoring than for operational benchmarking, resource planning, and sustained multi-institutional performance comparison.

**Table 2.** Evidence Maturity, Implementation Gaps, and Research Priorities across Federated Healthcare Operations Domains

| Application domain       | Current evidence maturity | What the literature demonstrates                                                                                                                                 | What remains under-developed                                                                                          | Practical risk if implemented prematurely                                                                           | Priority for future research                                                                                |
|--------------------------|---------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------|
| Operational benchmarking | Low to emerging           | Distributed architectures could support comparison of length of stay, throughput, discharge delay, occupancy, and operational variation without raw data pooling | Few studies directly evaluate live benchmarking dashboards, league tables, or recurrent management-facing comparisons | Hospitals may compare unstable or poorly harmonized indicators and misinterpret variation as performance difference | Conduct live multi-hospital pilots using standardized definitions and risk-adjusted operational indicators  |
| Quality monitoring       | Moderate                  | Federated mortality, readmission, and adverse-outcome modeling shows strong transferable value for distributed quality surveillance                              | Routine indicator tracking, longitudinal quality monitoring, and governance-tested reporting remain limited           | Quality indicators may be technically feasible but not trusted or integrated into improvement workflows             | Evaluate recurring federated quality-monitoring cycles linked to quality committees and improvement actions |
| Resource planning        | Low                       | Length-of-stay and outcome prediction studies provide partial foundations                                                                                        | Direct evidence for staffing, bed allocation, equipment                                                               | Predictive outputs may not translate into actionable planning decisions                                             | Link federated forecasts to real planning workflows, staffing                                               |

|                                            |                                                       |                                                                                                                                         |                                                                                                                                       |                                                                                                                 |                                                                                                                         |
|--------------------------------------------|-------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------|
|                                            |                                                       | for capacity forecasting and bed-demand estimation                                                                                      | utilization, queue management, and capacity planning is sparse                                                                        | or resource allocation changes                                                                                  | decisions, and capacity-management outcomes                                                                             |
| Multi-institutional performance comparison | Low to moderate                                       | Federated platforms and privacy-preserving scoring frameworks show that comparative analytics can be technically organized across sites | Executive-facing dashboards, transparent risk adjustment, dispute resolution, and sustained benchmarking networks are rarely reported | Comparative outputs could damage trust, reputation, or participation if perceived as unfair                     | Develop governance-tested comparison frameworks with output review, audit trails, and transparent case-mix adjustment   |
| Risk adjustment and fairness               | Emerging                                              | Federated models can incorporate patient-level covariates locally while estimating shared adjustment parameters                         | Communication of uncertainty, fairness constraints, and site heterogeneity to operational leaders remains weak                        | Unadjusted or opaque comparisons may penalize hospitals serving more complex or disadvantaged populations       | Establish fairness-aware federated benchmarking methods that report case-mix, uncertainty, and subgroup stability       |
| Privacy and threat modeling                | Emerging but inconsistent                             | Secure aggregation, encryption, multiparty computation, and differential privacy can strengthen privacy beyond data-local training      | Threat models, leakage testing, privacy-budget reporting, and small-site disclosure protections are inconsistently described          | Institutions may overestimate privacy protection and underestimate risks from repeated queries or model updates | Require standardized reporting of threat models, privacy mechanisms, privacy-utility trade-offs, and leakage risks      |
| Data harmonization                         | Moderate as prerequisite, weak in deployment evidence | Common data models and standardized extraction logic can improve comparability across sites                                             | Operational definitions such as discharge delay, staffing intensity, and throughput remain locally variable                           | Benchmarks may reflect documentation and coding variation rather than true performance differences              | Create operational common data definitions for federated benchmarking networks                                          |
| Governance and implementation              | Low to emerging                                       | Open-source platforms and implementation frameworks show that federated infrastructure can be deployed across healthcare networks       | Long-term cost, governance durability, maintenance burden, institutional withdrawal, and dispute management are rarely studied        | Technically functional systems may fail due to mistrust, unclear accountability, or unsustainable workload      | Use implementation science and economic evaluation to study sustainability, participation, cost, and operational impact |

## Limitations

### Review limitations

This review has several limitations. The search was restricted to English-language peer-reviewed publications and may have missed technical reports, standards documents, and local implementation evaluations not indexed in the selected databases. Heterogeneity in study design, application domain, privacy method, and reporting quality prevented meta-analysis and required narrative synthesis [5, 8, 10]. In addition, many included studies were clinically oriented, so their implications for operational benchmarking were interpreted cautiously rather than treated as direct evidence.

### Evidence base limitations

The underlying evidence base is also limited by small networks, retrospective validation, simulated federation, and incomplete reporting of privacy guarantees. Several influential studies demonstrate feasibility across institutions, but fewer provide long-term data on sustainability, maintenance cost, governance disputes, or operational decision impact [11, 14, 15]. Vendor and platform-driven implementations may accelerate adoption, but they can also obscure comparative evaluation if reporting is not standardized [12, 23]. The literature therefore supports cautious optimism rather than definitive claims about routine federated operational benchmarking.

## Comparison with prior reviews

Prior reviews of federated learning in healthcare have largely emphasized clinical artificial intelligence, biomedical prediction, medical imaging, genomics, and digital health modeling rather than operational management. For example, distributed deep learning in medical imaging showed early promise for multi-institutional collaboration, and subsequent systematic reviews synthesized federated learning applications across biomedical data types and clinical use cases [10, 25]. Reviews of healthcare federated learning architectures similarly focused on model development, data privacy, and clinical prediction rather than benchmarking of hospital throughput, occupancy, or discharge performance [8, 26]. This review therefore extends prior work by foregrounding operational and quality analytics as management-facing applications of the same privacy-preserving infrastructure.

This review also differs from earlier work by treating privacy-preserving healthcare analytics as broader than federated averaging alone. Several prior reviews identified federated learning as the dominant paradigm, but the evidence base increasingly includes secure aggregation, homomorphic encryption, multiparty computation, differential privacy, and privacy-first governance frameworks [5, 6, 10]. Recent methodological reviews further show that implementation choices, data harmonization, privacy threat models, and workflow integration are central to whether federated systems can be trusted in healthcare settings [13, 27]. These issues are especially important for operational benchmarking because comparative outputs may affect institutional reputation, resource allocation, and managerial accountability.

A major finding of this review is the discrepancy between the maturity of clinical federated learning and the underdevelopment of operational federated analytics. Clinical studies have demonstrated cross-site prediction for mortality, readmission, imaging, and disease-specific outcomes, while operational applications such as bed-capacity forecasting, discharge-delay benchmarking, staffing analytics, and comparative dashboards remain sparse [13, 14, 18]. Privacy-preserving health research has shown that decentralized analytics can support public health and biomedical discovery, but translation to routine hospital management has not kept pace [9,28]. The field therefore appears ready to move from clinical proof-of-concept toward operational learning networks, provided governance and standardization challenges are addressed.

## Recommendations

### Recommendations for research, hospital consortia, vendors, and regulators

Researchers should prioritize multi-site operational pilots that evaluate federated benchmarking for length of stay, occupancy, throughput, readmission, mortality, safety indicators, and resource utilization using standardized definitions.

Studies should report privacy mechanisms, threat models, privacy budgets where applicable, communication requirements, data harmonization assumptions, and implementation context rather than focusing only on predictive utility [10, 27]. Operational endpoints should be selected with hospital managers and quality leaders so that models produce interpretable outputs linked to actual decisions. Reporting standards should also distinguish simulated federation from live multi-institutional deployment because the operational and governance implications differ substantially [3, 11].

Hospital consortia should establish governance frameworks before technical rollout. Data use agreements, audit rights, output-review procedures, dispute mechanisms, and rules for comparative reporting are necessary to prevent federated benchmarking from becoming a source of mistrust [23, 24]. Institutions should also agree on common operational definitions, risk-adjustment variables, minimum data-quality standards, and procedures for handling small-site disclosure risks. Without this groundwork, even technically strong privacy-preserving systems may fail to achieve sustained participation.

Vendors and health information technology developers should build deployable federated nodes that operate inside hospital firewalls with minimal local engineering burden. Open-source and platform-oriented work shows that reusable infrastructure can make federated analysis more practical, particularly when combined with common data models and reproducible workflows [11, 12, 22]. Systems should support secure logging, local approval controls, privacy-preserving aggregation, and transparent reporting of what information leaves each site. For operational analytics, usability and maintenance requirements may be as important as model accuracy because participating hospitals often have limited informatics capacity.

Regulators and public reporting bodies should provide guidance on acceptable privacy-utility trade-offs for aggregated operational analytics. Federated benchmarking differs from patient-level clinical AI because its outputs often concern institutions, units, or service lines rather than individual treatment recommendations [6, 11]. Guidance should therefore address how much privacy protection is required for aggregated hospital performance indicators, how small-cell risks should be handled, and when federated outputs may be used for public accountability. Clear policy expectations would reduce uncertainty for hospitals considering participation in privacy-preserving benchmarking networks.

## Research gaps

### Real-world operational federated networks

Only a limited number of studies describe live, sustained, multi-hospital federated systems with direct relevance to operational benchmarking. Implementation-oriented platforms and decentralized healthcare networks show feasibility, but most evidence remains focused on model development rather than ongoing operational comparison [12, 13]. Future research should evaluate whether federated analytics can support recurring benchmark cycles, management review meetings, and continuous quality improvement. Such work should also document failure modes, including sites withdrawing, definitions changing, data pipelines breaking, or comparative outputs being contested.

### Economic evaluations

Economic evaluation is a major gap because no cited study provided a full cost-effectiveness analysis comparing federated operational benchmarking with centralized registries or traditional data-sharing models. Federated systems may reduce privacy and legal risks, but they can increase infrastructure, coordination, governance, and maintenance costs [24, 25]. Studies should evaluate total cost of ownership, implementation burden, staff time, security review costs, and the value of faster or more trusted benchmarking. Without economic evidence, health systems may struggle to justify federated infrastructure for operational analytics even when the technical case is strong.

### Equitable participation and data quality

Equitable participation remains underexplored, particularly for smaller hospitals, rural facilities, and organizations with weaker data infrastructure. Federated learning can protect local data, but it does not automatically correct for

missingness, coding variation, unequal feature availability, or limited technical capacity [21, 23]. Smaller sites may contribute less stable estimates or face higher disclosure risk when outputs are stratified too narrowly. Future studies should examine how federated benchmarking networks can include diverse hospitals without reinforcing existing disparities in digital maturity and analytic power.

## Implications

### Implications for research, practice, and policy

For research practice, operational federated analytics requires an applied agenda that treats sustainability, equity, governance, and interpretability as core outcomes. Methodological work has advanced rapidly, but the next generation of studies should assess whether privacy-preserving systems improve organizational learning, not only whether they approximate centralized model performance [7, 28]. Researchers should also develop benchmark-specific evaluation metrics, including stability of rankings, fairness of risk adjustment, robustness to missing data, and acceptability to participating institutions. These metrics are necessary because operational benchmarking can influence institutional behavior in ways that conventional predictive modeling metrics do not capture.

For healthcare practice, hospital alliances should begin with lower-sensitivity operational metrics to build trust before progressing to more reputationally sensitive comparative indicators. Federated pilots could initially examine descriptive utilization patterns, aggregate length-of-stay distributions, or de-identified capacity trends before moving toward risk-adjusted quality indicators and public-facing performance comparisons [18, 20]. This staged approach would allow institutions to test governance, data harmonization, and privacy-preserving workflows in a controlled manner. It would also help leaders understand how federated outputs should be interpreted and acted upon.

For policy, federated analytics offers a possible pathway to reconcile transparency, privacy, and institutional learning in healthcare quality reporting. Public agencies and professional bodies could encourage federated approaches when centralized data pooling is legally, politically, or operationally difficult [9, 26]. However, policy frameworks should require transparent governance, reproducible methods, bias assessment, and meaningful protections against re-identification or punitive misuse. If developed carefully, federated operational analytics could support accountable benchmarking while preserving legitimate institutional and patient privacy interests.

## Conclusion

Federated analytics provides a viable path to collaborative operational improvement without compromising the principle that sensitive healthcare data should remain under local institutional control. The reviewed literature shows that privacy-preserving computation can support multi-institutional learning while reducing reliance on centralized data pooling.

The technical foundation is strong, particularly for federated learning, secure aggregation, homomorphic encryption, multiparty computation, and privacy-first governance architectures. However, translation into routine hospital benchmarking, quality monitoring, resource planning, and multi-institutional performance comparison remains nascent.

The most important gaps are the absence of sustained real-world operational federated networks, the lack of rigorous prospective studies, and the limited evidence on cost, governance durability, and equitable participation. Addressing these gaps will require closer collaboration among researchers, health systems, technology developers, and regulators.

A concerted push is needed to move federated operational analytics from concept to everyday hospital management. With standardized definitions, transparent governance, strong privacy guarantees, and implementation-focused evaluation, federated analytics could become an important infrastructure for trusted healthcare performance improvement.

## Acknowledgements

None

## Conflict of interest

None

## Financial support

None

## Ethics statement

None

Received: 26 Sep 2025

Revised: 03 Nov 2025

Accepted: 28 Nov 2025

Published online: 25 February 2026

## 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

- Brisimi TS, Chen R, Mela T, Olshevsky A, Paschalidis IC, Shi W. Federated learning of predictive models from federated electronic health records. *Int J Med Inform.* 2018;112:59-67.
- Li Z, Roberts K, Jiang X, Long Q. Distributed learning from multiple EHR databases: contextual embedding models for medical events. *J Biomed Inform.* 2019;92:103138.
- Lee J, Sun J, Wang F, Wang S, Jun CH, Jiang X. Privacy-preserving patient similarity learning in a federated environment: development and analysis. *JMIR Med Inform.* 2018;6(2):e7744.
- 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.
- Zerka F, Barakat S, Walsh S, Bogowicz M, Leijenaar RT, Jochems A, et al. Systematic review of privacy-preserving distributed machine learning from federated databases in health care. *JCO Clin Cancer Inform.* 2020;4:184-200.
- Froelicher D, Troncoso-Pastoriza JR, Raisaro JL, Cuendet MA, Sousa JS, Cho H, et al. Truly privacy-preserving federated analytics for precision medicine with multiparty homomorphic encryption. *Nat Commun.* 2021;12(1):5910.

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.

Antunes RS, André da Costa C, Küderle A, Yari IA, Eskofier B. Federated learning for healthcare: systematic review and architecture proposal. *ACM Trans Intell Syst Technol.* 2022;13(4):1-23.

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.

Crowson MG, Moukheiber D, Arévalo AR, Lam BD, Mantena S, Rana A, et al. A systematic review of federated learning applications for biomedical data. *PLOS Digit Health.* 2022;1(5):e0000033.

Aouedi O, Sacco A, Piamrat K, Marchetto G. Handling privacy-sensitive medical data with federated learning: challenges and future directions. *IEEE J Biomed Health Inform.* 2023;27(2):790-803.

Mullie L, Afilalo J, Archambault P, Bouchakri R, Brown K, Buckeridge DL, et al. CODA: an open-source platform for federated analysis and machine learning on distributed healthcare data. *J Am Med Inform Assoc.* 2024;31(3):651-65.

Hu J. Implementing federated learning in healthcare. *arXiv.* 2024.

Vaid A, Jaladanki SK, Xu J, Teng S, Kumar A, Lee S, et al. Federated learning of electronic health records to improve mortality prediction in hospitalized patients with COVID-19: machine learning approach. *JMIR Med Inform.* 2021;9(1):e24207.

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.

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

Xu X, Peng H, Bhuiyan MZ, Hao Z, Liu L, Sun L, et al. Privacy-preserving federated depression detection from multisource mobile health data. *IEEE Trans Ind Inform.* 2022;18(7):4788-97.

Huang L, Shea AL, Qian H, Masurkar A, Deng H, Liu D. Patient clustering improves efficiency of federated machine learning to predict mortality and hospital stay time using distributed electronic medical records. *J Biomed Inform.* 2019;99:103291.

Alvarez-Romero C, Martinez-Garcia A, Vega JT, Díaz-Jiménez P, Jiménez-Juan C, Nieto-Martín MD, et al. Predicting 30-day readmission risk for patients with chronic obstructive pulmonary disease through a federated machine learning architecture on FAIR data: development and validation study. *JMIR Med Inform.* 2022;10(6):e35307.

Li S, Ning Y, Ong ME, Chakraborty B, Hong C, Xie F, et al. FedScore: a privacy-preserving framework for federated scoring system development. *J Biomed Inform.* 2023;146:104485.

Kim TH, Yu JY, Jang WS, Heo SC, Sung M, Hong J, et al. PPFL: a personalized progressive federated learning method for leveraging different healthcare institution-specific features. *iScience.* 2024;27(10):110930.

Zhang W, Zhou T, Lu Q, Wang X, Zhu C, Sun H, et al. Dynamic-fusion-based federated learning for COVID-19 detection. *IEEE Internet Things J.* 2021;8(21):15884-91.

Mateus P, Moonen J, Beran M, Jaarsma E, van der Landen SM, Heuvelink J, et al. Data harmonization and federated learning for multi-cohort dementia research using the OMOP common data model: a Netherlands consortium of dementia cohorts case study. *J Biomed Inform.* 2024;155:104661.

Sinaci AA, Gencturk M, Alvarez-Romero C, Erturkmen GB, Martinez-Garcia A, Escalona-Cuaresma MJ, et al. Privacy-preserving federated machine learning on FAIR health data: a real-world application. *Comput Struct Biotechnol J.* 2024;24:136-45.

Eden R, Chukwudi I, Bain C, Barbieri S, Callaway L, de Jersey S, et al. A scoping review of the governance of federated learning in healthcare. *NPJ Digit Med.* 2025;8(1):427.

Chang K, Balachandar N, Lam C, Yi D, Brown J, Beers A, et al. Distributed deep learning networks among institutions for medical imaging. *J Am Med Inform Assoc.* 2018;25(8):945-54.

Teo ZL, Jin L, Liu N, Li S, Miao D, Zhang X, et al. Federated machine learning in healthcare: a systematic review on clinical applications and technical architecture. *Cell Rep Med.* 2024;5(2):101428.

Zhang F, Kreuter D, Chen Y, Dittmer S, Tull S, Shadbahr T, et al. Recent methodological advances in federated learning for healthcare. *Patterns.* 2024;5(6):100991.

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.