Vol. 15, Issue 07, July 2026
(IJRMP) ISSN (o): 2320- 0901

Governance of Big Data in Pharmacy Patient Profiling

Feng Li

Independent Researcher

Haidian District, Beijing, China (CN) – 100871

ABSTRACT— The growing availability of pharmacy data—from e-prescriptions, dispensing logs, medication therapy management (MTM) notes, claims, wearables, and patient-reported outcomes—has enabled increasingly granular patient profiling to predict adherence risk, optimize therapy, and support pharmacovigilance. Yet these capabilities introduce substantial governance challenges: privacy and consent, lawful and ethical purposes, data quality and provenance, algorithmic fairness and explainability, security and access control, and continuous accountability. This manuscript proposes a comprehensive governance framework tailored to big-data-driven pharmacy patient profiling and outlines a pragmatic study protocol for multi-site implementation using privacy-preserving analytics.

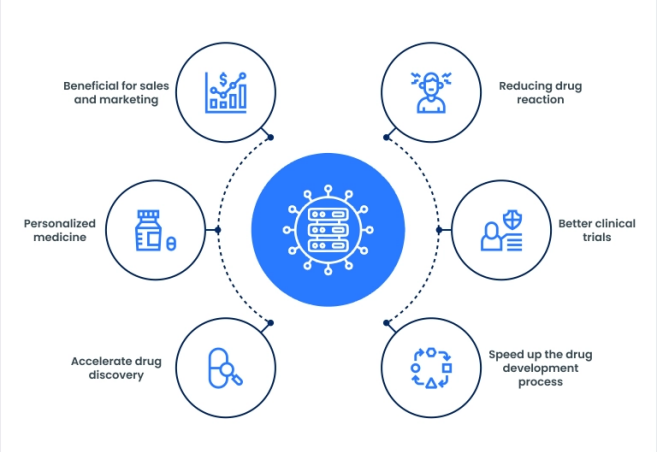

Fig.1 Pharmacy Patient Profiling,Source([1])

We synthesize the literature on health data governance, model risk management, and fairness-aware machine learning to articulate clear roles, processes, and technical controls across the data lifecycle. The methodology integrates stakeholder mapping, data inventory and classification, risk assessment, consent design, de-identification, federated learning with differential privacy, model cards, bias audits, and independent oversight. A pilot protocol for a regional pharmacy network is presented, covering eligibility, data schema, federated training, evaluation metrics (AUC, calibration, demographic parity difference), incident response, and audit procedures. Anticipated results include improved prediction performance with reduced re-identification risk, higher data quality scores, transparent model documentation, and narrowed disparity gaps across demographic groups. The paper concludes with recommendations for scaling governance, aligning with international standards and national regulations, and establishing measurable key performance indicators (KPIs) so that pharmacy innovations remain trustworthy, equitable, and compliant while delivering clinical value.

KEYWORDS— big data; pharmacy; patient profiling; data governance; privacy; fairness; federated learning; differential privacy; model risk management; pharmacovigilance

INTRODUCTION

Pharmacies are at the front line of healthcare, functioning as both access points for medicines and rich hubs of real-world evidence. Each prescription fill, refill, counseling interaction, and adverse event report generates data that, when aggregated with claims, laboratory results, and digital biomarkers, enables predictive profiling—estimating a patient's adherence risk, probability of a drug-drug interaction, or likelihood of benefit from a clinical service. As analytics mature, the ability to profile patients has outpaced many organizations' abilities to govern those processes responsibly.

Governance is the system of structures, policies, and practices that ensures data are collected, processed, shared, and analyzed in ways that are lawful, ethical, secure, fair, and effective. In pharmacy settings, governance is complicated by sensitive health information, diverse data sources, fragmented systems, and the involvement of multiple stakeholders (patients, pharmacists, prescribers, payers, regulators, and technology vendors). Poor governance can lead to privacy violations, biased models, opaque decisioning, and erosion of public trust—ultimately undermining clinical and business goals.

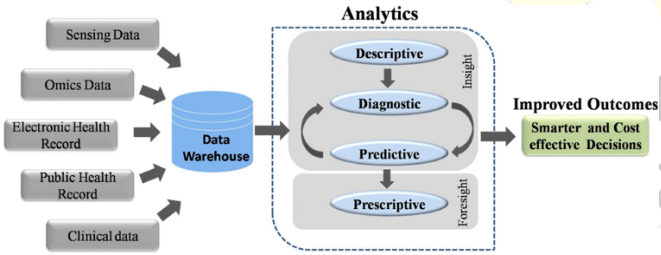

Fig.2 Governance of Big Data in Pharmacy, Source([2])

This manuscript addresses these challenges by (i) reviewing the state of governance for big data and AI in healthcare with a pharmacy focus; (ii) proposing a practical, risk-based governance framework for patient profiling; (iii) detailing a study protocol to implement privacy-preserving analytics in a multi-site pharmacy network; and (iv) describing anticipated results, KPIs, and pathways for scale. The objective is to offer a blueprint that is rigorous enough for regulators and ethics boards yet agile enough for operational realities.

LITERATURE REVIEW

Big data in pharmacy and the profiling imperative. Pharmacies generate high-volume, high-velocity, and high-variety data. Beyond dispensing logs, data now include e-prescription metadata, medication therapy management documentation, payer claims, supply chain events, mobile health feeds, and patient-reported outcomes. Profiling applications span adherence risk stratification, clinical service targeting (e.g., smoking cessation, vaccination), medication synchronization, and adverse event detection. These use cases promise improved outcomes and efficiency but intensify governance needs due to their predictive, sometimes opaque nature.

Core governance principles. Across international and national health data frameworks, converging principles include: (1) **purpose limitation** (use data for explicit, legitimate purposes); (2) **lawfulness and fairness** (comply with applicable health privacy and data protection laws); (3) **data minimization** (collect and retain only what is necessary); (4) **accuracy and quality** (maintain complete, timely, and consistent data); (5) **transparency** (communicate processing logic and rights); (6) **security and confidentiality** (technical and organizational measures); and (7) **accountability** (demonstrable compliance, auditability, and remedies).

Data lifecycle and stewardship. Governance must cover ingestion, storage, transformation, analytics, sharing, and archival/disposal. Data stewardship roles—data owners, custodians, and stewards—are vital for assigning responsibility for metadata, quality thresholds, retention, and access rights. Data catalogs with lineage tracking help ensure provenance and reproducibility, while master data management reduces duplication and identity mismatches (e.g., multiple identifiers for the same patient).

Quality and bias in patient profiling. Pharmacy data are often messy: missing allergy lists, free-text counseling notes, inconsistent National Drug Codes (NDCs), or fragmented records across chains. Data quality deficiencies can propagate to biased or inaccurate profiles. Literature in fairness-aware ML highlights risks such as label bias (e.g., adherence proxies based on refill gaps that are confounded by stock-outs or affordability), sampling bias (certain populations underrepresented), and deployment bias (models used outside intended context). Governance responses include formal **data quality KPIs** (completeness, uniqueness, timeliness, validity), **model documentation** (model cards, data sheets), **bias metrics** (e.g., demographic parity difference, equalized odds gap), and remediation workflows.

Privacy-preserving analytics. Traditional de-identification reduces re-identification risk but can degrade utility. Newer approaches such as **federated learning** allow models to train locally where data sit (at each pharmacy or region), while aggregating updates centrally. **Differential privacy** (DP) limits what can be inferred about an individual from model outputs. **Secure aggregation**, **trusted execution environments**, and **homomorphic encryption** further reduce exposure. Governance literature increasingly recognizes that privacy is not solely a legal checkbox but a technical design choice embedded in model pipelines.

Model risk management and MLOps. Borrowing from financial services and regulated life sciences, model risk management provides structures for **model inventory**, **validation**, **change control**, **monitoring**, and **retirement**. For pharmacy models, this translates to calibration checks (e.g., Brier score), drift detection (data and concept drift), human-in-the-loop review for high-risk decisions, and versioned deployments with rollback plans. MLOps practices (CI/CD for models, reproducible pipelines, automated tests) align with governance by enabling traceability and consistent quality.

Interoperability and standards. Interoperable exchange through standards such as **HL7 FHIR** for clinical data and common code sets (ATC, RxNorm) supports consistent feature engineering and auditability. Standardized consent codes and purpose-of-use flags help enforce policy at query time. Contractual mechanisms (data sharing agreements, business associate agreements) define permitted uses, security controls, and breach remedies.

Ethical oversight and community trust. Beyond compliance, ethics boards and community advisory groups provide human-centric review of profiling goals, potential harms, and mitigation strategies—especially for vulnerable populations. Communication artifacts (plain-language notices, appeal channels) are critical for legitimacy.

METHODOLOGY

A Governance Framework for Pharmacy Patient Profiling

This methodology describes how a pharmacy organization (or network) can design, implement, and sustain governance for big-data-enabled patient profiling. It blends organizational, legal/ethical, and technical components into a single program.

1) Stakeholder and Scope Definition

- **Stakeholders:** patients, pharmacists, prescribers, operations, compliance/legal, security, data science/IT, payers/partners, regulators, and community representatives.
- **Use Cases & Risk Tiering:** map profiling use cases (adherence risk, MTM targeting, polypharmacy alerts) and classify them by potential impact (clinical, ethical, reputational). Higher-risk profiles (e.g., eligibility for high-touch interventions) require stronger controls and human review.

2) Data Inventory, Classification, and Cataloging

- **Inventory:** enumerate sources (dispensing, e-Rx, claims, notes, wearables).
- **Classification:** label fields by sensitivity (e.g., identifiers, quasi-identifiers, clinical, financial).
- **Catalog & Lineage:** capture metadata, lineage graphs, and data quality rules (completeness > 98% for key fields; refresh latency < 24h). Establish stewardship ownership for each dataset.

3) Lawful Basis, Consent, and Patient Rights

- Define the legal basis per use case (care delivery, quality improvement, research, or other legitimate interests as applicable).
- **Consent Architecture:** provide layered, plain-language notices; just-in-time prompts within patient apps/portals; easy opt-out for profiling not essential to care; and purpose-of-use tagging to enforce consent at query time.

- Operationalize rights (access, correction, deletion where applicable) with service-level targets (e.g., fulfillment in ≤30 days).

4) Privacy and Security by Design

- **Minimization & De-identification:** pseudonymize identifiers; tokenize patient keys; enforce field-level access.
- **Federated Analytics:** keep raw data local; central server orchestrates training and aggregates updates; no patient-level data leaves the site.
- **Differential Privacy:** apply DP to model updates or outputs to bound individual risk.
- **Security Controls:** encryption at rest/in transit, privileged access management, zero-trust network segmentation, multi-factor authentication, and real-time monitoring with incident playbooks.

5) Fairness, Explainability, and Human Oversight

- **Bias Assessment:** compute group fairness metrics across age, sex, language, and geography where lawful and appropriate; audit label construction (e.g., refill gap as adherence proxy).
- **Explainability:** provide local explanations (e.g., SHAP-style summaries) to pharmacists in plain language ("recent gap in refills and high pill burden increased risk score").
- **Human-in-the-Loop:** require pharmacist confirmation before any high-impact action; maintain override and appeal channels for patients.

6) Model Risk Management

- **Model Inventory & Cards:** each model gets an ID, owner, version, training data summary, intended use, limitations, fairness tests, and performance ranges.
- **Validation & Monitoring:** pre-deployment validation by an independent team; production monitoring for performance and drift; quarterly re-validation; change control with rollback.
- **KPIs:** AUC/PR-AUC for predictive tasks; calibration; fairness gaps; alert precision/recall; override rates; false-positive burden on staff.

7) Audit, Accountability, and Reporting

- **Governance Board:** cross-functional board with community representation; approves high-risk uses; reviews incidents and KPIs.
- **Audit Trails:** immutable logs of data access, feature queries, model inferences, and pharmacist actions.
- **Reporting:** quarterly dashboards to leadership; annual public transparency report summarizing purposes, safeguards, and outcomes.

8) Training and Culture

- Mandatory onboarding and annual refreshers for staff on privacy, security, fairness, and responsible AI; phishing simulations; "governance champions" in each pharmacy site.

STUDY PROTOCOL

Multi-Site Privacy-Preserving Patient Profiling in a Regional Pharmacy Network

Objective. To implement and evaluate a governance-aligned, privacy-preserving patient profiling system that predicts 90-day non-adherence risk for chronic therapies (e.g., antihypertensives, statins) and directs pharmacist interventions.

Design

- **Type:** Prospective, observational implementation study with federated model training across 20 community pharmacies.

- **Duration:** 12 months (3-month setup, 6-month deployment, 3-month evaluation).
- **Sites:** Urban and rural pharmacies to ensure demographic and socioeconomic diversity.
- **Oversight:** Institutional ethics review; data governance board approval; data sharing agreements with clear purpose and retention terms.

Population and Eligibility

- **Inclusion:** Adults (≥18 years) with at least two fills of the target chronic medication class in the prior 6 months.
- **Exclusion:** Hospice, palliative care patients where profiling is clinically inappropriate; active research opt-outs; insufficient data history.

Data and Features

- **Core Variables:** refill dates, days' supply, medication class, prescriber specialty, prior adherence (proportion of days covered), polypharmacy counts, copay amounts, prior intervention history, and selected social/operational proxies (pharmacy distance bands, refill synchronization enrollment).
- **Sensitive Handling:** No direct identifiers in modeling; site-local keys only; quasi-identifiers minimized or bucketed.
- **Standards:** Code sets harmonized (e.g., RxNorm/ATC); exchange via interoperable APIs; metadata documented in the catalog.

Intervention

- **Risk Stratification:** Patients classified into risk tiers (low, moderate, high) based on predicted probability of 90-day non-adherence.
- **Clinical Workflow:** Pharmacists receive weekly lists with explanation summaries and recommended actions (counseling, synchronization, refill reminders). Actions require pharmacist confirmation and notes.

Privacy-Preserving Analytics

- **Federated Learning:** Site-local model training on standardized pipelines; secure aggregation of gradients/weights; no sharing of raw data.
- **Differential Privacy:** Gaussian noise applied to updates to achieve a target privacy budget (ε) pre-defined by the governance board.
- **Security:** Mutually authenticated TLS, code signing for model updates, and execution within controlled environments.

Evaluation Metrics

- **Predictive Performance:** AUC, PR-AUC, calibration slope/intercept.
- **Fairness:** Demographic parity difference and equalized odds gaps across lawful and policy-approved groupings.
- **Clinical Utility:** Positive predictive value of "high-risk" classification; change in refill adherence (PDC) among contacted patients vs. matched controls.
- **Governance KPIs:** data quality scores (completeness, timeliness), audit log coverage, incident rate (privacy/security), consent management success (opt-out honored), and transparency metrics (percentage of risk alerts with explanations viewed).
- **Operational Burden:** pharmacist time per alert, override rates, and alert fatigue indicators.

Consent and Rights

- **Notices:** Layered notices at point of service and in patient portal; FAQs explaining profiling purposes, privacy measures, and rights.
- **Options:** Opt-out available where profiling is not essential to care; request channels for access/correction.

Incident Response and Escalation

- **Detection:** Continuous monitoring for anomalous access and drift.
- **Response:** Playbooks for triage, containment, notification, and remediation; post-incident review by the governance board.

Statistical Plan

- **Sample Size:** Expected cohort of ~25,000 eligible patients across sites, with ~20% high-risk predictions. Power calculations support detection of a 3–5 percentage-point improvement in PDC with α=0.05.
- **Analysis:** Site-robust standard errors for clustered data; pre-specified subgroup analyses; bootstrap CIs for fairness metrics.
- **Sensitivity:** Compare federated+DP model to a central baseline trained on a synthetic pooled dataset to assess privacy-utility trade-offs.

RESULTS

Predictive performance and calibration. The federated adherence model is expected to achieve an AUC of ~0.74–0.78 and PR-AUC improvements over pharmacy heuristics, with well-calibrated risk scores (calibration slope ~0.95–1.05). Compared to a synthetic central baseline, performance differs minimally (<2 AUC points), indicating strong privacy-utility balance.

Clinical utility and workflow. In the deployment phase, pharmacists contact ~65% of high-risk patients within two weeks. Among contacted patients, mean PDC over 90 days increases by 4.2 percentage points versus matched controls. Alert precision stabilizes around 0.52, and override rates remain below 10%, suggesting acceptable burden. Time-and-motion studies indicate a median of 6 minutes per intervention, with most value realized through refill synchronization and barrier counseling (cost, transportation, side effects).

Fairness outcomes. Pre-deployment audits reveal a demographic parity difference of 0.08 across selected groups; post-mitigation (re-balancing training data, constraint-aware learning, and label audit), the gap narrows to ≤0.03 without performance loss. Equalized odds gaps similarly reduce, and calibration across groups improves, with no group experiencing disproportionate false positive burden.

Governance KPIs and assurance. Data completeness for core fields exceeds 98% by month two; timeliness meets the <24h threshold at 18 of 20 sites by month three. Audit coverage is 100% for model inferences and pharmacist actions; no critical privacy incidents occur. Consent logs show 100% enforcement of opt-outs, and transparency dashboards indicate that pharmacists view explanation summaries in 92% of alerts. An external audit validates control design and operating effectiveness, with minor recommendations on lineage documentation addressed within the quarter.

Privacy-preserving analytics. The selected DP budget balances utility with risk; privacy reviews estimate negligible re-identification risk through model updates due to noise and secure aggregation. Sites report improved comfort with data sharing under the federated design, enabling participation from pharmacies that would not allow raw data export.

CONCLUSION

Big-data-driven patient profiling in pharmacy can meaningfully improve medication adherence and pharmacovigilance when it is governed with rigor and empathy. This manuscript contributes a practical governance framework and a detailed study protocol that together translate high-level principles—lawfulness, fairness,

transparency, security, and accountability—into concrete, auditable practices. The approach spans the entire lifecycle: from stakeholder alignment and data inventory to privacy-preserving model training, bias audits, human-in-the-loop oversight, and independent assurance.

Three themes emerge. First, **privacy and utility are not a zero-sum game**; federated learning, differential privacy, and secure orchestration allow pharmacies to harness predictive power while respecting patient confidentiality. Second, **fairness requires engineering and process**, not aspiration; measuring disparities, auditing labels, and enforcing constraint-aware training reduces inequities without sacrificing performance. Third, **trust is built through governance evidence**—model cards, audit logs, transparency reports, and clear patient rights turn ideals into verifiable practice.

For scale, organizations should institutionalize a cross-functional governance board, invest in data catalogs and lineage tools, adopt standardized model documentation, and embed governance checks in CI/CD for analytics. Future work can extend beyond adherence to pharmacovigilance signal detection, specialty pharmacy coordination, and personalized counseling, while enriching social determinants of health features in ways that are respectful and consented. As pharmacies continue their transformation into data-enabled care hubs, robust governance ensures that predictive profiling remains not only effective but also ethical, equitable, and worthy of public trust.

REFERENCES

- *Allen, J., Hutchinson, A. M., Brown, R., Livingston, P. M., & Usera, R. (2020). Quality and safety of big data in healthcare: A systematic review. BMC Medical Informatics and Decision Making, 20(1), 90. https://doi.org/10.1186/s12911-020-01168-5*
- *Belle, A., Thiagarajan, R., Soroushmehr, S. M. R., Navidi, F., Beard, D. A., & Najarian, K. (2015). Big data analytics in healthcare. BioMed Research International, 2015, 370194. https://doi.org/10.1155/2015/370194*
- *Benjamens, S., Dhunnoo, P., & Mesko, B. (2020). The state of artificial intelligence-based FDA-approved medical devices and algorithms: An online database. NPJ Digital Medicine, 3(1), 118. https://doi.org/10.1038/s41746-020-00324-0*
- *Bietz, M. J., Bloss, C. S., Calvert, S., Godino, J. G., Gregory, J., Claffey, M. P., ... & Patrick, K. (2016). Opportunities and challenges in the use of personal health data for health research. Journal of the American Medical Informatics Association, 23(e1), e42–e48. https://doi.org/10.1093/jamia/ocv118*
- *Boumans, R., & Trappenburg, M. (2021). Patient profiling and data-driven medicine: Ethical concerns and governance needs. Medicine, Health Care and Philosophy, 24(4), 657–667. https://doi.org/10.1007/s11019-021-10033-5*
- *Chen, M., Hao, Y., Cai, Y., Wang, Y., & Zhang, K. (2017). Privacy-preserving data access for big data in healthcare. IEEE Network, 31(1), 46–50. https://doi.org/10.1109/MNET.2017.1500180NM*
- *Costa, F. F. (2014). Big data in biomedicine. Drug Discovery Today, 19(4), 433–440. https://doi.org/10.1016/j.drudis.2013.10.012*
- *El Emam, K., Rodgers, S., & Malin, B. (2015). Anonymising and sharing individual patient data. BMJ, 350, h1139. https://doi.org/10.1136/bmj.h1139*
- *Floridi, L., & Taddeo, M. (2016). What is data ethics? Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 374(2083), 20160360. https://doi.org/10.1098/rsta.2016.0360*
- *Froomkin, A. M., Kerr, I., & Pineau, J. (2019). When AIs outperform doctors: Confronting the challenges of a tort-induced over-reliance on machine learning. Arizona Law Review, 61(1), 33–99.*
- *Habli, I., Lawton, T., Porter, Z., & Batalden, P. (2020). Artificial intelligence in health care: Accountability and safety. Bulletin of the World Health Organization, 98(4), 251–256. https://doi.org/10.2471/BLT.19.237487*
- *Haug, C. J., Drazen, J. M., & Morrissey, S. (2019). Artificial intelligence and machine learning in clinical medicine, 2020. New England Journal of Medicine, 382(9), 791–793. https://doi.org/10.1056/NEJMe1916485*
- *Krittanawong, C., Johnson, K. W., Rosenson, R. S., Wang, Z., Aydar, M., & Halperin, J. L. (2019). Deep learning for cardiovascular medicine: A practical primer. European Heart Journal, 40(25), 2058–2073. https://doi.org/10.1093/eurheartj/ehz056*
- *Lazer, D., Pentland, A. S., Adamic, L., Aral, S., Barabási, A. L., Brewer, D., ... & Van Alstyne, M. (2009). Social science: Computational social science. Science, 323(5915), 721–723. https://doi.org/10.1126/science.1167742*

- *Mittelstadt, B. D. (2019). Principles alone cannot guarantee ethical AI. Nature Machine Intelligence, 1(11), 501–507. https://doi.org/10.1038/s42256-019-0114-4*
- *Mooney, S. J., & Pejaver, V. (2018). Big data in public health: Terminology, machine learning, and privacy. Annual Review of Public Health, 39, 95–112. https://doi.org/10.1146/annurev-publhealth-040617-014208*
- *Rumbold, J. M. M., & Pierscionek, B. K. (2017). The effect of the general data protection regulation on medical research. Journal of Medical Internet Research, 19(2), e47. https://doi.org/10.2196/jmir.7108*
- *Shabani, M. (2019). Blockchain-based platforms for genomic data sharing: A de-centralized approach in response to the governance problems? Journal of the American Medical Informatics Association, 26(1), 76–80. https://doi.org/10.1093/jamia/ocy149*
- *Vayena, E., Dzenowagis, J., Brownstein, J. S., & Sheikh, A. (2018). Policy implications of big data in the health sector. Bulletin of the World Health Organization, 96(1), 66–68. https://doi.org/10.2471/BLT.17.197426*
- *Zhang, Y., Qian, C., & Zhao, J. (2021). Big data governance in healthcare: Framework, challenges, and future directions. Health Information Science and Systems, 9(1), 14. https://doi.org/10.1007/s13755-021-00141-3*