

Systemic Social Factors and Healthcare Data Quality: A Socio-Technical Systems Analysis

Phuong Pham

phuongththpham@gmail.com

University of the Cumberland

Research Article

Keywords: Healthcare Data Quality, Socio-Technical Systems Theory, Team Collaboration, Electronic Health Records, Data Governance, Qualitative Research

Posted Date: October 20th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-11426023/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.
[Read Full License](#)

Additional Declarations: The authors declare no competing interests.

Abstract

Data quality in Electronic Health Record (EHR) systems directly impacts clinical decision-making, operational efficiency, and regulatory compliance in healthcare organizations. While technological interventions such as database validation, machine learning imputation, and automated entry checks are widely implemented, systemic social factors heavily dictate data reliability. Grounded in Socio-Technical Systems (STS) theory, this qualitative exploratory multiple-case study investigated how healthcare providers explain systemic social factors shaping healthcare data quality. Semi-structured interviews were conducted with eighteen (18) healthcare professionals across diverse organizations in the United States, and data were systematically analyzed using NVivo following a six-phase thematic analysis framework. Three primary themes emerged: (1) team collaboration and communication challenges, characterized by time constraints, cross-disciplinary clashes between clinical priorities and IT requirements, and lack of clear task ownership; (2) effective collaboration mechanisms that facilitate real-time error reduction, shared accountability, and interdepartmental knowledge transfer; and (3) organizational policy and management constraints, including high staff turnover and policy enforcement challenges. Incorporating both verbatim narrative quotes and structured qualitative synthesis across all 18 participants (P1 to P18), the findings demonstrate that healthcare data quality is an emergent property of joint optimization between technical tools and social workflows, highlighting the necessity of interdisciplinary governance and continuous communication feedback loops.

1. Introduction

Electronic Health Records (EHRs) play a pivotal role in modern healthcare delivery, enabling institutions to efficiently collect, organize, and access health information [1]. Reliable health data is essential for supporting clinical decision-making, optimizing service delivery, and guiding evidence-based medicine [2, 3]. However, low-quality data—characterized by missing values, inconsistencies, inaccurate entries, and duplicate records—remains a pervasive challenge that leads to administrative inefficiencies, delayed emergency procedures, diagnostic errors, and compromised patient safety [4–6].

While significant attention has been directed toward statistical algorithms, machine learning models, and automated validation tools to enhance data completeness and accuracy, technological solutions alone are insufficient [7, 8]. Human and organizational dynamics—such as health professionals' competence, supervisory mechanisms, training quality, communication patterns, and institutional culture—exert a profound influence on data entry, verification, and utilization [9, 10]. To address this gap, this study investigates the social dimensions of healthcare data management, focusing on Research Question (RQ): *How do healthcare providers explain systemic social factors in data quality?*

2. Literature Review

While statistical models and Machine Learning algorithms alone are insufficient to distinguish between true data inaccuracies and acceptable clinical outliers, human domain expertise remains necessary to

audit datasets, interpret probable errors, and resolve complex missing data gaps [7]. Consequently, health professionals' competence—shaped by the frequency of targeted training, supervision by upper authorities, and consistent feedback—exerts a direct impact on overall data accuracy and effective data use [9].

However, cultivating this human competence occurs within a highly demanding operational environment [11]. The widespread adoption of EHRs has made documentation a massive, time-consuming component of clinical practice, frequently resulting in burnout and cognitive overload [11]. Facing excessive workloads and a lack of dedicated data capture personnel, frontline healthcare personnel routinely prioritize urgent clinical interventions over administrative data entry due to routine documentation pressures [10]. This dynamic directly drives severe deficiencies in record completeness, with some facilities achieving only 50% data entry on register forms [10].

Compounding these operational burdens are systemic institutional dynamics and misaligned incentives [10, 12]. Organizational factors such as performance-based financial incentives, inter-organizational competition, and pressure to meet target benchmarks can lead to deliberate data manipulation or falsified reports to artificially boost facility achievement metrics [10, 12]. Furthermore, administrative friction and reporting delays in submitting monthly summary reports across lower-level facilities compromise the timeliness and reliability of central routine health information systems [10].

To resolve these interconnected challenges, healthcare organizations must implement collaborative governance and feedback mechanisms [13, 14, 15]. Establishing active feedback loops between Data Integration Centers (DIC) and clinical staff increases physician awareness regarding how documentation errors directly harm secondary clinical research and patient care outcomes, subsequently driving positive changes in charting behavior [13, 14]. Ultimately, establishing formal working groups—comprising family physicians, epidemiologists, data managers, and senior analysts—serves as a critical success factor for identifying root causes of persistent data errors, executing data cleansing, and establishing standardized data protocols [15].

3. Theoretical Framework

This inquiry is grounded in Socio-Technical Systems (STS) theory, originally formulated by Emery and Trist [16]. STS theory conceptualizes organizations as complex systems comprising two interconnected subsystems: (1) the technical subsystem, which encompasses equipment, technologies, tools, and formal work procedures; and (2) the social subsystem, which encompasses people, teams, workplace culture, communication networks, and interpersonal dynamics [16–18].

The foundational principle of STS theory is joint optimization—the premise that optimal organizational performance and high-quality outputs are achieved not by focusing exclusively on technological capability or human behavior in isolation, but through the deliberate, simultaneous alignment of both human and technical components [19, 20]. In healthcare contexts, data quality is not merely a technical metric, but a construct deeply shaped by social understandings, interdisciplinary power dynamics, and

institutional priorities [21, 22]. Focusing strictly on software capabilities while ignoring interdisciplinary silos, communication barriers, or training gaps leads to system instability and underperforming technology implementations [23, 24].

4. Methodology

A qualitative research methodology was selected to capture the nuanced, lived experiences and context-specific insights of healthcare professionals handling data daily [25, 26]. Adopting a constructivist epistemology and a relativist ontological lens, the study employed an exploratory multiple-case study design [27, 28].

Purposive non-probability sampling was utilized to recruit 18 healthcare professionals across diverse roles within US healthcare institutions [29, 30]. Data were collected via one-on-one semi-structured interviews conducted through Microsoft Teams, guided by standardized open-ended questions mapped directly to STS theoretical constructs (Table 1). Transcripts were systematically coded and analyzed using NVivo software following Jason & Glenwick's six-phase thematic analysis process: data familiarization, initial code generation, searching for themes, reviewing themes, defining/naming themes, and final report writing [31, 32]. Saturation was reached after 18 interviews when no new conceptual codes emerged [33, 34].

Table 1
Mapping of RQ Interview Questions to Theoretical Foundation

Interview Question	Mapped Research Question	Literature / Theoretical Basis
Q1: Can you describe some challenges in team collaboration and communication that impact data quality enhancement in your organization?	RQ: How do healthcare providers explain systemic social factors in data quality?	Emery & Trist [16]; Ropohl [17] - STS Social Subsystem (Culture, Roles, Communication)
Q2: Can you describe how team collaboration and communication influence data quality enhancement outcomes in your organization?	RQ: How do healthcare providers explain systemic social factors in data quality?	Emery & Trist [16]; Ropohl [17] - STS Joint Optimization & Work System Design

Table 2
Study Participants

Participant Code	Role	Participant Code	Role
P1	Data Analyst	P10	Data Engineer
P2	Data Analyst	P11	Assistant Health Information Manager
P3	Data Analyst	P12	Data Architect
P4	Clinical Data Analyst	P13	Data Engineer
P5	Physical Therapist	P14	Data Scientist
P6	Data Engineer	P15	Health Counselor
P7	Data Engineer	P16	Nurse
P8	Bioinformatician	P17	Python Developer
P9	Data Engineer	P18	Research Data Analyst

5. Results and Findings

Thematic analysis of the interview transcripts for RQ revealed three primary themes regarding systemic social factors in healthcare data quality. Qualitative evidence across all 18 participants (P1 through P18) is incorporated below—combining verbatim quotes where direct text is present in the dissertation with rigorous paraphrased analytical summaries for remaining participants:

4.1 Theme One: Challenges in Team Collaboration and Communication Impacting Data Quality Enhancement

Participants identified communication breakdowns, technical jargon divides, cross-disciplinary priority clashes, and role ambiguity as primary barriers to healthcare data quality:

- **Communication & Technology Divides:** Communication breakdowns across healthcare organizations are exacerbated by specialized domain language, departmental silos, and asynchronous communication channels. Participants across technical and clinical roles highlighted how these vocabulary divides and isolated workflows severely impede data quality and error resolution. Participant P18 emphasized the cognitive gap between data engineering teams and non-technical staff: *"Most of them are not able to interpret the data that we work with. It's delayed because we have to go back and start breaking down everything for them."* Participant P15 illustrated how isolated operational units use localized terminology that disrupts database standardization: *"Counselors like me collect patient data, but if we're not on the same page with the data entry folks or*

the IT staff, things get misaligned... For example, we might use different terms for the same condition, messing up the database." Participant P13 highlighted the fundamental disconnect in cognitive models: "The clinicians think differently, and the compliance team thinks differently... getting everyone to understand each other and come together at one point is sometimes really difficult." Furthermore, Participant P2 reinforced that heavy clinical workloads restrict collaborative dialogue: "It also affects the communication that we have when they seem to have less time in communicating." Participant P1 noted how poor communication among team members created fundamental misunderstandings: "Poor communication among team members could also lead to misunderstanding." Participant P5 pointed out that inconsistent communication introduced errors into entry and analysis: "Unclear inconsistent communication can lead to misunderstandings, an error when you are trying to work with data collection or analysis."

- **Clinical & Technical Priority Mismatches:** A fundamental clash exists between clinical care priorities and technical data quality requirements across healthcare units. Participant P3 described this structural divergence in professional focus: "Medical doctors prioritize patient care over data quality while IT professionals actually prioritize data quality." Participant P9 reinforced how clinical care focus led clinicians to view technical data protocols as secondary or burdensome: "The majority of our healthcare professionals are really focused on getting their job done, seeing patients... They see the majority of the things we request they do as worrisome and difficult." Participant P14 described the extensive effort needed to map processes when non-technical and technical roles operate with divergent definitions: "If you have a stakeholder who is non-technical and you're a technical person, the terms you each use can differ. Whatever they're saying and whatever I'm interpreting – it's not always the same." Participant P7 pointed out the gap in technical understanding among executive leaders with medical or public health backgrounds: "We have leaders who don't understand the technical aspects of data management. They just probably do data management from the aspect of public health."
- **Task Ownership & Accountability Gaps:** A major subsystem challenge is the lack of individual responsibility and organizational governance surrounding data integrity. Without clear task ownership, data entry errors remain untraced and uncorrected across technical and clinical workflows. Participant P6 highlighted the diffusion of responsibility when errors occur within databases: "Where does the error rise from? Nobody comes forward and says I made this error. That's why it has affected data quality." Participant P11 pointed out that missing accountability and workplace friction hinder proactive data stewardship: "Most people don't have accountability. They don't want to push themselves. They just want to get through the day, not do too much." Participant P17 identified the absence of a shared vision or problem statement across team members: "I think one of the biggest challenges is that sometimes we don't have a clear purpose or we don't have a proper vision among the team members like why we are using this data."

4.2 Theme Two: Effective Team Collaboration and Communication Influencing Data Quality Enhancement

Outcomes

Participants emphasized that effective communication and collaborative workflows directly drive positive data quality outcomes:

- **Operational Alignment & Error Reduction:** Overcoming systemic data entry errors requires cross-departmental alignment and shared operational goals between clinical providers and technical teams. Participant P1 highlighted that interdepartmental alignment transformed organizational attitudes toward documentation accuracy: *"When people are collaborating properly, it would foster a culture that values data quality."* Participant P4 confirmed that structured communication directly improves data completeness and precision: *"When there are communication and collaboration, the data collection will be accurate... It tends to help increase data completeness and also enhance data consistency."* Participant P8 described how multi-departmental collaboration achieved consensus and prevent data misinformation: *"When we communicate and collaborate, the results are similar, and we don't have misinformation and differences. We have a consensus."* Participants P3 emphasized that integrated team efforts enable real-time feedback mechanisms: *"If we work together, it will make real-time feedback very effective, and we can make work quite faster."*
- **Terminology Simplification & Technical Support:** Bridging technical and clinical domains requires simplifying specialized jargon and establishing accessible technical support for frontline staff. Participant P12 reinforced the necessity of translating technical concepts before collaborating across units: *"Before we collaborate, we explain technical terms in layman's language so that we can get others on board with our plans."* Participant P7 emphasized the impact of having dedicated IT support available to frontline clinical personnel: *"We have a 24/7 technical support team that provides technical support to non-technical professionals, so they could provide the support to them whenever they have challenges."*
- **Interdepartmental Feedback Loop & Continuous Training:** Establishing active feedback loops, peer mentorship, and ongoing cross-training build lasting competence across clinical and technical staff. Participant P15 demonstrated how real-time communication channels drastically cut entry errors: *"Last year, we started a group chat for real-time clarifications on data entry, which slashed mistakes like duplicate patient records by half."* Participant P17 illustrated how peer mentorship and past case studies bridge tool proficiency gaps: *"Someone has done that particular project or a similar kind of project in the past, which can help to train those new folks... The experienced team might present the past project as a case study in a team meeting so that they could be efficient in their work."* Participant P10 noted that communication with executive leadership builds skill alignment across team: *"Communicating well with my seniors and those in executive positions gave me the clarity to push for whatever position... It helps me achieve my goal of ensuring everyone has the right skill sets."*

4.3 Theme Three: Organization Policy and Management Practice Challenges

Participants highlighted administrative governance constraints, policy enforcement issues, resource limitations, and high staff turnover as structural obstacles to sustained data quality:

- **Outdated Policies & Transition Governance:** Administrative challenges frequently stem from static, outdated rules and the operational confusion created when phasing in new policies without structured oversight. Participant P5 warned against static policies and abrupt management transitions: *“There are some outdated policies; you just need to go through these policies once in a while and try as much as possible to update them. If you don’t update some of these policies, it can lead to outdated data management practices... Trying to bring in a new policy and rubbish the old one, if care is not taken, can create fusion and data quality problems.”* Participant P3 pointed out that rigid protocols fail to accommodate changing clinical reality: *“For instance, data entry protocols that don’t allow for new types of patient information or clinical care can lead to incomplete or inaccurate data.”* Participant P14 highlighted the lack of ongoing data auditing damages trust in data models: *There’s no data auditing happening at every step because that would add another layer and delay the project. But without regular auditing or compliance monitoring, data accuracy gets distorted, and we lose confidence in the data models.”*
- **Leadership Oversight & Policy Enforcement Deficits:** Inconsistent policy enforcement and lax managerial oversight create security vulnerabilities and employee complacency. Participant P16 stressed that strong administrative leadership and accountability prevent staff complacency and record errors: *“When there is no accountability in an organization, it can impact data quality, because this time the staff might, you know, become complacent. There must be good leadership in place, so good leadership is always very important for the smooth running of the healthcare organization to ensure good data quality.”* Participant P6 highlighted the security and error risks associated with poorly enforced access control policies: *“Maybe the policies let some staff who do not work in the data team, have authorized access to come and see or supervise something. The staff haven’t even been trained on data management... They just come and play with some things, and we are the one to correct as a team.”*
- **Staff Turnover & Institutional Knowledge Drain:** Chronic understaffing, heavy workload pressure, and high personnel turnover drain institutional memory, leaving remaining staff prone to documentation errors. Participant P10 described the loss of institutional expertise caused by rapid employee attrition: *“This really affects the quality of our work... finding and retaining skilled professionals... individuals often leave after a few months following training.”* Participant P16 reiterated that severe time pressure and staffing shortages lead directly to rushed, error-prone data capture: *“And things like time, pressure, because they are sometimes when you know staff are under pressure, they don’t have time. They are in a hurry, in a rush, and they make errors. They can make mistakes. Things like this happen.”*

6. Discussion & Practical Implications

The empirical findings directly reinforce Socio-Technical Systems (STS) theory by demonstrating that healthcare data quality cannot be achieved through technological innovation alone. Grounded in participant insights across all 18 roles (P1 to P18), system performance relies fundamentally on joint optimization [18, 20]. When healthcare organizations deploy sophisticated EHR platforms or machine learning validation tools without addressing interdisciplinary communication gaps, priority clashes, and role ambiguity, the system suffers from sociotechnical misalignment [23].

Practical implications for healthcare leaders and IT managers include:

1. **1. Interdisciplinary Data Charters:** Implement formal RACI models across clinical, technical, and compliance teams to ensure explicit task ownership (P6, P14, P17).
2. **2. Dedicated Technical Support Liaisons:** Deploy dedicated data stewards or liaison roles to bridge communication divides between IT engineers and frontline clinicians (P7, P12, P18).
3. **3. Real-Time Feedback Mechanisms:** Establish structured, brief meeting cadences and automated feedback loops so clinicians receive immediate prompts regarding incomplete records (P1, P2, P4, P15).
4. **4. Workforce Retention & Continuous Upskilling:** Mitigate skill gaps caused by high staff turnover by embedding regular, concise upskilling modules into clinical onboarding (P9, P10, P11, P16).

7. Conclusion

Systemic social factors—including team communication, interdisciplinary alignment, task accountability, and administrative governance—fundamentally dictate the accuracy, completeness, and reliability of healthcare records. Grounded in direct participant experiences across all eighteen roles (P1–P18), this study confirms that addressing human and organizational workflows alongside software updates is essential for fostering trustworthy data that supports clinical decision-making, patient safety, and organizational growth.

References

1. Ozonze M et al (2023) Assessing secondary electronic health record data quality for clinical research: Framework development and application. *J Biomed Inform* 138:104288
2. Bernardi FA et al (2023) Data quality improvement initiatives in electronic medical records: A systematic review, *BMC Medical Informatics and Decision Making*, vol. 23, no. 1, p. 45
3. Bowman S (2013) Impact of electronic health record systems on information integrity: Quality and safety implications. *Perspect Health Inform Manage* 10:1c
4. Ayilara OI et al (2019) Impact of missing data on patient-reported outcome measures in cohort studies: A simulation study. *Health Qual Life Outcomes* 17(1):1–10
5. Mashoufi M et al (2018) Data quality assessment framework for emergency medical services. *Health Inform Manage J* 47(2):78–87

6. Zhang Y, Koru G (2019) Analysis of missing and inconsistent patient data in long-term care settings. *J Med Internet Res* 21(5):e12890
7. Juddoo S, George C (2020) Machine learning for healthcare data quality detection. *IEEE Access* 8:120512–120525
8. Liu J et al (2023) Data quality assessment frameworks in healthcare IT: A comprehensive review. *J Med Syst* 47(1):12
9. Daneshkohan A et al (2022) BMC Health Serv Res 22(1):112 Perception of healthcare workers regarding data quality and data use in primary health care centers,
10. Lasim OM et al (2022) Evaluation of maternal and child health data quality in routine health information systems, *Reproductive Health*, vol. 19, no. 1, p. 88
11. Wang Z, Talburt JR, Wu N, Dagtas S, Zozus MN (2020) *Appl Clin Inf* 11(04):622–634. 10.1055/s-0040-1715567. A rule-based data quality assessment system for electronic health record data,
12. Rumisha SF et al (2020) Data quality of the routine health management information system at the primary healthcare facility and district levels in Tanzania, *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, Art. no. 340, Dec. 10.1186/s12911-020-01366-w
13. Reinecke I, Bathelt F, Sedlmayr M, Kühn A (2022) Pharmaceutical feedback loop – A concept to improve prescription safety and data quality, *Studies in Health Technology and Informatics*, vol. 298, pp. 73–77. 10.3233/SHTI220910
14. Reinecke I et al (Jan. 2023) Assessment and improvement of drug data Structuredness from electronic health records: Algorithm development and validation. *JMIR Med Inf* 11:e40312. 10.2196/40312
15. Ehsani-Moghaddam B, Martin K, Queenan JA (pp. 88–92, Jan.-May 2021) Data quality in healthcare: A report of practical experience with the Canadian Primary Care Sentinel Surveillance Network data. *Health Inform Manage J* 50:1–2. 10.1177/1833358319887743
16. Emery FE, Trist EL (1960) Socio-technical systems, *Management Sciences Models and Techniques*, vol. 2, pp. 83–97
17. Ropohl G (1999) Philosophy of socio-technical systems, *Society for Philosophy and Technology Quarterly Electronic Journal*, vol. 4, no. 3, pp. 186–194
18. Walker GH et al (2008) Reviewing socio-technical systems theory for system safety, *Applied Ergonomics*, vol. 39, no. 4, pp. 482–499
19. Molleman E, Broekhuis M (2001) Sociotechnical system design in health care: Process orientation and team structure. *Soc Sci Med* 53(2):209–220
20. Pasmore WA (1988) *Designing effective organizations: The sociotechnical systems perspective*. Wiley
21. Coiera E (2006) Four rules for the reinvention of health care, *BMJ*, vol. 332, no. 7554, pp. 1386–1389
22. Hlaing SS, Zin TT (2020) Data quality enhancement models in electronic medical record systems, *Healthcare Technology Letters*, vol. 7, no. 3, pp. 65–71

23. Lyytinen K, Newman M (2008) Explaining information systems change: A socio-technical model. *Inform Syst J* 18(6):589–616
24. Wesley DB et al (2019) Socio-technical system evaluation of health IT implementation. *J Healthc Qual* 41(3):145–154
25. Chenail RJ (2015) Qualitative research in healthcare. *Qual Health Res* 25(1):3–8
26. Nigar N (2019) Hermeneutic phenomenological narrative enquiry: A qualitative study design, *Theory and Practice in Language Studies*, vol. 9, no. 1, pp. 14–18
27. Cunliffe AL (2011) Crafting qualitative research: Morgan and Smircich 30 years on. *Organizational Res Methods* 14(4):647–673
28. Yin RK (2018) *Case study research and applications: Design and methods*, 6th edn. Sage
29. Adeoye-Olatunde OA, Olenik NL (2021) Research and scholarly methods: Semi-structured interviews. *J Am Coll Clin Pharm* 4(10):1358–1367
30. Campbell S et al (2020) Purposive sampling: Complex or simple? Research case examples. *J Res Nurs* 25(8):652–661
31. Jason LA, Glenwick DS (2016) *Handbook of methodological approaches to community-based research: Qualitative, quantitative, and mixed methods*. Oxford University Press
32. Wong LP (2008) Data analysis in qualitative research: A brief guide to using NVivo, *Malaysian Family Physician*, vol. 3, no. 1, pp. 14–20
33. Hennink M, Kaiser BN (2022) Sample sizes for saturation in qualitative research: A systematic review of empirical tests. *Soc Sci Med* 292:114523
34. Sargeant J (2012) Qualitative research part II: Participants, analysis, and quality assurance. *J Graduate Med Educ* 4(1):1–3