Journal of Information & Systems Management

DLINE JOURNALS

ISSN: 2230 – 8776

JISM 2026: 16 (3)
https://doi.org/10.6025/jism/2026/16/3/93-109

A Normalized Relational Schema for Hospital Management Systems: Design Principles, Network Centralization, and Utilization Phenotypes

Koodichimma Ibe-Ariwa
Cardiff Metropolitan University
United Kingdom

ABSTRACT

Background: Healthcare organizations increasingly require information systems that support not only clinical data storage but also efficient retrieval, integration, and advanced analytics. Poorly designed database schemas introduce redundancy, update anomalies, and integrity risks that compromise both operational efficiency and data quality.

Objective: This study presents a normalized relational database schema for hospital management and demonstrates its capacity to support diverse operational and patient level analyses.

Methods: A fourteen table relational schema was designed applying first, second, and third normal forms (1NF–3NF) with primary and foreign key constraints to ensure referential integrity across patient demographics, staff management, departmental organization, physical resources, appointments, clinical records, surgical events, shift scheduling, and billing. The schema was then evaluated through mediation analysis, moderation analysis, bipartite patient doctor network analysis, four class Gaussian mixture latent profile analysis, group comparisons with effect sizes, and robustness/sensitivity checks.

Results: Mediation of payment through appointment completion was not statistically supported (bootstrap 95% CI:- 53.45 to +6.08). Moderation by age and department was non-significant. The patient doctor network exhibited strong centralization, with one physician accounting for 505 weighted links and 426 unique patients among 726 patients and 1,000 appointment edges. Four distinct utilization phenotypes emerged. Appointment modality showed negligible practical effects on payment or completion ($\eta^2 = 0.0012$; Cramér's V = 0.056). Findings were robust across alternative specifications.

Conclusion: The normalized schema functions as an integrated analytical foundation bridging data management and data services, supporting network analytics, patient phenotyping, and reusable clinical phenotype development for future hospital information infrastructures.

Keywords: Hospital Management System, Normalized Relational Schema, Database Normalization, Mediation Analysis, Network Centralization, Latent-Profile Analysis, Electronic Phenotyping, Appointment Utilization, Referential Integrity

Received: 3 March 2026, Revised: 13 June 2026, Accepted: 30 June 2026

1. Introduction

Electronic health records (EHRs) and other routinely collected healthcare data sources contain a substantial volume of information that can be used to improve the health of individuals and populations [1]. The secondary analysis of these routinely collected data has become increasingly important, particularly as healthcare systems have generated larger and more diverse digital datasets; this trend was further accelerated by the COVID-19 pandemic [2]. Consequently, healthcare organizations increasingly require information systems that can support not only the storage and management of clinical data but also its efficient retrieval, integration, analysis, and reuse.

Effective hospital information systems depend on well structured data models that capture the complex inter relationships among patients, clinicians, resources, and administrative processes. Poorly designed schemas often introduce redundancy, update anomalies, and integrity risks, which can compromise both operational efficiency and data quality in healthcare settings.

This paper describes a normalized relational database schema specifically designed for hospital management. The schema addresses key functional areas: patient demographics and admissions, staff management (doctors, nurses, and support personnel), departmental organization, physical resources (rooms, wards, and beds), appointment scheduling, clinical documentation (vitals, diagnoses, and treatments), surgical records, shift scheduling, and basic billing. Normalization principles (1NF, 2NF, and 3NF) are systematically applied to eliminate redundancy and enforce referential integrity through primary and foreign keys.

The complete schema, together with sample data insertion scripts, is publicly available under the MIT license on Kaggle and GitHub. The primary goals of the resource are educational, supporting instruction in database design and SQL within a healthcare domain, and demonstrative, providing a clean reference model for health information system prototyping.

2. Motivation and Early Work

Healthcare institutions, including hospitals, clinics, and blood banks, can benefit substantially from healthcare management systems that integrate operational and clinical information. Such systems can support medical personnel, including doctors and nurses, by enabling them to manage routine activities more efficiently and by improving patient service delivery through features such as more effective access to information and faster appointment scheduling [3]. The value of these systems therefore extends beyond administrative efficiency to encompass broader improvements in healthcare delivery and service coordination.

The secure exchange of medical information has also been associated with improvements in patients' quality of life, care, and treatment. However, achieving interoperability across the healthcare ecosystem remains a persistent challenge, particularly because the exchange and integration of healthcare information introduce significant security and privacy risks [4]. These challenges emphasize the importance of database architectures that can simultaneously support data consistency, efficient information exchange, analytical use, and secure management of sensitive healthcare information.

Designing database systems that provide both efficient data management and effective data services remains an enduring challenge in healthcare. In many healthcare environments, data management and data services have traditionally been treated as two relatively separate functions. Data services encompass activities such as information retrieval, search, joins, statistical data extraction, and basic data mining operations, whereas data management focuses on constructing error-tolerant, consistent, and non redundant database systems. This separation can create a gap between the operational requirements of data management and the analytical requirements of data services, leading to rigid database structures and schemas that are insufficiently adaptable to advanced analytics [5]. Addressing this gap requires database designs that preserve the advantages of normalization while also enabling richer representations of relationships among healthcare entities.

Within this broader context, electronic phenotyping has increasingly emerged as a distinct area of health data research [6, 7]. Electronic phenotypes and the algorithms used to define them are not merely technical components of an analysis; they represent important first order outputs of health data research. The validity and usefulness of downstream analyses depend substantially on the quality and clarity of the definitions on which they are based. Well defined phenotypes can also function as reusable building blocks across studies and research settings. For example, a phenotype definition developed to identify smoking status in one research study may subsequently be reused by other researchers who require smoking status as a lifestyle risk factor in different analyses [8, 9, 10]. Thus, the systematic representation, management, and reuse of phenotype definitions are important considerations in the development of healthcare information systems.

A practical approach to bridging the gap between normalized relational data management and richer analytical representations was demonstrated by Park, who constructed a graph structure from an abstracted healthcare relational database management system (RDBMS). The study showed that a healthcare graph can be automatically generated from a normalized relational database through the proposed “3NF Equivalent Graph” (3EG) transformation [5]. This approach illustrates how the structural discipline of a normalized relational schema can be retained while relationships embedded in the database can be represented in a graph oriented form that is potentially more suitable for network analysis, relationship exploration, and advanced data services.

Complementing this relational to graph perspective, an Enhanced Entity Relationship Model (EERM) has been proposed for healthcare management systems to represent and manage major hospital functions. The model encompasses the reception section, casualty section, medical treatment details, and employee information, thereby providing a structured representation of the entities and relationships required for hospital wide management [11]. Such an approach highlights the importance of designing healthcare information systems around the full range of institutional activities rather than treating individual administrative or clinical processes as isolated components.

At the same time, conventional relational database approaches face increasing pressure from the volume, variety, and complexity of modern healthcare data. Feng [12] reviewed emerging approaches, including NoSQL databases, cloud computing methods, and AI-enabled systems, in response to the limitations of traditional relational listing methods. These limitations include difficulties in handling unstructured data within structured schemas, scalability challenges as data grows, and restrictions on advanced analytics that rely on querying and data ingestion. The review also considered hybrid approaches and emerging technologies including blockchain, federated learning, edge computing, and semantic knowledge graphs that are contributing to the development of next generation healthcare data infrastructures. These developments suggest that future healthcare database architectures may increasingly combine normalized relational foundations with complementary technologies capable of supporting distributed, heterogeneous, and analytically intensive workloads.

The development and management of clinical phenotypes from EHRs introduces another important dimension of healthcare data infrastructure. Creating computable phenotypes can be resource intensive, which has encouraged the development of phenotype libraries intended to facilitate the reuse of established definitions. Nevertheless, these libraries differ considerably in their target audiences, functionality, and practical utility. Jacqueline described the development of the Centralized Interactive Phenomics Resource (CIPHER), a publicfacing phenotype knowledge base designed to support clinical and health services research. The platform was developed to collect and catalogue EHR-based computable phenotype algorithms from different healthcare systems, scale metadata management, facilitate phenotype discovery, and support integration with tools and user workflows. Phenomics experts were involved in the development and testing of the site [13]. This work demonstrates the growing need for systematic resources that make phenotype definitions discoverable, interpretable, reusable, and accessible across research environments.

There is therefore a widely recognized need for dedicated resources that support the sharing and reuse of electronic phenotypes, and several initiatives have been established to develop phenotype libraries and repositories [14-19]. Chapman [20] proposed requirements for phenotype libraries organized around five key areas: (1) authoring and modeling, (2) refining and logging, (3) implementation, (4) validation, and (5) sharing. These requirements provide a useful framework for assessing whether existing phenotype resources can support the complete lifecycle of phenotype development and reuse rather than merely storing isolated phenotype definitions.

Evaluation of existing libraries against these requirements has shown that, despite substantial progress, important gaps remain in both functionality and scope [20-22]. These gaps indicate that the next generation of healthcare information systems should not focus exclusively on normalized data storage or on isolated phenotype repositories. Instead, they should seek to integrate robust relational database design, relationship-aware representations, interoperability and security mechanisms, scalable analytical infrastructure, and reusable phenotype management. Such integration can help connect operational healthcare data management with advanced analytical services while preserving data quality, consistency, and interpretability.

Overall, the literature points toward a convergent requirement for healthcare information architectures that can support both reliable data management and flexible analytical use. Normalized relational schemas provide an important foundation for reducing redundancy and maintaining data integrity, while graph-based transformations, enhanced entity relationship models, cloud and NoSQL technologies, AI-enabled systems, and semantic knowledge representations provide complementary mechanisms for addressing complex analytical and interoperability requirements. At the phenotype level, reusable libraries and knowledge bases further strengthen the capacity of healthcare researchers to develop, validate, share, and reuse computable definitions. A journal ready healthcare management architecture should therefore be understood not simply as a database for storing hospital records, but as an integrated data infrastructure capable of connecting hospital operations, enabling secure information exchange, providing analytical services, and supporting reusable clinical phenotypes [23-28].

3. Database Design and Schema Description

3.1 Design Principles: The schema was constructed following classical relational design methodology. Entities were identified from core hospital workflows. Functional dependencies were analyzed, and the resulting relations were decomposed to satisfy 1NF (atomic values), 2NF (full functional dependency on the primary key), and 3NF (elimination of transitive dependencies). Foreign key constraints enforce referential integrity, while primary keys uniquely identify each tuple.

3.2 Entity Overview: The schema comprises fourteen interrelated tables. The principal entities and their attributes are summarized below.

• Department (dept_Id PK, dept_Name): Master list of clinical and support specialties (approximately 30 departments in the sample data, including Cardiology, Neurology, Pediatrics, Orthopedics, Oncology, Emergency Medicine, and others).

• Room (room_No PK, dept_Id FK, room_Type): Physical rooms categorized by type (Consultation, Super Deluxe, Deluxe, Standard, Emergency) and linked to departments.

• Doctor (doct_Id PK, dept_Id FK, name attributes, Gender, contact_No, surgeon_Type, office_No FK): Clinician profiles with departmental affiliation and office assignment.

• Nurse and Helpers: Parallel staff tables linked to departments, capturing identity and contact information.

• Ward and Bed: Hierarchical resource structures (Ward linked to Department; Bed linked to Ward).

• Patients (patient_Id PK, demographic and contact attributes including date of birth and address).

• BedRecords and RoomRecords: Admission/discharge transactions recording assigned resources, staff, dates, charges, and payment mode (Insurance, Card, Cash).

• Appointment: Scheduling records linking patients and doctors, with reason, date, payment details, modality, and status.

• MedicalRecord: Clinical encounter data including visit date, vital signs (weight, height, blood pressure, temperature), diagnosis, treatment, and next visit date.

• StaffShift: Temporal scheduling of doctors, nurses, and helpers.

• SurgeryRecord: Operative events capturing patient, surgeon, room, supporting staff, procedure type, timing, and notes.

3.3 Relationships and Integrity: Referential integrity is maintained through foreign key constraints (e.g., doctors and rooms reference departments; admissions reference patients, beds/rooms, and assigned staff). The design supports efficient joins for common analytical queries such as occupancy rates, staff workload, departmental activity, and revenue by payment modality.

An accompanying Entity Relationship Diagram (ERD) visualizes the complete set of relationships and is available in the associated repository.

3.4 Implementation and Sample Data: The schema is distributed as a complete SQL script that creates the database, defines all tables with constraints, and populates them with sample records. Sample data include multiple departments, varied room types, patient demographics, appointment and admission transactions, medical records with vital signs and diagnoses, and surgical events from recent years. The modest scale is intentional, prioritizing clarity and educational usability over volume.

3.5 Applications and Use Cases:

The schema is particularly well suited for:

• Instruction in relational database design and normalization within computer science, information systems, or health informatics curricula.

• Hands-on SQL practice involving multi-table joins, aggregations, and integrity constraint testing.

• Rapid prototyping of hospital information system components or dashboard mock-ups.

• Illustrative case studies comparing normalized versus denormalized designs in healthcare contexts.

It can also serve as a controlled environment for developing and testing queries related to resource utilization, appointment patterns, and basic operational metrics.

4. Methods

4.1 Study Design and Analytical Framework

This study used a quantitative, database oriented analytical design to evaluate a normalized hospital management schema and to examine patterns of hospital service utilization within the resulting relational data structure. The analytical framework combined database design principles with statistical, mediation, moderation, network, latent profile, and robustness analyses. The approach was designed to move beyond assessment of schema structure alone by demonstrating how a normalized hospital database can support operational and patient level analyses.

4.2 Mediation Analysis

A classical mediation model was estimated to evaluate whether appointment completion transmitted the association between online appointment status and payment amount. The model estimated a path from online appointment to completion, the b path from completion to payment, the total effect (c), the direct effect (c'), and the indirect effect (ab). The indirect effect was evaluated using a bootstrap 95% confidence interval. Mediation was considered supported only when the bootstrap interval for the indirect effect did not include zero.

4.3 Moderation Analysis

Two interaction models were estimated. The first tested the interaction between patient age and online appointment status in predicting payment amount. The second tested the interaction between department and online appointment status in predicting appointment completion. Interaction coefficients and their associated significance levels were examined to determine whether the relationship between online booking and the outcomes differed systematically by age or department.

4.4 Patient Doctor Network Analysis

A bipartite patient doctor appointment network was constructed, with patients and doctors as the two node types and appointment relationships as the edges. Degree and weighted degree centrality were calculated to identify concentration of appointment activity. The analysis was used to characterize the distribution of patient doctor interactions and to identify highly connected clinicians who may represent operational hubs.

4.5 Latent-Profile Analysis

Patient utilization patterns were examined using a four class Gaussian mixture model based on appointment volume, completion, inpatient status, surgery, and payment. Because the indicators are continuous and the method is a Gaussian mixture approach rather than a conventional categorical latent class analysis (LCA), the resulting groups are interpreted as continuous density utilization profiles rather than strictly discrete latent classes. The profiles were subsequently characterized according to their dominant patterns of service use and payment.

4.6 Group Comparisons and Effect Sizes

Differences in payment amount across appointment modes (Call, In-Person, and Online) were assessed using both one way analysis of variance (ANOVA) and the Kruskal Wallis test. The association between appointment mode and appointment completion was evaluated using a chi-square test. Effect sizes were reported using eta-squared (η^2), Cramér's V, and pairwise Cohen's d to distinguish statistical significance from practical importance.

4.7 Robustness and Sensitivity Analysis

The association between online appointment status and payment was re-estimated under alternative specifications to evaluate the stability of the findings. These included HC3-robust standard errors, log-transformed payment, and exclusion of the highest 1% and 5% of payment observations. Consistency of coefficient direction and statistical inference across these specifications was used to assess sensitivity to heteroscedasticity, outcome transformation, and influential observations.

4.8 Analytical Interpretation

The analyses were interpreted jointly rather than as isolated statistical exercises. Mediation addressed the potential mechanism linking online booking to payment through completion; moderation assessed whether this relationship varied across patient age or department; network analysis examined structural concentration in provider utilization; latent profiling identified heterogeneous patient service phenotypes; and robustness analyses evaluated the stability of the primary payment association. This integrated approach was intended to demonstrate the analytical utility of the normalized schema for operational and research oriented hospital analytics.

5. Research Issues and Challenges

Healthcare database research faces a fundamental tension between data integrity and analytical flexibility. Highly normalized relational schemas are effective for reducing redundancy and maintaining consistency, but complex analytical questions may require repeated joins across multiple entities. Conversely, highly denormalized or specialized analytical structures may improve query convenience while increasing redundancy and the risk of inconsistency. The present schema addresses this issue by retaining a normalized relational foundation while demonstrating its capacity to support higher level analytical representations.

Interoperability represents a second major issue. Hospital data are generated across clinical, administrative, scheduling, resource, and financial processes, and these domains do not necessarily use identical structures or semantics. The literature reviewed in the manuscript identifies interoperability and secure information exchange as persistent challenges [4]. The schema therefore provides a structured internal representation, but broader interoperability with external healthcare systems remains an important implementation issue.

A further challenge concerns the transition from structured relational records to complex and unstructured healthcare information. The present schema primarily represents structured hospital management information. Modern healthcare environments increasingly incorporate heterogeneous sources, and the literature identifies limitations of traditional relational approaches in handling unstructured data, scalability, and advanced analytics [12]. This creates a need for hybrid architectures that preserve relational integrity while enabling graph, cloud, NoSQL, semantic, and AI-oriented processing.

Electronic phenotyping introduces an additional methodological challenge because phenotype definitions can vary in their construction, validation, portability, and metadata. The literature emphasizes the importance of authoring, refining, implementation, validation, and sharing throughout the phenotype lifecycle [20]. Consequently, a hospital database intended for research reuse should provide mechanisms that enable operational records to be transformed into transparent, reusable phenotype definitions.

The empirical analyses also reveal an operational issue: appointment demand is highly concentrated around a small number of doctors. Such concentration can affect scheduling resilience, workload distribution, and service continuity. The existence of utilization profiles further indicates that patients do not interact with hospital services uniformly. These findings suggest that a hospital information system should support both structural monitoring of provider networks and patient-level segmentation rather than relying exclusively on aggregate counts.

5.1 Research Gaps

First, existing work on healthcare databases often separates database design from downstream analytical use. The present study identifies the need for a more integrated perspective in which normalization, referential integrity, operational querying, network analysis, patient phenotyping, and statistical modeling are considered within a common data architecture [5]

Second, there remains a gap between normalized relational schemas and relationship-oriented representations. Park's 3NF Equivalent Graph transformation demonstrates that relationships embedded in a normalized database can be represented as a graph [5], but further work is required to establish how such representations can be systematically integrated into routine hospital analytics.

Third, phenotype library research has identified continuing gaps in functionality and scope despite the development of multiple phenotype resources [20-22]. The present framework highlights the need to connect phenotype development with the underlying operational database so that computable phenotypes can be derived, validated, documented, and reused more systematically.

Fourth, much of the practical value of hospital information systems is evaluated through operational functionality rather than through integrated empirical analysis of utilization patterns. The present results demonstrate that a normalized schema can support mediation, moderation, network, latent profile, group comparison, and robustness analyses. However, longitudinal validation and larger real world datasets are needed to determine the generalizability of these analytical patterns.

Finally, the current framework does not yet fully address privacy preserving interoperability, distributed learning, semantic integration, or real-time analytics. These areas are particularly relevant given the emerging role of federated learning, edge computing, blockchain, and semantic knowledge graphs in healthcare infrastructures [12].

6. Results

6.1 Base Table

The base table reports the path coefficients from a mediation analysis examining the pathway Online appointment → Appointment completion → Payment amount. It presents the point estimates, associated p-values (where available), and the bootstrap 95% confidence interval for the indirect effect.

Path	Estimate	p_value	Bootstrap_CI_2.5%	Bootstrap_CI_97.5%
a: Online → Completion	0.269163983	0.084384134	-	-
b: Completion → Payment	60.43939102	0.112815488	-	-
c: Total effect	36.46984372	0.267813115	-	-
c2 : Direct effect	39.41058527	0.232335945	-	-
ab: Indirect effect	-16.26810724	-	-53.4471488	6.084642947

Table 1. Basic details

Online appointments are associated with a modest reduction in the probability of appointment completion (path a ≈ − 0.269; odds ratio ≈ 0.76). Completed appointments, in turn, are associated with substantially higher payment amounts (path b ≈ + 60.4 monetary units). The total and direct effects of online booking on payment are positive, yet neither reaches conventional statistical significance. The indirect (mediated) effect is negative (-16.27), but its bootstrap confidence interval includes zero (- 53.45 to +6.08). Consequently, there is no conclusive evidence of mediation: any tendency for online appointments to reduce completion (and thereby lower payments) is offset by a positive direct association, resulting in a net positive but non-significant total effect.

6.2 Mediation analysis

Model: Online appointment → appointment completion → payment amount

Online → completion: coefficient - 0.269, OR ≈ 0.764

Completion → payment: + 60.44 payment units

Total effect of online appointment: +36.47

Direct effect: +39.41

Indirect effect: -16.27

Bootstrap 95% CI for indirect effect: approximately - 53.45 to +6.08

Thus, the mediation pathway is not statistically significant because the bootstrap interval includes zero.

A classical mediation model was estimated with online appointment status as the independent variable, appointment completion as the mediator, and payment amount as the outcome.

The analysis reveals a compensatory pattern. Online booking slightly lowers the likelihood of completion, which would be expected to reduce revenue; however, the direct pathway from online booking to payment is positive and larger in magnitude. Because the bootstrap interval for the indirect effect straddles zero, mediation cannot be claimed. The results suggest that any revenue advantage of online appointments operates primarily through a direct channel rather than through higher completion rates.

6.3 Moderation

Age × online appointment → payment: interaction 0.404, p = 0.727

Department × online appointment → completion: overall interaction p = 0.580

Neither moderation effect is statistically significant.

Moderation Analysis

Two interaction tests were performed: (i) Age × Online appointment predicting payment amount, and (ii) Department × Online appointment predicting appointment completion.

Neither interaction reached statistical significance (Age × Online: interaction coefficient = 0.404, p = 0.727; Department × Online: overall p = 0.580). The association between online booking and the outcomes of interest does not appear to be conditioned by patient age or by the clinical department in which the appointment occurs. The main-effect patterns observed in the mediation model are therefore relatively homogeneous across these subgroups.

6.4 Network analysis

The patient doctor network is strongly centralized:

726 patients

9 doctors

1,000 appointment links

155 patients interacted with more than one doctor.

Doctor 1027 is the dominant network hub, with 505 weighted links and 426 unique patients.

This indicates substantial concentration of patient–doctor interactions around a small number of doctors.

A bipartite patient doctor appointment network was constructed from 1,000 appointment links involving 726 unique patients and 9 doctors. Degree and weighted degree centrality metrics were computed.

The network is highly centralized. A single physician (Doctor 1027) accounts for 505 weighted links and serves 426 unique patients more than half of all patients in the network. Only 155 patients (approximately 21%) consulted more than one doctor. This extreme concentration implies that a small number of clinicians absorb the majority of appointment volume. Such centralization may create operational bottlenecks, elevate the risk of capacity strain for hub physicians, and limit the resilience of the appointment system if a high-volume provider becomes unavailable.

6.5 Latent-profile analysis

A four class patient utilization profile was estimated using appointment, completion, inpatient, surgery, and payment indicators. The resulting groups broadly represent:

- Low utilization/low payment patients

- Higher appointment/high payment patients without surgery

- Surgery oriented lower payment patients

- Higher appointment/high payment patients with surgery

The analysis is technically a Gaussian mixture latent profile proxy, rather than a conventional categorical LCA; this distinction should be retained in a journal paper.

A four class Gaussian mixture model (a continuous-indicator proxy for latent class analysis) was fitted to patient level indicators of appointment volume, completion, inpatient status, surgery, and payment amount.

Four clinically interpretable utilization phenotypes emerged:

1. Low utilization / low payment patients

2. High appointment / high payment patients without surgery

3. Surgery oriented patients with comparatively lower payments

4. High appointment / high payment patients who also undergo surgery

These profiles demonstrate substantial heterogeneity in how patients interact with hospital services. The distinction between high volume surgical and non-surgical high payers, as well as the existence of a low-engagement group, can inform differentiated care pathways, resource allocation, and targeted outreach strategies. Because the solution is based on a Gaussian mixture rather than a traditional categorical LCA, the profiles should be interpreted as continuous density clusters rather than discrete latent classes.

6.6 Statistical significance and effect sizes

- Payment by appointment mode: ANOVA $p = 0.542$, $\eta^2 = 0.0012$

- Kruskal Wallis: $p = 0.538$

- Appointment mode × appointment status: $X^2 = 6.232$, $p = 0.398$, Cramér's V = 0.056

- Pairwise payment effects were also very small:

o Call vs In Person: Cohen's d ≈ 0.034

o Call vs Online: - 0.050

o In Person vs Online: - 0.083

Overall, the effect-size framework suggests very weak practical differences among appointment modes, even after examining several statistical formulations.

Differences in payment amount across appointment modes (Call, In-Person, Online) were tested with both parametric (ANOVA) and non-parametric (Kruskal–Wallis) procedures. The association between appointment mode and completion status was examined using a X^2 test. Standardized effect sizes (η^2, Cramér's V, Cohen's d) were calculated.

None of the omnibus tests reached statistical significance (ANOVA p = 0.542, η^2 = 0.0012; Kruskal Wallis p = 0.538; X^2 p = 0.398, Cramér's V = 0.056). Pairwise Cohen's d values were uniformly negligible ($|d| \leq 0.083$). Collectively, these results indicate that appointment modality exerts, at most, a trivial practical influence on either payment amount or completion probability. Any observed differences are unlikely to be operationally meaningful.

6.7 Robustness/sensitivity

The online appointment coefficient remained positive but statistically nonsignificant under:

- Baseline HC3-robust OLS

- Log-transformed payment

- Exclusion of the highest 1% of payments

- Exclusion of the highest 5% of payments

This indicates that the substantive conclusion is robust to outcome transformation and influential/high-payment observations.

6.7.1 Robustness / Sensitivity Analysis

The coefficient on online appointment status in the payment regression was re-estimated under four alternative specifications: HC3 robust standard errors, log transformed payment, exclusion of the top 1% of payments, and exclusion of the top 5% of payments.

Across all specifications the online appointment coefficient remained positive yet statistically non significant. The substantive conclusion that online booking is not associated with a reliable difference in payment after accounting for other factors is therefore robust to heteroscedasticity, outcome transformation, and the influence of extreme high payment observations.

Figure 1. Network Centrality Description. Visualization of the bipartite patient doctor appointment network, typically rendered as a node size or edge weight diagram highlighting degree or weighted degree centrality. Interpretation. The figure underscores the extreme centralization already quantified in the network metrics: one physician node dominates both connectivity and patient volume. The visual imbalance indicates potential single points of failure and the need for load balancing or capacity planning interventions.

6.8 Latent Profile

Figure 2. Latent Profile Description. Graphical summary (e.g., profile plot or radar chart) of the four latent utilization classes across the indicators of appointment volume, completion, inpatient stay, surgery, and payment. Interpretation. The plot makes the distinct behavioral signatures of the four groups immediately

apparent, facilitating communication of patient heterogeneity to clinical and administrative stakeholders and supporting the design of stratified service models.

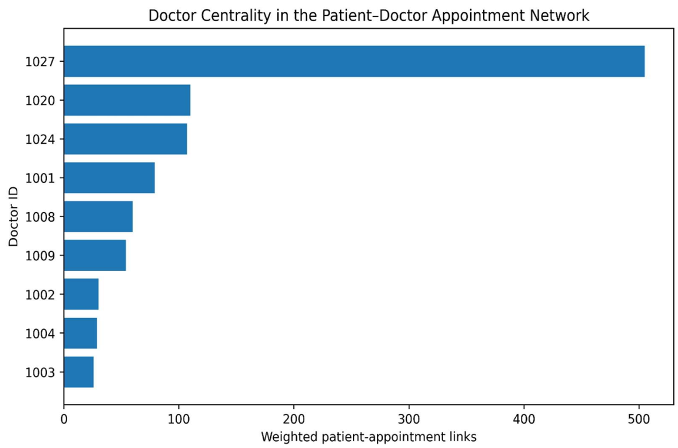

Figure 1. Network Centrality

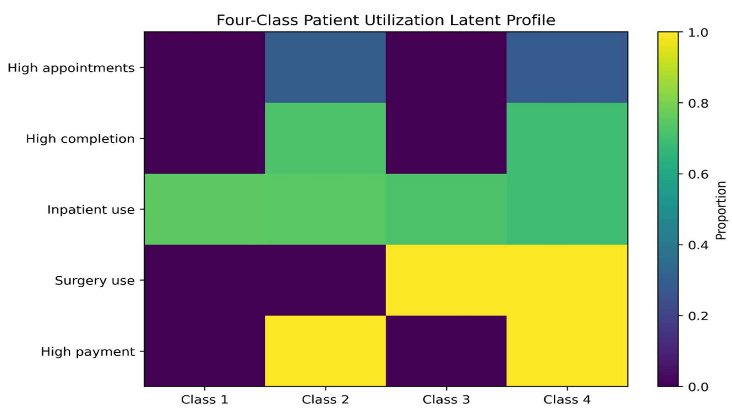

Figure 2. Latent Profile

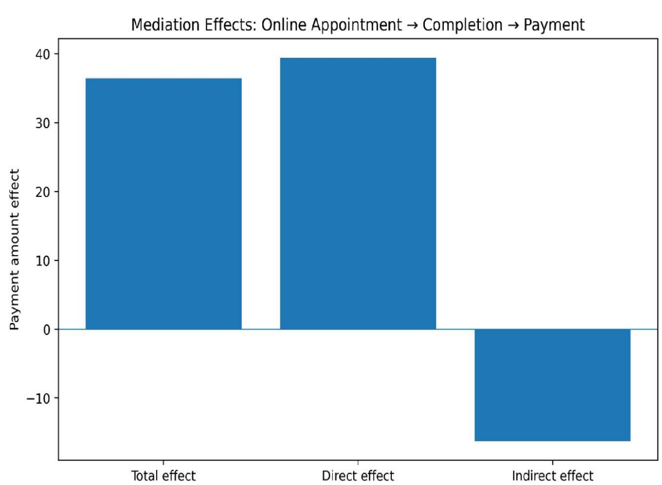

Figure 3. Mediation Effect

Path diagram or coefficient plot displaying the a, b, c, c′, and ab paths together with the bootstrap confidence interval for the indirect effect. The figure visually confirms the absence of statistically supported mediation: the indirect effect interval crosses zero, while the direct and total effects remain positive but non significant. It clarifies that any revenue association with online appointments operates predominantly through the direct pathway.

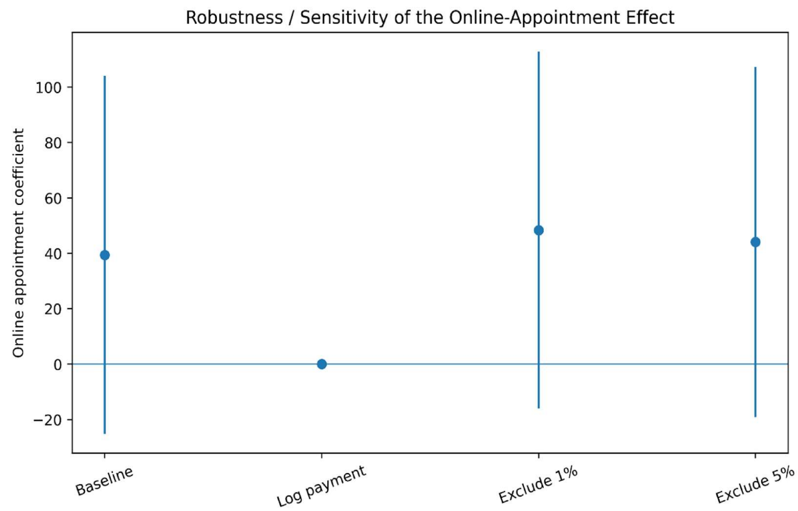

Figure 4. Robustness

Description. Coefficient plot or forest-style display of the online-appointment effect on payment under the baseline and three sensitivity specifications. Interpretation. The consistent direction and persistent non-significance of the coefficient across all robustness checks reinforce the stability of the primary finding and reduce concern that the result is an artifact of model specification or outlier influence.

7. Discussion

The findings demonstrate that the normalized hospital schema is not merely a storage structure but can function as an analytical foundation for investigating hospital utilization and operational relationships. The database design organizes patients, providers, departments, physical resources, appointments, clinical encounters, surgeries, shifts, and payment related records into a referentially coherent structure. This provides a foundation for both routine operational queries and more advanced statistical and network based analyses.

The mediation analysis did not provide statistically conclusive evidence that appointment completion mediates the relationship between online booking and payment. Online appointment status was associated with a modest reduction in the probability of completion, whereas completion was positively associated with payment. The indirect effect was negative (-16.27), but its 95% bootstrap confidence interval (approximately -53.45 to +6.08) included zero. The total and direct effects were positive but not statistically significant. The result therefore suggests a compensatory pattern rather than a confirmed mediation mechanism.

The moderation analyses similarly indicated that the observed relationships were relatively homogeneous across the examined subgroups. The age × online appointment interaction was not significant (0.404, $p = 0.727$), and the department × online appointment interaction was also not significant ($p = 0.580$). Thus, the data do not support the proposition that patient age or department systematically changes the relationship between online booking and the specified outcomes.

In contrast, the network analysis revealed a strong structural pattern. The network comprised 726 patients, 9 doctors, and 1,000 appointment links. Doctor 1027 accounted for 505 weighted links and 426 unique patients, while only 155 patients interacted with more than one doctor. This concentration indicates that hospital appointment activity is not evenly distributed across clinicians. From an operational perspective, such centralization may create capacity constraints and potential single points of failure, highlighting the value of network analytics for hospital resource planning.

The latent-profile analysis provided complementary evidence of substantial heterogeneity in patient utilization. Four broad utilization profiles were identified: low utilization/low payment patients; high-appointment/high-payment patients without surgery; surgery oriented patients with comparatively lower payments; and high-appointment/high payment patients who also undergo surgery. These profiles suggest that aggregate utilization measures may obscure clinically and operationally meaningful differences among patient groups. The profiles may therefore support differentiated service planning, although they should be interpreted as Gaussian-mixture density clusters rather than conventional categorical latent classes.

The appointment-mode analyses showed little evidence of practically meaningful differences. ANOVA ($p = 0.542$, $\eta^2 = 0.0012$), the Kruskal–Wallis test ($p = 0.538$), and the chi-square analysis ($p = 0.398$, Cramér's $V = 0.056$) were all non-significant. Pairwise Cohen's d values were also negligible (approximately -0.083 to 0.034). These findings indicate that appointment modality alone contributes very little to variation in payment or completion within the analyzed data.

The robustness analyses strengthen this interpretation. The online appointment coefficient remained positive but not statistically significant under HC3 robust estimation, log transformed payments, and the exclusion of the highest 1% and 5% of payment observations. Thus, the principal conclusion is not readily attributable to heteroscedasticity, outcome scale, or extreme payment observations.

Taken together, the results illustrate the analytical breadth enabled by the schema. The same relational foundation can support causal pathway exploration, subgroup testing, provider network analysis, patient utilization phenotyping, conventional group comparisons, and sensitivity analysis. This is consistent with the broader literature emphasizing the need to bridge data management and data services rather than treating them as orthogonal activities [5] . It also aligns with the growing importance of reusable electronic phenotypes and phenotype repositories [6, 7].

8. Limitations

Several limitations should be considered. First, the sample and schema are explicitly described as modest in scale and educational in purpose. The database is therefore suitable for demonstrating relational design and analytical workflows but cannot by itself establish generalizability to large, heterogeneous hospital systems.

Second, the empirical analyses are based on the variables represented in the available dataset. Important clinical, socioeconomic, temporal, and organizational factors may not be represented, limiting the extent to which observed associations can be interpreted as comprehensive explanations of hospital utilization.

Third, the mediation analysis is observational. The statistically non significant indirect effect should not be interpreted as proof that no causal mechanism exists. Likewise, the direct and total effects should be understood as associations rather than causal effects.

Fourth, the utilization profiles were obtained using a Gaussian-mixture model. Although these profiles are useful for describing heterogeneous utilization patterns, they should not be presented as conventional categorical latent classes [20].

Finally, the present framework does not provide a complete evaluation of security, privacy, interoperability standards, real time performance, or deployment at institutional scale. These aspects require dedicated implementation and validation studies.

9. Implications and Future Research

The study has implications for health informatics education, hospital information-system prototyping, and research oriented database design. A normalized schema can provide a transparent foundation for SQL training while simultaneously supporting realistic analytical workflows involving resource utilization, appointment behavior, provider networks, and patient phenotypes.

Future work should evaluate the schema using larger and longitudinal hospital datasets, introduce standardized interoperability mechanisms, and test performance under realistic transaction and analytical workloads. Extensions should also investigate graph based representations derived systematically from the relational schema, semantic knowledge models, privacy preserving analytics, federated learning, and AI-assisted clinical and operational decision support. The phenotype component could be extended through explicit metadata, provenance, validation status, versioning, and reusable computable definitions, consistent with the requirements identified for phenotype libraries [20].

A further research direction is to investigate whether the observed provider centralization persists across longer observation periods and whether targeted load-balancing strategies can reduce concentration without adversely affecting continuity or patient outcomes. Similarly, the four utilization profiles could be externally validated and examined for stability across hospitals, departments, and patient populations. Such studies would help determine whether the identified phenotypes are reproducible and operationally actionable.

10. Conclusion

This study presents a normalized relational schema for hospital management and demonstrates how a structured database foundation can support a broad range of healthcare analytics. The design applies 1NF, 2NF, and 3NF principles and uses primary and foreign keys to maintain data integrity across hospital entities and workflows. Empirical analyses show no statistically conclusive mediation of payment through appointment completion, no significant moderation by age or department, strong concentration in the patient doctor appointment network, four distinct utilization profiles, negligible differences across appointment modalities, and stable results under multiple sensitivity specifications. More broadly, the study highlights the value of integrating database design with analytical services, network representations, and reusable patient phenotypes. The resulting framework provides a foundation for future development toward interoperable, scalable, secure, and analytically capable hospital information infrastructures.

References

[1] GOV.UK. (2022). *Data saves lives: reshaping health and social care with data*. https://www.gov.uk/government/publications/data saves lives reshaping health and social care with data/data saves lives reshaping-health and social care with data.

[2] Manktelow, M., Iftikhar, A., Bucholc, M., et al. (2022). Clinical and operational insights from data-driven care pathway mapping: A systematic review. *BMC Medical Informatics and Decision Making, 22*(1), 43. https://doi.org/10.1186/s12911-022-01756-2.

[3] Hajdari, A. (2023). Designing a relational database schema for a Kosovo hospital management system. In *The Software Principles of Design for Data Modeling* (p. 194–210). IGI Global Scientific Publishing.

[4] Villarreal, E. R. D., García-Alonso, J., Moguel, E., Alegría, J. A. H. (2023). Blockchain for healthcare management systems: A survey on interoperability and security. *IEEE Access, 11*, 5629–5652. https://doi.org/10.1109/ACCESS.2023.3236505.

[5] Park, Y., Shankar, M., Park, B. H., Ghosh, J. (2014). Graph databases for large scale healthcare systems: A framework for efficient data management and data services. *2014 IEEE 30th International Conference on Data*

Engineering Workshops (p. 12–19). https://doi.org/10.1109/ICDEW.2014.6818295.

[6] Wei, W. Q., Denny, J. C. (2015). Extracting research quality phenotypes from electronic health records to support precision medicine. *Genome Medicine, 7*(1), 41. https://doi.org/10.1186/s13073-015-0166.

[7] Moskowitz, A., Chen, K. (2016). Defining the patient cohort. In *Secondary Analysis of Electronic Health Records*. Springer. http://www.ncbi.nlm.nih.gov/books/NBK543631/.

[8] Pikoula, M., Quint, J. K., Nissen, F., et al. (2019). Identifying clinically important COPD sub-types using data-driven approaches in primary care population based electronic health records. *BMC Medical Informatics and Decision Making, 19*(1), 86. https://doi.org/10.1186/s12911-019-0805-0.

[9] Riesco, J. A., Alcázar, B., Trigueros, J. A., et al. (2017). Active smoking and COPD phenotype: Distribution and impact on prognostic factors. *International Journal of Chronic Obstructive Pulmonary Disease, 12*, 1989–1999. https://doi.org/10.2147/COPD.S135344.

[10] Pujades Rodriguez, M., George, J., Shah, A. D., et al. (2015). Heterogeneous associations between smoking and a wide range of initial presentations of cardiovascular disease in 1,937,360 people in England: Lifetime risks and implications for risk prediction. *International Journal of Epidemiology, 44*(1), 129–141. https://doi.org/10.1093/ije/dyu218.

[11] Samydurai, A., Revathi, K., Karthikeyan, L., Vanathi, B., Devi, K. (2022). An enhanced entity model for converting relational to non-relational documents in hospital management system based on cloud computing. *IETE Technical Review, 39*(6), 1449–1462.

[12] Feng, Z. (2025). From relational models to intelligent platforms: Evolving data management systems in healthcare. *ITM Web of Conferences, 78*, 03024. EDP Sciences.

[13] Honerlaw, J., Ho, Y. L., Fontin, F., Murray, M., Galloway, A., Heise, D., Connatser, K., Davies, L., Gosian, J., Maripuri, M., Russo, J., Sangar, R., Tanukonda, V., Zielinski, E., Dubreuil, M., Zimolzak, A. J., Panickan, V. A., Cheng, S. C., Whitbourne, S. B., Cho, K. (2024). Centralized Interactive Phenomics Resource: An integrated online phenomics knowledgebase for health data users. *Journal of the American Medical Informatics Association, 31*(5), 1126–1134. https://doi.org/10.1093/jamia/ocae042.

[14] Springate, D. A., Kontopantelis, E., Ashcroft, D. M., et al. (2014). ClinicalCodes: An online clinical codes repository to improve the validity and reproducibility of research using electronic medical records. *PLOS ONE, 9*(6), e99825. https://doi.org/10.1371/journal.pone.0099825.

[15] CALIBER. (n.d.). *Phenotypes*. Retrieved October 27, 2022, from https://www.caliberresearch.org/portal/phenotypes.

[16] Denaxas, S., Gonzalez-Izquierdo, A., Direk, K., et al. (2019). UK phenomics platform for developing and validating electronic health record phenotypes: CALIBER. *Journal of the American Medical Informatics Association, 26*(12), 1545–1559. https://doi.org/10.1093/jamia/ocz105.

[17] ClinicalCodes Repository. (n.d.). Retrieved October 27, 2022, from https://clinicalcodes.rss.mhs.man.ac.uk.

[18] OpenCodelists. (n.d.). Retrieved October 9, 2022, from https://www.opencodelists.org.

[19] Payne, R. (2017). *CPRD codes: severe mental illness* [Data set]. University of Bristol. https://data.bris.ac.uk/data/dataset/7ymb92btsycl11ip5t9pi9zvr.

[20] Chapman, M., Mumtaz, S., Rasmussen, L. V., et al. (2021). Desiderata for the development of next generation

electronic health record phenotype libraries. *GigaScience, 10*(9), giab059. https://doi.org/10.1093/gigascience/giab059.

[21] Azimaee, M., Victor, J. C., Vermeulen, M., et al. (2018). A national concept dictionary. *International Journal of Population Data Science, 3*(4). https://doi.org/10.23889/ijpds.v3i4.709.

[22] OHDSI. (n.d.). *Phenotype library*. Retrieved October 9, 2022, from https://www.ohdsi.org/resources/libraries/phenotype-library.

[23] OHDSI. (2022). *PhenotypeLibrary* [GitHub repository]. Retrieved October 9, 2022, from https://github.com/OHDSI/PhenotypeLibrary.

[24] Strongman, H., Bhaskaran, K., Matthews, A., et al. (2019). *Code lists used in* "Medium and long term risks of specific cardiovascular diseases among 108, 215 survivors of 20 adult cancers: population based cohort study using multiple linked UK electronic health records databases" [Data set]. London School of Hygiene Tropical Medicine. https://datacompass.lshtm.ac.uk/id/eprint/1113.

[25] CPRD @ Cambridge. (n.d.). Codes Lists (GOLD). Primary Care Unit, University of Cambridge. Retrieved October 26, 2022, from https://www.phpc.cam.ac.uk/pcu/research/research-groups/crmh/cprd_cam/codelists/v11.

[26] Thayer, D. S., Mumtaz, S., Elmessary, M. A., Scanlon, I., Zinnurov, A., Coldea, A. I., Scanlon, J., Chapman, M., Curcin, V., John, A., DelPozo Banos, M., Davies, H., Karwath, A., Gkoutos, G. V., Fitzpatrick, N. K., Quint, J. K., Varma, S., Milner, C., Oliveira, C., ... Jefferson, E. (2024). Creating a next generation phenotype library: The health data research UK Phenotype Library. *JAMIA Open, 7*(2), ooae049. https://doi.org/10.1093/jamiaopen/ooae049

[27] This number was not assigned to a reference in your file.

[28] World Health Organization. (2021). *In focus 2021*. https://www.who.int/publications-detail-redirect/9789240053557.