Natural Resources for Human Health
eISSN: 2583-1194

Original Research

Received 07 Jan 2026
Revised 11 Jan 2026
Accepted 29 Jan 2026

Natural Resources for Human Health 2026, Vol. 6 (1s): 803-817

KEYWORDS:
Clinical coding; ICD-10; Coding accuracy; Medical records; Documentation quality; Systematic review

Accuracy of ICD-10 Diagnosis Coding in Hospital Inpatient Records and its Relationship with Medical Record Documentation Quality: A Systematic Review

Abeer Shoei Faqih, Khadijah Shoei Faqih, Noof Hassan Alzahrani, Abdullah Ahmed Alamri, Meshari Fahad Almunimi, Hiba Abdulrahman A. Sobahi, Hamad Abdulrahman Albathi, Maha Adnan Alsalmy, Tahani Mohammad Balkhyour, Ahlam Sami Basudan, Fatimah Sabri W. Ahmed

King Abdullah Medical City, Makkah, Saudi Arabia
Corresponding author: Abeer Shoei Faqih

ABSTRACT:

Background: Coded hospital data underpin reimbursement, quality monitoring, and research, yet the accuracy of coder-assigned International Classification of Diseases, 10th Revision (ICD-10) diagnosis codes vary widely. We synthesized record-level audits of ICD-10 diagnosis coding in hospital inpatient records and the documentation-related error sources reported by their authors.

Methods: PubMed was searched for English-language studies published from 1 January 2009 to 8 December 2025. Candidate reports from the original review and targeted citation checks were also assessed. Eligible inpatient audits compared ICD-10 diagnosis coding with independent re-coding or chart review. Risk of bias was appraised with the Joanna Briggs Institute prevalence checklist. Results were synthesized narratively because outcome definitions and units of analysis differed.

Results: Fifteen studies (16 reports) from eight countries were included. Principal-diagnosis results varied by setting and matching rule. Four-character agreement was 26.3% in a South African pediatric audit. A Malaysian audit found a 49.8% primary-diagnosis error rate. Studies also identified incomplete documentation and coder-related errors.

Conclusions: Principal diagnosis coding accuracy varied across inpatient audits. Incomplete documentation and coder-related errors were recurring findings. Hospitals should combine routine blinded coding audits with documentation standards, including for dictated and transcribed reports, and structured feedback between clinicians, transcription staff, and coders.

1. INTRODUCTION

Clinical coding converts the diagnoses documented by clinicians in a patient's health record into standardized alphanumeric codes, most commonly using the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10) and its national modifications (World Health Organization, 2016). The resulting data feed hospital reimbursement and case-mix systems, quality and safety indicators, disease surveillance, health services planning, and a growing volume of observational research (O'Malley et al., 2005; Quan et al., 2008).

Because these uses assume that each code reflects the patient's true condition, coding accuracy is conventionally defined as agreement between the codes assigned in routine practice and those assigned independently by an expert reviewer working from the same record (O'Malley et al., 2005). Accuracy is more than a technical concern. Countries differ in how the principal, or main, diagnosis is defined, which can by itself alter coded outcomes (Quan et al., 2014), and errors in principal diagnosis selection or in the capture of secondary diagnoses can change severity adjustment and payment groupings.

Earlier systematic reviews of routinely collected discharge data in Great Britain reported that roughly four in five diagnosis codes were accurate, with wide variation between studies and settings (Campbell et al., 2001; Burns et al., 2012). Those syntheses were limited to a single country's health system and to literature published before 2012, and they predate most of the audits conducted in other health systems since.

Studies of the determinants of coding quality point to two broad sources of error: the coder and the source document. Coder characteristics, including experience, training, and workload, have been examined in relation to the validity of discharge data (Hennessy et al., 2010), and qualitative work with coders identifies incomplete chart documentation, variable interpretation of chart content, and production quotas as principal barriers to data quality (Lucyk et al., 2017). Documentation, in turn, reaches the coder through physicians' notes, discharge summaries, and dictated and transcribed reports, so shortcomings in clinical documentation propagate directly into coded data.

Coding practice also differs across settings. A national survey of diagnosis-related group coding in Thailand, for example, illustrates how coder backgrounds, training, and workflows vary between hospitals (Pongpirul et al., 2011), and reliability studies have examined the extent to which independent coders agree when assigning ICD-10 codes to the same record (Stausberg et al., 2008). Together, these observations suggest that pooled estimates of accuracy should be interpreted alongside the setting in which coding takes place.

To our knowledge, no recent synthesis has compared the accuracy of coder-assigned ICD-10 diagnosis codes from record-level audits across health systems while also describing the documentation-related explanations that audit authors report. Such a synthesis is particularly relevant to health information management teams, in which coders and medical transcription staff jointly shape the record from which codes are derived.

We therefore conducted a systematic review of record-level audit studies published from 2009 to 8 December 2025 to (i) estimate the accuracy of ICD-10 diagnosis coding in hospital inpatient records against an independent reference standard and (ii) describe the documentation-related and coder-related error sources reported by the audits.

2. MATERIALS AND METHODS

2.1 Study design and reporting

This systematic review was reported with reference to the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 statement (Page et al., 2021). The protocol was not prospectively registered.

2.2 Eligibility criteria

The review question was framed with the following elements:

Population: inpatient discharge records from hospitals of any size or specialty.

Exposure: diagnosis codes assigned by coders in routine practice with ICD-10 or a national ICD-10 modification (for example, ICD-10-AM or ICD-10-TM).

Comparator: independent re-coding or case-note review by an expert reviewer, used as the reference standard.

Outcomes: primary, the accuracy, agreement, or error proportion of diagnosis codes (principal diagnosis where reported); secondary, documentation-related and coder-related error sources reported by the authors.

Study designs: randomized trials, cohort, case-control, cross-sectional, and case-series studies published in English from January 2009 to 8 December 2025.

Studies were excluded if they addressed procedure coding only or used ICD-9 only; validated codes for a single disease; relied on surveys, questionnaires, or coding exercises rather than record-level audit; evaluated automated or natural-language-

processing coding; concerned emergency department or outpatient records; or were reviews, protocols, uncontrolled before-and-after quality-improvement projects, conference abstracts, or grey literature. Mixed-setting studies were eligible only when inpatient results could be separated.

2.3 Information sources and search strategy

The retrospective search was run through PubMed (NCBI E-utilities) in November 2025 for English-language reports published from 1 January 2009 to 8 December 2025. The complete query, source counts, and decisions on reports assessed for eligibility appear in Appendix S1. Candidate reports from the earlier web-based manuscript search were carried forward after checking for overlap with PubMed. Targeted citation checks identified additional reports, including a companion publication for the Malaysian audit.

2.4 Study selection

Title and abstract screening was conducted for the reconstructed search. Potentially eligible reports were checked against the published full text when it was accessible. Published abstracts were used for journal articles when the full text was inaccessible and the abstract established eligibility. Multiple reports from the same dataset were counted as one study.

2.5 Data extraction

Extracted items were country and setting, design, sample size, classification, reference standard, accuracy and error outcomes, financial impact, documentation-related and coder-related error sources, and the authors' conclusions. For journal articles assessed from published abstracts, extraction was restricted to the details supported by the abstract. Reported counts and denominators were retained with the original unit of analysis.

2.6 Risk-of-bias assessment

Risk of bias was appraised with the nine-item Joanna Briggs Institute checklist for studies reporting prevalence data (Munn et al., 2015), adapted to coding audits: sampling frame (Q1), sampling method (Q2), sample size of at least 200 records (Q3), description of subjects and setting (Q4), coverage of the analyzed sample (Q5), validity of the reference standard (Q6), reliable measurement (Q7), appropriate statistical analysis (Q8), and adequate completeness or response (Q9). Each item was rated yes (Y), no (N), or unclear (U). Studies with seven or more "yes" ratings were classed as low risk, five or six as moderate, and four or fewer as high, with a downgrade of one level when the sampling frame, sampling method, or reference standard was rated "no."

2.7 Data synthesis

Results were synthesized narratively by outcome and unit of analysis. The synthesis distinguished exact code agreement from record-level coding errors, changes to diagnoses within hospital spells, and condition-specific sensitivity or predictive values. A pooled estimate and country-income subgroup analysis were considered during preparation of the review. The outcomes and matching rules differed sufficiently that the final synthesis reports study-specific results.

2.8 Certainty of evidence

The certainty of the principal-diagnosis evidence was considered using GRADE (Guyatt et al., 2008), with attention to risk of bias, inconsistency, indirectness, and imprecision. Observational audit evidence started at low certainty.

3. RESULTS AND DISCUSSION

3.1 Results

3.1.1 Study selection

PubMed returned 346 records and 8 additional reports were identified from the earlier candidate set and targeted citation checks, after matching against PubMed records. After title and abstract screening, 318 PubMed records were excluded. A total of 36 reports were assessed for eligibility and 20 were excluded, leaving 16 reports of 15 studies. **Figure 1** presents the selection flow. Appendix S1 documents the source and decision for each report assessed for eligibility.

3.1.2 Study characteristics

The 15 included studies met the 8 December 2025 publication cutoff and came from eight countries. Most were single-hospital or two-hospital audits. Multicenter studies included an audit across seven English NHS trusts and a Thai national audit covering 534 hospitals. Reference standards included re-coding by senior coders, clinicians, auditors, or expert panels (Table 1).

Table 1. Characteristics of the included studies.

Study	Country / setting	Design	Records audited	Classification	Reference standard / reviewers
(Cheng et al., 2009)	Australia (Victoria); one hospital, surgical discharges	Retrospective internal audit (cross-sectional)	752 surgical inpatient discharges (6 months)	Not stated in accessed record	Blind re-coding of each episode
(Farzandipour et al., 2010)	Iran (Kashan); teaching hospitals	Cross-sectional	370 randomly selected records	ICD-10	Blind re-coding used as gold standard
(Nouraei et al., 2015)	UK; two surgical centers	Two-center audit (cross-sectional)	>30,000 case notes	UK national coding standards	Audit team review; random sample of 400 re-audited independently
(Nouraei et al., 2016)	UK; teaching and district hospital; emergency medical admissions	Retrospective audit	8,888 discharges over 3 years	UK national coding standards	Coding accuracy audit team
(Heywood et al., 2016)	UK (North West England); seven NHS trusts; emergency general surgery	Multicenter retrospective audit	208 discharges (Mar–Aug 2013)	ICD-10 and OPCS-4	Blinded re-coding by a senior coder and a clinician
(Chongthawonsatid, 2017)	Thailand; 534 hospitals (505 public, 29 private)	Retrospective cross-sectional (national audit data)	118,971 records	ICD-10-TM	Trained auditors (physicians, nurses, coders) reviewed records
(Zafirah et al., 2018)	Malaysia; teaching hospital	Cross-sectional	464 records (four disciplines)	ICD-10 (diagnoses); ICD-9-CM (procedures)	Independent senior coder; casemix expert verification
(Tsopra et al., 2018)	UK (Leeds); three respiratory wards	Prospective cohort (comparative)	107 consecutive discharges	ICD-10	Responsible consultant with a coder, using case notes
(Daniels et al., 2021)	South Africa (Cape Town); children's referral hospital	Retrospective cross-sectional	396 folders analyzed (of 450 sampled)	ICD-10 (4-character)	Principal investigator's re-coding from the folder

Study	Country / setting	Design	Records audited	Classification	Reference standard / reviewers
(Oweghoro and Oluwole, 2022)	Nigeria (Ibadan); tertiary teaching hospital	Retrospective cross-sectional	1,882 records (of 2,000 sampled)	ICD-10	Blind re-coding by an independent senior coder
(Khorrami et al., 2022)	Iran (Bandar Abbas); teaching hospital	Cross-sectional	500 records (injury and poisoning)	ICD-10 (compared with ICD-11)	Two independent auditors with at least 3 years' experience
(Zarei et al., 2023)	Iran; two general hospitals	Prospective cross-sectional	1,578 ICD-10 codes	ICD-10 (compared with ICD-11)	Five mentors; their codes served as gold standard
(Hong et al., 2009)	Vietnam, Da Nang General Hospital	Retrospective audit	205 pediatric injury discharges	ICD-10	External re-coding in Vietnam and Australia
(Haliasos et al., 2010)	UK, neurosurgery	Multidisciplinary audit	386 episodes	UK national coding standards	Neurosurgeon review with experienced coder consultation
(Saizan et al., 2021)	Malaysia, two Ministry of Health hospitals	Cross-sectional	226 audited cases	ICD-10 (MalaysianDRG)	Senior Ministry of Health coders; case-mix expert verification

ICD-10-AM, International Classification of Diseases, 10th Revision, Australian Modification; ICD-10-TM, Thai Modification; OPCS-4, Office of Population Censuses and Surveys classification of interventions and procedures, version 4.

3.1.3 Accuracy and error findings

The audits reported differing results for principal diagnosis coding, all diagnoses, and coding or grouping changes per record. The coding agreement for all diagnoses was 89.1% across 1,578 ICD-10 codes in an Iranian audit. Financial changes after re-coding varied between settings. Table 2 presents the results in their original units.

Table 2. Key quantified findings of the included studies.

Study	Key quantified findings
(Cheng et al., 2009)	About 16% of 752 audited episodes showed a diagnosis-related group (DRG) change; the net revenue effect was an increase of about AU$575,300; 56% of DRG-change cases were due to documentation issues.
(Farzandipour et al., 2010)	Principal diagnosis accuracy 77.3% (286/370); 84 errors (22.7%), of which 28 (33.3%) were major.
(Nouraei et al., 2015)	At least one coding change in 51% of patients (15,402/30,127). The audit reported 3,911 changes to primary diagnoses (13%).

Study	Key quantified findings
(Nouraei et al., 2016)	At least one coding change in 55% of discharges (4,889/8,888); the primary diagnosis changed in at least one consultant episode in 16.8% of hospital spells (1,496/8,888); income variance +£816,977 (+5.0%).
(Heywood et al., 2016)	At least one error in 93.3% of cases (194/208); primary diagnosis errors 30.8% (64/208); primary procedure errors 21.9% (30/137).
(Chongthawonsatid, 2017)	Discharge-summary principal diagnoses: sensitivity 7.3–37.9%, PPV 9.2–60.7%, specificity 97.2–99.8%. Coding assessments: sensitivity 31.1–69.4%, PPV 43.8–89.0%, specificity 99.0–99.9%.
(Zafirah et al., 2018)	At least one error in 89.4% of records (415/464); secondary diagnoses 81.3% (377/464); primary diagnoses 49.8% (231/464); DRG changed in 74.0% of records with errors (307/415); potential income loss RM654,303.91.
(Tsopra et al., 2018)	Exact 4-character match of the primary diagnosis: 35% (summary only), 50% (with case notes), 43% (point-of-care with a doctor). Inaccuracy of all codes: 70%, 58%, 60%. Incorrect HRG: 42%, 31%, 35%.
(Daniels et al., 2021)	Primary diagnosis agreement 26.3% at 4 characters and 34.3% at 3 characters; secondary diagnoses 14.9% and 27.7%.
(Oweghoro and Oluwole, 2022)	The authors reported 82.5% correct coding (1,552/1,882) and 98.7% with no discharge summary. The reported categories for incorrect codes used inconsistent percentages and were not combined.
(Khorrami et al., 2022)	98% of records had at least one error (23% major only, 46% minor only, 31% both); 81% of records had adequate documentation for ICD-10 coding.
(Zarei et al., 2023)	ICD-10 accuracy 89.1% across 1,578 codes (lowest by chapter 55.2%).
(Hong et al., 2009)	Main-diagnosis discrepancy was 32%–40% at three characters and 57%–60% at four characters, depending on the coder pair.
(Haliasos et al., 2010)	At least one error in 71/386 episodes (18.4%); 36 diagnosis errors; HRG changed in 40 episodes.
(Saizan et al., 2021)	Coding errors in 57/226 cases (25.2%); inaccurate clinical documentation in 140/226 (61.9%).

DRG, diagnosis-related group; HRG, Healthcare Resource Group; PPV, positive predictive value.

3.1.4 Conclusions and reported error sources

Studies identified absent, incomplete, or ambiguous documentation and difficulties transferring information between clinicians and coders. In the Nigerian audit, nearly all records lacked a discharge summary. In the respiratory-ward study, coders working from the discharge summary alone were least accurate, and adding case notes or medical input improved accuracy. The reported sources of error and authors' conclusions are summarized in Table 3.

Table 3. Authors' conclusions and reported documentation- and coder-related error sources.

Study	Authors' main conclusion	Reported documentation-related and coder-related error sources
(Cheng et al., 2009)	Inadequate documentation is a major driver of miscoding, with consequences for funding.	Documentation issues accounted for most DRG changes.
(Farzandipour et al., 2010)	Most principal diagnosis codes were accurate; coder attention and better documentation can raise quality.	Memory-based coding, abbreviations, and incomplete review of the record increased errors; record readability was not significantly related to accuracy.
(Nouraei et al., 2015)	Coding inaccuracy is common in surgical administrative data, with implications for outcomes management.	Not reported in the accessed abstract.
(Nouraei et al., 2016)	Accuracy depends on high-fidelity information transfer between clinicians and coders; greater clinician involvement is recommended.	Subjective handover of information; ambiguity in strategic diagnoses such as pneumonia and renal failure.
(Heywood et al., 2016)	Errors are multifactorial; surgeons should collaborate with coding departments.	Coder error and the need for clinical interpretation of notes.
(Chongthawonsatid, 2017)	Validity of principal diagnoses and codes was low; training of physicians and coders should be strengthened.	Incomplete physician documentation; coders from diverse backgrounds; reimbursement incentives.
(Zafirah et al., 2018)	Intensive re-training and close monitoring of coders are needed to prevent loss of hospital income.	Coder performance, especially in secondary diagnosis capture.
(Tsopra et al., 2018)	Accuracy improved when coders used case notes or medical input in addition to the discharge summary; summary quality should improve.	Coders' limited medical knowledge; missing codes were the most frequent error; ambiguity between past and current diagnoses; abbreviations.
(Daniels et al., 2021)	Reliability of administrative discharge data was poor.	Lack of dedicated and adequately trained coding personnel.
(Oweghoro and Oluwole, 2022)	Absent discharge summaries are counterproductive; coders need specialist training and clinicians should write summaries.	No discharge summary in nearly all records; on-the-job training only; staff rotation; 84.9% of errors attributed to coders.
(Khorrami et al., 2022)	Documentation was less complete for ICD-11 than for ICD-10; documentation guidelines, templates, and audits are advised.	Incomplete medical record documentation.
(Zarei et al., 2023)	Coder training and documentation guidelines are essential.	Coder mistakes (79.5% of ICD-10 errors) and poor documentation (15.8%).
(Hong et al., 2009)	Coding agreement was limited.	Documentation of external causes of injury was frequently incomplete.

Study	Authors' main conclusion	Reported documentation-related and coder-related error sources
(Haliasos et al., 2010)	Clinical involvement can improve the accuracy of neurosurgical coding.	The findings supported clinician–coder review.
(Saizan et al., 2021)	Documentation and coding errors affected estimated treatment costs.	The audit assessed documentation and coding errors separately.

3.1.5 Principal diagnosis coding accuracy

Farzandipour et al. (2010) reported 77.3% of principal diagnoses coded correctly (286/370). Daniels et al. (2021) reported 26.3% agreement at four characters and 34.3% at three characters. Hong et al. (2009) reported four-character discrepancies of 57%–60%, depending on the coder comparison. Zafirah et al. (2018) found primary-diagnosis coding errors in 49.8% of records (231/464). In the UK audits, Heywood et al. (2016) identified a primary-diagnosis error in 64 of 208 cases (30.8%), and Nouraei et al. (2016) reported at least one primary-diagnosis change in 1,496 of 8,888 hospital spells (16.8%). These results measure different aspects of coding performance.

3.1.6 Certainty of evidence

The certainty of evidence on principal diagnosis accuracy was very low. Estimates varied across audits, matching rules and sampling frames differed, and several studies had unclear methods for sampling or assessing the reliability of the reference standard. These results support variation between the audited settings, but provide limited evidence for a common accuracy estimate.

3.1.7 Risk of bias

The most frequent concerns were unclear sampling methods, limited information on the reliability of the reference standard, and incomplete reporting. The Thai national audit sampled records already flagged as abnormal. Ratings based on published abstracts are provisional. Table 4 presents the study-level appraisal.

Table 4. Risk-of-bias assessment (Joanna Briggs Institute checklist for prevalence studies, adapted to coding audits).

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Yes (n)	Overall risk
(Cheng et al., 2009)	Y	U	Y	U	Y	Y	U	Y	U	5	Moderate
(Farzandipour et al., 2010)	Y	Y	Y	Y	Y	Y	U	Y	U	7	Low
(Nouraei et al., 2015)	Y	U	Y	U	Y	Y	Y	U	U	5	Moderate
(Nouraei et al., 2016)	Y	U	Y	Y	Y	Y	U	Y	U	6	Moderate
(Heywood et al., 2016)	Y	U	Y	Y	Y	Y	U	Y	U	6	Moderate
(Chongthawonsatid, 2017)	N	N	Y	Y	Y	Y	U	Y	U	5	High†
(Zafirah et al., 2018)	Y	Y	Y	Y	U	Y	Y	Y	U	7	Low
(Tsopra et al., 2018)	Y	Y	N	Y	Y	Y	Y	Y	Y	8	Low
(Daniels et al., 2021)	Y	Y	Y	Y	Y	U	U	Y	Y	7	Low
(Oweghoro and Oluwole, 2022)	Y	Y	Y	Y	U	Y	U	N	Y	6	Moderate
(Khorrami et al., 2022)	U	U	Y	Y	Y	Y	U	U	U	4	High
(Zarei et al., 2023)	Y	U	Y	Y	Y	Y	U	Y	U	6	Moderate
(Hong et al., 2009)	Y	Y	Y	Y	Y	Y	U	Y	U	7	Low

Study	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Yes (n)	Overall risk
(Haliasos et al., 2010)	U	U	Y	Y	Y	Y	U	Y	U	5	Moderate
(Saizan et al., 2021)	U	Y	Y	Y	Y	Y	U	Y	Y	7	Low

Y, yes; N, no; U, unclear. Q1, sampling frame appropriate; Q2, sampling method appropriate; Q3, adequate sample size (200 or more records); Q4, subjects and setting described; Q5, sufficient coverage of the analyzed sample; Q6, valid reference standard; Q7, reliable measurement; Q8, appropriate analysis; Q9, adequate completeness or response. †Downgraded one level because the sampling frame and sampling method were rated "no."

3.2 Discussion

Clinical coding is the step at which a clinician's description of illness becomes data, and the quality of that step determines how far coded data can be trusted for payment, benchmarking, and research. This review brought together record-level audits from eight countries and found substantial variation in principal diagnosis coding results across settings and matching rules. This pattern is consistent with earlier syntheses of British discharge data that reported generally good but highly variable accuracy (Campbell et al., 2001; Burns et al., 2012). The present review extends that finding to a wider range of health systems.

The specialty audits that make up much of the literature point in the same direction. Audits have examined coding in otolaryngology, oral surgery, and neurosurgery (Nouraei et al., 2009; Naran et al., 2014; Haliasos et al., 2010). Coding practice under payment by results also has practical implications for surgeons (Jameson and Reed, 2007). The shared message is that coding discrepancies carry budgetary as well as informational consequences. The included studies add a nuance to this message. In some hospitals, errors lowered income, whereas in others, re-coding would have raised it, so inaccuracy cannot simply be equated with financial loss. What the errors have in common is that they weaken the link between a patient's clinical picture and the data that represent it, which matters for case-mix adjustment and quality indicators regardless of the direction of the financial effect.

Differences in audit populations, reference standards, and matching rules limit comparisons by country income group. Coding in the United Kingdom operates within a national payment system, is performed by coders who hold a national qualification (Tsopra et al., 2018), and is subject to routine clinical audit (Burns et al., 2012). A South African children's hospital audit attributed poor reliability to a shortage of dedicated, adequately trained coders (Daniels et al., 2021). A Nigerian audit described a unit whose staff learned coding on the job and rotated out as experience accumulated (Oweghoro and Oluwole, 2022). In most studies that reported both, secondary diagnoses were captured less accurately than principal diagnoses, (Zafirah et al., 2018), because every additional diagnosis must be found, judged relevant, and coded, and each step is an opportunity for omission.

Mechanistically, the studies converge on the source document. Coders do not observe patients; they translate what has been written. When the discharge summary is absent, as in the Nigerian audit (Oweghoro and Oluwole, 2022), or is the only document consulted, as in the respiratory-ward study (Tsopra et al., 2018), coders must infer diagnoses from drug charts, laboratory values, and fragments of notes, and they omit or misclassify what they cannot interpret. Providing the full case notes, or a clinician's interpretation of the summary, improved accuracy in that study, which shows that the information was in the record but not in a form a non-clinical coder could use. Abbreviations, ambiguous entries, and confusion between past and current diagnoses produced the same effect in two audits (Farzandipour et al., 2010; Zarei et al., 2023). These are exactly the failures at which dictation, transcription, and editing intervene, because a transcribed report passes through a human reader before it reaches the coder, and that step can either resolve ambiguity or introduce it. Wider evidence supports this reading. Deficiencies in discharge summaries and medical documentation can alter diagnosis-related group assignment (Chin et al., 2013). Other studies documented variation in discharge summaries written by junior physicians (Legault et al., 2012), differences in information captured by electronic and handwritten summaries (Callen et al., 2008), and gaps in information transfer at discharge (Kripalani et al., 2007).

Coder-related mechanisms form the second theme. Memory-based coding, hurried review of the record, limited clinical knowledge, and production pressure all erode accuracy. Qualitative work has identified chart documentation, variable interpretation, and quotas as the principal barriers (Lucyk et al., 2017), and coder characteristics have been examined as

determinants of validity (Hennessy et al., 2010). In one audit, less experienced coders made fewer but more serious errors, suggesting that experience may affect the type of error (Farzandipour et al., 2010). Payment rules can also shape coding, a point made explicitly in the Thai national audit (Chongthawonsatid, 2017). The encouraging implication is that these mechanisms are modifiable. A study of a specialty-specific electronic health record found that the documentation system influenced ICD-10 coding (Kortüm et al., 2016). Nigerian general hospitals reported sustained improvements in coding accuracy after introducing an electronic diagnostic terminology tool alongside coder training (Olagundoye et al., 2021).

For health information management practice, these findings argue for treating coding quality as a shared responsibility of coders, clinicians, and documentation staff. Regular blinded audits, feedback to clinicians, minimum standards for discharge summaries, and structured clinician–coder communication are consistent with the clinician–coder collaboration recommended in two UK audits (Heywood et al., 2016; Nouraei et al., 2016). In hospitals that rely on dictation and transcription, transcription staff can contribute by flagging unexplained abbreviations and ambiguous diagnostic statements before reports reach coders, and coders can reciprocate by reporting the diagnostic terms that most often cause difficulty.

3.2.1 Strengths and limitations

This review has limitations that readers should weigh. The reconstructed database search used PubMed and was supplemented by known reports and targeted citation checks. The protocol was not registered, and study selection did not use independent duplicate screening. Data available only in published abstracts were limited to what could be verified there. Definitions and denominators differed across studies, including three- versus four-character agreement and records versus codes versus hospital spells. One national audit sampled records already flagged as abnormal. The restriction to English-language reports may have excluded relevant audits.

4. CONCLUSION

Across 15 studies, the accuracy of ICD-10 principal diagnosis coding varied by setting and by how agreement was measured. Differences in audit methods precluded a reliable pooled estimate. Incomplete or ambiguous documentation and coder-related errors recurred across audits. Hospitals should standardize discharge summaries, including dictated and transcribed reports, and use blinded audits to provide feedback to clinicians, transcription staff, and coders. Future audits should use consistent definitions and denominators and measure documentation quality prospectively.

FUNDING

No external funding was received for this work.

ACKNOWLEDGEMENTS

None.

AUTHOR CONTRIBUTIONS

All authors equally contributed to this work (Conceptualization, Methodology, Data Curation, Formal Analysis, Writing – Original Draft, Writing – Review & Editing, and Supervision).

CONFLICT OF INTEREST

The authors declare no conflict of interest.

Ethics Statement

This study is a systematic review of previously published, de-identified data and did not involve new human or animal subjects. Formal ethics approval was therefore not required.

Data Availability Statement

The data extraction tables supporting this review are available from the corresponding author upon reasonable request. The included reports are identified in the references.

REFERENCES

[1] Burns, E.M., Rigby, E., Mamidanna, R., Bottle, A., Aylin, P., Ziprin, P., Faiz, O.D., 2012. Systematic review of discharge coding accuracy. Journal of Public Health 34 (1), 138–148. https://doi.org/10.1093/pubmed/fdr054

[2] Callen, J.L., Alderton, M., McIntosh, J., 2008. Evaluation of electronic discharge summaries: a comparison of documentation in electronic and handwritten discharge summaries. International Journal of Medical Informatics 77, 613–620. https://doi.org/10.1016/j.ijmedinf.2007.12.002

[3] Campbell, S.E., Campbell, M.K., Grimshaw, J.M., Walker, A.E., 2001. A systematic review of discharge coding accuracy. Journal of Public Health Medicine 23 (3), 205–211.

[4] Cheng, P., Gilchrist, A., Robinson, K.M., Paul, L., 2009. The risk and consequences of clinical miscoding due to inadequate medical documentation: a case study of the impact on health services funding. Health Information Management Journal 38 (1), 35–46. https://doi.org/10.1177/183335830903800105

[5] Chin, N., Perera, P., Roberts, A., Nagappan, R., 2013. Review of medical discharge summaries and medical documentation in a metropolitan hospital: impact on diagnostic-related groups and weighted inlier equivalent separation. Internal Medicine Journal 43, 767–771. https://doi.org/10.1111/imj.12084

[6] Chongthawonsatid, S., 2017. Validity of principal diagnoses in discharge summaries and ICD-10 coding assessments based on national health data of Thailand. Healthcare Informatics Research 23 (4), 293–303. https://doi.org/10.4258/hir.2017.23.4.293

[7] Daniels, A., Muloiwa, R., Myer, L., Buys, H., 2021. Examining the reliability of ICD-10 discharge coding in Red Cross War Memorial Children's Hospital administrative database. South African Medical Journal 111 (2), 137–142. https://doi.org/10.7196/SAMJ.2021.v111i2.14686

[8] Farzandipour, M., Sheikhtaheri, A., Sadoughi, F., 2010. Effective factors on accuracy of principal diagnosis coding based on International Classification of Diseases, the 10th revision (ICD-10). International Journal of Information Management 30 (1), 78–84. https://doi.org/10.1016/j.ijinfomgt.2009.07.002

[9] Guyatt, G.H., Oxman, A.D., Vist, G.E., Kunz, R., Falck-Ytter, Y., Alonso-Coello, P., Schünemann, H.J., 2008. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. BMJ 336 (7650), 924–926.

[10] Haliasos, N., Rezajooi, K., O'Neill, K.S., Van Dellen, J., Hudovsky, A., Nouraei, S.A.R., 2010. Financial and clinical governance implications of clinical coding accuracy in neurosurgery: a multidisciplinary audit. British Journal of Neurosurgery 24, 191–195.

[11] Hennessy, D.A., Quan, H., Faris, P.D., Beck, C.A., 2010. Do coder characteristics influence validity of ICD-10 hospital discharge data? BMC Health Services Research 10, 99.

[12] Heywood, N.A., Gill, M.D., Charlwood, N., Brindle, R., Kirwan, C.C., 2016. Improving accuracy of clinical coding in surgery: collaboration is key. Journal of Surgical Research 204 (2), 490–495.

[13] Hong, T.T., Walker, S.M., McKenzie, K., 2009. The quality of injury data from hospital records in Vietnam. Health Information Management Journal 38 (1), 15–21. https://doi.org/10.1177/183335830903800103

[14] Jameson, S., Reed, M.R., 2007. Payment by results and coding practice in the National Health Service: the importance for orthopaedic surgeons. The Journal of Bone and Joint Surgery. British Volume 89, 1427–1430. https://doi.org/10.1302/0301-620X.89B11.19609

[15] Khorrami, F., Alipour, J., Karami, N.A., Hayavi-Haghighi, M.H., Chahooei, M.K., 2022. Quality of documentation of medical records and coding accuracy of ICD-10 versus ICD-11. Journal of Health Administration 25 (3), 150–170. https://doi.org/10.22034/25.3.150

[16] Kortüm, K., Hirneiß, C., Müller, M., Babenko, A., Kampik, A., Kreutzer, T.C., 2016. The influence of a specific ophthalmological electronic health record on ICD-10 coding. BMC Medical Informatics and Decision Making 16, 100. https://doi.org/10.1186/s12911-016-0340-1

[17] Kripalani, S., LeFevre, F., Phillips, C.O., Williams, M.V., Basaviah, P., Baker, D.W., 2007. Deficits in communication and information transfer between hospital-based and primary care physicians: implications for patient safety and continuity of care. JAMA 297, 831–841. https://doi.org/10.1001/jama.297.8.831

[18] Legault, K., Ostro, J., Khalid, Z., Wasi, P., You, J.J., 2012. Quality of discharge summaries prepared by first year internal medicine residents. BMC Medical Education 12, 77. https://doi.org/10.1186/1472-6920-12-77

[19] Lucyk, K., Tang, K., Quan, H., 2017. Barriers to data quality resulting from the process of coding health information to administrative data: a qualitative study. BMC Health Services Research 17, 766. https://doi.org/10.1186/s12913-017-2697-y

[20] Munn, Z., Moola, S., Lisy, K., Riitano, D., Tufanaru, C., 2015. Methodological guidance for systematic reviews of observational epidemiological studies reporting prevalence and cumulative incidence data. International Journal of Evidence-Based Healthcare 13 (3), 147–153.

[21] Naran, S., Hudovsky, A., Antscherl, J., Howells, S., Nouraei, S.A.R., 2014. Audit of accuracy of clinical coding in oral surgery. British Journal of Oral and Maxillofacial Surgery 52, 735–739.

[22] Nouraei, S.A.R., O'Hanlon, S., Butler, C.R., Hadovsky, A., Donald, E., Benjamin, E., Sandhu, G.S., 2009. A multidisciplinary audit of clinical coding accuracy in otolaryngology: financial, managerial and clinical governance considerations under payment-by-results. Clinical Otolaryngology 34 (1), 43–51. https://doi.org/10.1111/j.1749-4486.2008.01863.x

[23] Nouraei, S.A.R., Hudovsky, A., Frampton, A.E., Mufti, U., White, N.B., Wathen, C.G., Sandhu, G.S., Darzi, A., 2015. A study of clinical coding accuracy in surgery: implications for the use of administrative big data for outcomes management. Annals of Surgery 261 (6), 1096–1107. https://doi.org/10.1097/SLA.0000000000000851

[24] Nouraei, S.A.R., Virk, J.S., Hudovsky, A., Wathen, C., Darzi, A., Parsons, D., 2016. Accuracy of clinician-clinical coder information handover following acute medical admissions: implication for using administrative datasets in clinical outcomes management. Journal of Public Health 38 (2), 352–362.

[25] Olagundoye, O., van Boven, K., Daramola, O., Njoku, K., Omosun, A., 2021. Improving the accuracy of ICD-10 coding of morbidity/mortality data through the introduction of an electronic diagnostic terminology tool at the general hospitals in Lagos, Nigeria. BMJ Open Quality 10 (1), e000938. https://doi.org/10.1136/bmjoq-2020-000938

[26] O'Malley, K.J., Cook, K.F., Price, M.D., Wildes, K.R., Hurdle, J.F., Ashton, C.M., 2005. Measuring diagnoses: ICD code accuracy. Health Services Research 40 (5 Pt 2), 1620–1639. https://doi.org/10.1111/j.1475-6773.2005.00444.x

[27] Oweghoro, B.M., Oluwole, E.T., 2022. A review of clinical coding accuracy at University College Hospital, Ibadan, Nigeria. International Journal of Health Records & Information Management 5 (1), 10–14.

[28] Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., Shamseer, L., Tetzlaff, J.M., Akl, E.A., Brennan, S.E., Chou, R., Glanville, J., Grimshaw, J.M., Hróbjartsson, A., Lalu, M.M., Li, T., Loder, E.W., Mayo-Wilson, E., McDonald, S., McGuinness, L.A., Stewart, L.A., Thomas, J., Tricco, A.C., Welch, V.A., Whiting, P., Moher, D., 2021. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372, n71. https://doi.org/10.1136/bmj.n71

[29] Pongpirul, K., Walker, D.G., Rahman, H., Robinson, C., 2011. DRG coding practice: a nationwide hospital survey in Thailand. BMC Health Services Research 11, 290. https://doi.org/10.1186/1472-6963-11-290

[30] Quan, H., Li, B., Saunders, L.D., Parsons, G.A., Nilsson, C.I., Alibhai, A., Ghali, W.A., 2008. Assessing validity of ICD-9-CM and ICD-10 administrative data in recording clinical conditions in a unique dually coded database. Health Services Research 43 (4), 1424–1441. https://doi.org/10.1111/j.1475-6773.2007.00822.x

[31] Quan, H., Moskal, L., Forster, A.J., Brien, S., Walker, R., Romano, P.S., Sundararajan, V., Burnand, B., Henriksson, G., Steinum, O., Droesler, S., Pincus, H.A., Ghali, W.A., 2014. International variation in the definition of 'main condition' in ICD-coded health data. International Journal for Quality in Health Care 26 (5), 511–515. https://doi.org/10.1093/intqhc/mzu064

[32] Saizan, S., Jaudin, R., Yaacob, N.M., Sukeri, S., 2021. The MalaysianDRG casemix system: financial implications of inaccurate clinical documentation and coding error. Malaysian Journal of Medicine and Health Sciences 17 (1), 83–87.

[33] Stausberg, J., Lehmann, N., Kaczmarek, D., Stein, M., 2008. Reliability of diagnoses coding with ICD-10. International Journal of Medical Informatics 77 (1), 50–57. https://doi.org/10.1016/j.ijmedinf.2006.11.005

[34] Tsopra, R., Peckham, D., Beirne, P., Rodger, K., Callister, M., White, H., Jais, J.P., Ghosh, D., Whitaker, P., Clifton, I.J., Wyatt, J.C., 2018. The impact of three discharge coding methods on the accuracy of diagnostic coding and hospital reimbursement for inpatient medical care. International Journal of Medical Informatics 115, 35–42. https://doi.org/10.1016/j.ijmedinf.2018.03.015

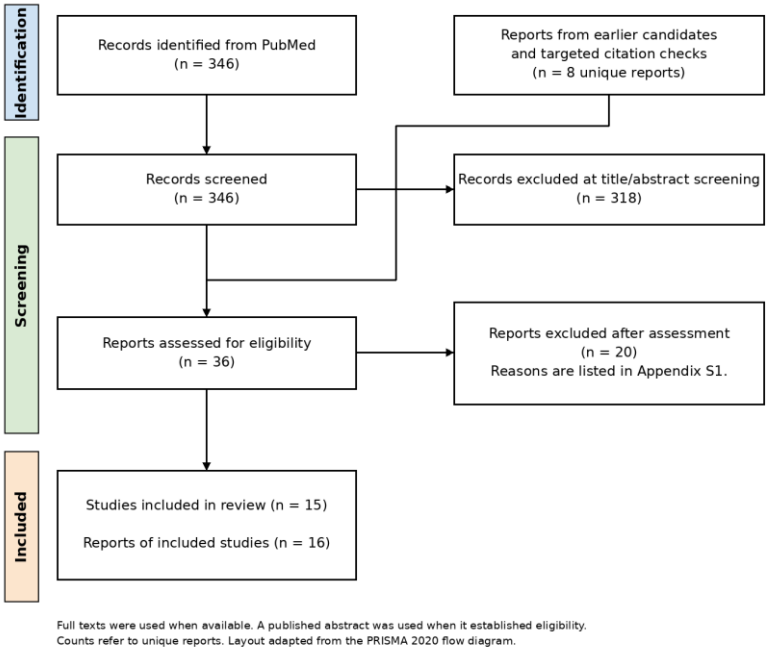
[35] World Health Organization, 2016. International statistical classification of diseases and related health problems, 10th revision, fifth ed. World Health Organization, Geneva.

[36] Zafirah, S.A., Nur, A.M., Puteh, S.E.W., Aljunid, S.M., 2018. Potential loss of revenue due to errors in clinical coding during the implementation of the Malaysia diagnosis related group (MY-DRG®) Casemix system in a teaching hospital in Malaysia. BMC Health Services Research 18, 38. https://doi.org/10.1186/s12913-018-2843-1

[37] Zarei, J., Golpira, R., Hashemi, N., Azadmanjir, Z., Meidani, Z., Vahedi, A., Bakhshandeh, H., Fakharian, E., Sheikhtaheri, A., 2023. Comparison of the accuracy of inpatient morbidity coding with ICD-11 and ICD-10. Health Information Management Journal 18333583231185355. https://doi.org/10.1177/18333583231185355

Figure legend

Figure 1. Study selection flow for the reconstructed search.



Appendix S1. Search and study selection

Retrospective PubMed search run in November 2025 through the NCBI E-utilities ESearch interface. The date and language limits are embedded in the exact query below. The published reports were assessed through PubMed, publishers, journal websites, and repositories as applicable.

("clinical coding"[tiab] OR "discharge coding"[tiab] OR "diagnostic coding"[tiab] OR "diagnosis coding"[tiab] OR "ICD-10 coding"[tiab] OR "ICD-10 discharge"[tiab]) AND (accura*[tiab] OR error*[tiab] OR audit*[tiab] OR agreement[tiab] OR

reliability[tiab]) AND (inpatient*[tiab] OR hospital*[tiab] OR discharge*[tiab]) AND english[la] AND 2009/01/01:2025/12/08[dp]

PubMed returned 346 unique PMIDs. 318 records were excluded at title and abstract screening. The earlier candidate set and targeted checking of citations or related publications contributed 8 distinct reports after matching against the PubMed results, using PMID, DOI, or normalized article title. The list below records all reports assessed for eligibility.

PubMed reports included

PMID 19293434 — Cheng et al.. Included.

PMID 25470740 — Nouraei et al. (surgery). Included.

PMID 25907271 — Nouraei et al. (acute medicine). Included.

PMID 27565087 — Heywood et al.. Included.

PMID 29181239 — Chongthawonsatid. Included.

PMID 29370785 — Zafirah et al. (2018). Included.

PMID 29779718 — Tsopra et al.. Included.

PMID 33944724 — Daniels et al.. Included.

PMID 37491819 — Zarei et al.. Included.

PMID 19293432 — Hong et al.. Included.

PMID 20210533 — Haliasos et al.. Included.

Additional reports included

10.1016/j.ijinfomgt.2009.07.002 — Farzandipour et al. (2010). Included.

Oweghoro and Oluwole, 2022 — Oweghoro and Oluwole. Included.

10.22034/25.3.150 — Khorrami et al.. Included.

Saizan et al., 2021, 17(1):83–87 — Saizan et al.. Included.

Zafirah et al., 2017, 17(2):19–28 — Zafirah et al. (2017), companion to the 2018 report. Included.

Reports excluded after eligibility assessment

PMID 39184245 — Impact of Inaccurate Clinical Coding on Financial Outcome: A Study in a local hospital in Najran, Saudi Arabia. Mixed settings including emergency and clinic records. Separate inpatient results unavailable.

PMID 40523856 — Exploring discrepancies in clinical coding between rural and urban hospitals in Aotearoa New Zealand in patients who underwent interhospital transfer. Compared codes from different hospital encounters without an independent reference standard.

PMID 38994277 — Adoption of network and plan-do-check-action in the international classification of disease 10 coding. Compared quality-management periods. An independent case-note standard for diagnosis coding could not be verified.

PMID 38689701 — Targeted education for clinicians and clinical coding staff improves the accuracy of clinical coding: A quality improvement project. Uncontrolled before-and-after quality-improvement study.

PMID 33674344 — Improving the accuracy of ICD-10 coding of morbidity/mortality data through the introduction of an electronic diagnostic terminology tool at the general hospitals in Lagos, Nigeria. Uncontrolled quality-improvement study using an electronic terminology tool.

PMID 33133832 — Clinical Coding Audit: No coding - No Income - No Hospital. Quality-improvement intervention assessed by tariff changes without independent comparison of diagnosis codes.

PMID 32198235 — Improving the quality of clinical coding and payments through student doctor-coder collaboration in a tertiary haematology department. Single-disease quality-improvement study and no independent diagnosis accuracy comparison.

PMID 31445263 — The impact of improving the quality of coding in the utilities of Diagnosis Related Groups system in a private healthcare institution. 14-year experience. Coding workflow changes. An independent diagnosis re-coding reference standard could not be verified.

PMID 31098532 — Clinician-directed improvement in the accuracy of hospital clinical coding. Reported aggregate code changes and tariffs. Diagnosis-specific coding accuracy could not be isolated.

PMID 30105889 — Level of accuracy of diagnoses recorded in discharge summaries: A cohort study in three respiratory wards. Compared diagnoses documented in summaries with a reference diagnosis list. The companion coding-method study assessed coder-assigned codes.

PMID 29673266 — Level of agreement between medical record and ICD-10-AM coding of mental health, alcohol and drug conditions in trauma patients. Validation of selected mental health, alcohol and drug comorbidity indicators.

PMID 27396219 — Accuracy of ICD-10 Coding System for Identifying Comorbidities and Infectious Conditions Using Data from a Thai University Hospital Administrative Database. Condition-specific validation of selected comorbidity and infection codes.

PMID 21849089 — Improved accuracy of co-morbidity coding over time after the introduction of ICD-10 administrative data. Condition-specific validation of predefined Charlson and Elixhauser indicators.

PMID 19260884 — A multidisciplinary audit of clinical coding accuracy in otolaryngology: financial, managerial and clinical governance considerations under payment-by-results. Combined inpatient and day-surgery cases without separable inpatient outcomes.

PMID 19471647 — Evaluation of factors influencing accuracy of principal procedure coding based on ICD-9-CM: an Iranian study. Procedure coding only using ICD-9-CM.

PMID 31395460 — Quality of coding within clinical datasets: A case-study using burn-related hospitalizations. Compared routine coding distributions without independent record review.

PMID 37891300 — Are we Getting It Right First Time? An exploration of clinical coding practices within a UK paediatric dental hospital department. Procedure coding only.

Farzandipour and Sheikhtaheri, 2009 archive report. Full report in Persian with an English abstract. The English-language report of the Kashan sample was included.

Naran et al., 2014, 52:735–739. Day-case oral surgery population only.

The application of quality control circle in improving the accuracy of ICD coding on medical records (PMC12121122). Uncontrolled before-and-after quality-improvement study.