JOURNAL MEDICAL INFORMATICS TECHNOLOGY
medinftech.org ISSN 2988-7003 (online)

Accuracy of Clinical Coding in Cesarean Section Cases Using ICD-10 and ICD-9-CM

Gama Bagus Kuntoadi[1], Indah Kristina[2], Hudiyati Agustini[3], and Rumondang Christin[4]*

[1,4] *STIKes Widya Dharma Husada Tangerang, Indonesia*
[2,3] *APIKES Bhumi Husada Jakarta, Indonesia*

ARTICLE HISTORY
Received: 23 May 26
Final Revision: 16 June 26
Accepted: 19 June 26
Online Publication: 30 June 26

KEYWORDS
Cesarean Section, Coding Accuracy, ICD-10, ICD-9-CM, Medical Records

CORRESPONDING AUTHOR
gamabaguskuntoadi@wdh.ac.id

DOI
10.37034/medinftech.v4i2.146

A B S T R A C T

Accurate clinical coding in cesarean section cases is essential for morbidity reporting, clinical data quality, and the validity of INA-CBG reimbursement claims under the national health insurance system. Obstetric cases are inherently complex, involving multiple diagnosis and procedure codes that must conform to ICD-10 and ICD-9-CM standards. Objective: This study aimed to analyze the accuracy of ICD-10 diagnosis coding and ICD-9-CM procedure coding across multiple obstetric coding components in cesarean section cases at a hospital in Jakarta, Indonesia. A descriptive quantitative study with a retrospective design was conducted on 95 medical records of obstetric and gynecological patients who underwent cesarean section in 2025, selected through total sampling. Coding accuracy was determined by comparing hospital-assigned codes against standard codes re-verified by researchers based on WHO ICD-10 Volume 2 and ICD-9-CM guidelines. Data were analyzed using frequency, percentage, mean, median, and standard deviation, following a Shapiro-Wilk normality test. The overall accuracy rate for ICD-10 diagnosis codes was 83.9% (266/317 codes) and for ICD-9-CM procedure codes was 91.5% (86/94 codes). Accuracy varied substantially across coding components: secondary diagnosis codes for cesarean delivery and delivery outcome codes each reached 90.5%, while codes for concomitant diagnoses reached only 34.7%. For procedures, cesarean section codes (74.x) achieved 85.3% accuracy, whereas non-cesarean section procedure codes showed accuracy of only 5.3%, predominantly due to undercoding of additional operative procedures documented in surgical reports. Coding accuracy for obstetric cesarean section cases was generally good to very good in aggregate; however, substantial gaps remain in the coding of concomitant diagnoses and additional procedures. This study contributes a component-level evaluation across diagnosis, delivery outcome, and procedure coding, offering a more granular assessment of coding accuracy than prior single-metric studies.

1. Introduction

Clinical coding accuracy is a foundational requirement for modern hospital information systems. Diagnosis and procedure codes generated from medical records serve as the primary basis for morbidity statistics, hospital performance evaluation, clinical audit, and critically in the Indonesian context the determination of reimbursement tariffs under the INA-CBG (Indonesia Case Base Groups) system within the Jaminan Kesehatan Nasional (JKN) framework. A single miscoded diagnosis in an obstetric record can shift a case into an entirely different severity group, directly affecting claim value, hospital revenue, and the integrity of national health data.

Cesarean section cases present a particular challenge for clinical coding staff. The World Health Organization reported that cesarean deliveries now account for more than 21% of all births globally, a figure projected to reach 29% by 2030 [1]. In Indonesia, data from the 2023 Survei Kesehatan Indonesia recorded a national cesarean section rate of 18.4%, with substantially higher rates in urban hospitals and private facilities [2]. This rising volume is not merely an administrative burden, it reflects a growing pool of clinically complex cases. Each cesarean section record typically requires the simultaneous coding of a primary obstetric condition, one or more comorbid diagnoses, the mode of delivery, the delivery outcome, and any additional operative procedures performed. This multi-code structure, governed by specific rules in ICD-10 Volume 2 Chapter XV and ICD-9-CM, leaves considerable room for error at each coding step.

The practical consequences of coding inaccuracies in cesarean section cases are measurable and documented. In the INA-CBG system, diagnostic and procedural codes directly determine case grouping and severity levels; undercoding of comorbidities or additional procedures can result in cases being assigned to lower-rate groups than clinically justified. Research at RSUD Pandan Arang, Boyolali, found that coding inaccuracies caused INA-CBG claims to be placed on hold, with delayed funding totaling Rp286,930,600 in a single quarter [3]. At RSI Sultan Agung Semarang, coding-related claim returns in the first quarter of 2024 represented 25 of 282 total returned files, disrupting hospital cash flow and increasing administrative workload [4]. Beyond the financial dimension, inaccurate obstetric codes distort national morbidity data affecting maternal health statistics, hospital benchmarking, and evidence-based policy formulation at the regional and national level.

Despite this recognized risk, the state of obstetric coding accuracy across Indonesian hospitals remains inconsistent and inadequately characterized at the national level. Available evidence suggests wide variability. A study at RSUD Sleman, Yogyakarta, found an overall diagnostic coding accuracy of 55% in cesarean section cases [5]. Research at RSU Premagana, Bali, reported accuracy rates ranging from 45.98% to 93.2% across different obstetric coding components [6]. A study at Jakarta Port Hospital found delivery outcome coding accuracy of 0%, despite 90% accuracy for maternal complication codes [7]. These figures, taken together, indicate that coding quality is highly uneven across institutions and across coding components within the same institution yet no study has systematically mapped this variability across multiple components within a single institutional setting using both ICD-10 and ICD-9-CM simultaneously.

Previous studies in this area share a common limitation: most evaluate coding accuracy as a single aggregate metric, without decomposing performance across the distinct coding components that constitute an obstetric record. Studies by Anggraini et al. [7], Yunawati [6], and Hidayah & Yunengsih [8] provide valuable institution-specific data but assess only overall accuracy rates or focus on a single code type (primary diagnosis or delivery code). None simultaneously evaluates primary diagnosis codes, concomitant diagnosis codes, delivery outcome codes, cesarean section procedure codes, and non-cesarean section procedure codes within the same sample. This gap is significant because, as the evidence above suggests, aggregate accuracy rates can mask critical weaknesses in specific components particularly in comorbidity coding and additional procedure coding which carry disproportionate weight in INA-CBG grouping and severity assignment.

This study addresses that gap directly. By analyzing 95 medical records of obstetric and gynecological cesarean section cases at a hospital in Jakarta, Indonesia, and evaluating coding accuracy across six distinct coding components using both ICD-10 and ICD-9-CM standards, this study provides a component-level diagnostic of coding quality that prior single-metric studies have not offered. The findings are intended to identify specific weak points in coding practice, support targeted competency development for obstetric coders, and contribute evidence for the design of coding audit systems in Indonesian hospitals.

Based on this background, this study aims to analyze the accuracy of ICD-10 diagnosis coding and ICD-9-CM procedure coding in obstetrics and gynecology cases involving cesarean sections. This study also aims to identify variations in accuracy levels across each coding component and to describe coding areas that remain prone to errors or undercoding in hospital medical record practices.

2. Research Method

This study employed a quantitative descriptive design with a retrospective approach. A descriptive design was chosen because the primary objective is to characterize the accuracy of coding practices as they currently exist, without experimental manipulation or intervention. The retrospective approach was appropriate given that the data source, inpatient medical records had already been generated during routine clinical care prior to the research period. This design is consistent with established methodological frameworks for evaluating clinical coding quality in health information management research.

The study was conducted at ABC Hospital Jakarta, a type C private hospital providing obstetric and gynecological services under the JKN scheme. Data collection was carried out from February to April 2025, covering inpatient medical records of cesarean section cases treated during the 2025 fiscal year. Access to medical records was granted following institutional approval and ethical clearance from the hospital's research ethics committee.

The study population comprised all inpatient medical records of obstetric and gynecological patients who underwent cesarean section at ABC Hospital Jakarta during the data collection period. Total sampling was applied, meaning all records meeting the inclusion criteria were included without further random selection, given that the total eligible population was accessible and countable.

Inclusion criteria: medical records were eligible if they belonged to inpatients who underwent cesarean section during the study period, contained a complete medical summary (resume medis), surgical report (laporan operasi), and nursing discharge notes, and had ICD-10 diagnosis codes and ICD-9-CM procedure codes assigned by the hospital coding staff.

Exclusion criteria: medical records were excluded if the surgical report was missing or illegible, preventing verification of the procedure code; if the medical summary was incomplete and failed to document the primary reason for hospital admission; or if the record was undergoing an active clinical audit or administrative dispute at the time of data collection.

Following application of these criteria, 95 medical records were included in the final analysis.

For the purposes of this study, a diagnosis or procedure code was classified as accurate if the code assigned by the hospital coder was identical to the standard code independently determined by the research team based on ICD-10 and ICD-9-CM guidelines, including correct selection of the principal condition, appropriate use of additional characters (4th and 5th characters), and compliance with ICD-10 Volume 2 coding rules. A code was classified as inaccurate if it differed from the standard code in any of the following respects: incorrect category selection, insufficient specificity (use of a non-specific code where a more specific code was available and documented), incorrect chapter assignment (e.g., use of Chapter XI codes for conditions requiring Chapter XV codes in obstetric patients), or complete omission of a code for a condition or procedure documented in the medical record (undercoding). Partial accuracy was not applied as a separate category; each individual code was assessed as either accurate or inaccurate to maintain objectivity and comparability with previous studies.

The dependent variable in this study was coding accuracy, measured across six coding components, while no independent variables were analyzed because the study employed a purely descriptive design:

1. Primary diagnosis code (ICD-10): the code assigned for the main condition that led to the decision to perform a cesarean section, coded from ICD-10 Chapter XV (O00–O9A).
2. Secondary diagnosis code for mode of delivery (ICD-10): the code specifying the type of cesarean delivery (e.g., O82, O82.1), required in all cesarean section records.
3. Concomitant/other secondary diagnosis codes (ICD-10): codes for comorbid conditions, complications, or other conditions affecting clinical management, coded according to ICD-10 Chapter XV priority rules and supplementary chapter instructions.
4. Delivery outcome code (ICD-10): the code from category Z37 documenting the outcome of delivery (e.g., Z37.0 single live birth).
5. Cesarean section procedure code (ICD-9-CM): codes from category 74.x specifying the type of cesarean procedure performed.
6. Non-cesarean section procedure codes (ICD-9-CM): codes for additional operative procedures performed concurrently with the cesarean section (e.g., adhesiolysis, blood transfusion, catheter placement), as documented in the surgical report.

Code re-verification was conducted by the principal researcher, who holds a Bachelor of Dental Medicine (S.KG.) and a Master of Hospital Administration (MARS.), with documented competency in ICD-10 and ICD-9-CM clinical coding and active involvement in medical record research and health information management education. The verification process was performed independently of the hospital's coding department, using only the information available in the medical record at the time of the original coding.

To assess the reliability of the verification process and minimize single assessor bias, a random subsample of 20 medical records (approximately 21% of the total sample) was independently re-coded by a second assessor, a certified medical record practitioner (Perekam Medis) with a minimum of three years of clinical coding experience in an accredited hospital. Inter-rater agreement between the two assessors was calculated using Cohen's Kappa statistic. A Kappa value ≥ 0.80 was set as the threshold for acceptable reliability, consistent with established benchmarks for coding agreement studies. In the event of disagreement between assessors, the discrepant cases were resolved through structured discussion with reference to the applicable ICD-10 and ICD-9-CM guidelines until consensus was reached.

Code verification was based on the International Statistical Classification of Diseases and Related Health Problems, Tenth Revision (ICD-10), Volumes

1–3 (WHO, 2016 edition) and the International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM), which serve as the reference standards for hospital coding practice in Indonesia under the regulations of the Ministry of Health. These editions represent the versions currently mandated for use in INA-CBG grouping in Indonesia.

Data collection followed a sequential procedure as follows:

1. Medical record identification: A list of all inpatient medical record numbers for cesarean section cases during the study period was obtained from the hospital's medical record information system.
2. Eligibility screening: Each record was reviewed against the inclusion and exclusion criteria. Records failing to meet inclusion criteria or meeting any exclusion criterion were documented and excluded from analysis.
3. Document review: For each eligible record, the following documents were reviewed: medical summary (resume medis), surgical report (laporan operasi), nursing discharge notes, and laboratory results where referenced in the discharge summary.
4. Standard code determination: The principal researcher independently determined the standard codes for each coding component based on the clinical information documented in the medical record, applying ICD-10 and ICD-9-CM guidelines.
5. Accuracy comparison: The hospital-assigned codes were compared against the standard codes determined in step 4. Discrepancies were recorded on the coding checklist instrument.
6. Inter-rater verification: For the 20-record subsample, the second assessor independently completed steps 3–5, and Cohen's Kappa was calculated.
7. Data entry and cleaning: Accuracy findings were entered into a structured spreadsheet. Each record was assigned an accuracy value (accurate or inaccurate) for each of the six coding components.

Two primary sources of potential bias were identified and addressed. First, documentation bias: the accuracy of code verification depends on the quality and completeness of medical record documentation. To control for this, only records with complete core documents (medical summary, surgical report, nursing notes) were included, and an explicit protocol for handling incomplete data was applied through the inclusion and exclusion criteria. Second, assessor interpretation bias: single-assessor code verification introduces the risk that the researcher's clinical knowledge or coding preferences may influence judgments. This was addressed through inter-rater reliability testing on 21% of the sample and through the use of structured coding checklists that constrained assessor judgment to documented information only.

Data were analyzed using descriptive statistical methods. Prior to selecting measures of central tendency, the Shapiro-Wilk normality test was applied to the following continuous numeric variables: (1) number of ICD-10 diagnosis codes per medical record; (2) number of ICD-9-CM procedure codes per medical record; (3) number of accurate diagnosis codes per medical record; (4) number of accurate procedure codes per medical record; (5) diagnosis code accuracy percentage per medical record; and (6) procedure code accuracy percentage per medical record. The Shapiro-Wilk test was selected over the Kolmogorov-Smirnov test given the sample size of $n = 95$, as it demonstrates superior statistical power for detecting non-normality in samples of $n \leq 2{,}000$ [5]. For variables with non-normal distributions, median, mode, and interquartile range (IQR) were reported as primary descriptors; mean and standard deviation were reported additionally for comparative purposes. All statistical analyses were performed using IBM SPSS Statistics Version 26.0 (IBM Corp., Armonk, NY, USA).

The primary data collection instrument was a structured coding checklist (lembar observasi) developed by the research team based on ICD-10 and ICD-9-CM guidelines. The checklist recorded, for each medical record: the codes assigned by the hospital coder for each of the six coding components, the standard codes determined by the researcher, the accuracy classification (accurate or inaccurate or uncodeable), and the type of inaccuracy where applicable (category error, specificity error, chapter assignment error, or omission or undercoding). The checklist was piloted on 10 records prior to main data collection to verify instrument clarity and completeness.

The study was conducted in accordance with the ethical principles governing health research involving medical records. Patient identifiers were removed from all analysis datasets. Medical record data were used exclusively for scientific purposes and were not disclosed to parties outside the research team. The study received ethical approval from the relevant institutional review board prior to data collection.

3. Result and Discussion

3.1. Inter-rater Reliability

Prior to the main analysis, inter-rater reliability was assessed using Cohen's Kappa on 20 randomly selected medical records independently reviewed by two assessors. The analysis yielded a Cohen's Kappa value of 0.87, indicating excellent agreement between assessors. This result demonstrates that the coding verification process was sufficiently reliable and consistent for use in the main study.

3.2. Research Result

This study analyzed 95 medical records of obstetric and gynecological patients who underwent cesarean section and met all inclusion criteria, obtained through total sampling. Of the 95 records, 59 (62.1%) showed complete agreement between the codes assigned by hospital coding staff and the standard codes established by the research team based on ICD-10 and ICD-9-CM guidelines. The remaining 36 records (37.9%) contained at least one coding discrepancy in one or more coding components. The descriptive characteristics and statistical distribution of coding variables are presented in Tables 1 and 2.

Table 1. Shapiro-Wilk Normality Test Results for Study Variables

Variable	n	Min	Max	W-Statistic	p-value & Decision
Number of ICD-10 diagnosis codes per medical record	95	0	4	0.6126	< 0.001 (Not Normal)
Number of ICD-9-CM procedure codes per medical record	95	0	2	0.5097	< 0.001 (Not Normal)
Number of accurate diagnosis codes per medical record	95	0	4	0.7996	< 0.001 (Not Normal)
Number of accurate procedure codes per medical record	95	0	2	0.4158	< 0.001 (Not Normal)
Diagnosis code accuracy percentage (%)	89*	0	100	0.7112	< 0.001 (Not Normal)
Procedure code accuracy percentage (%)	88**	0	100	0.3239	< 0.001 (Not Normal)

* n=89: 6 of 95 records had zero ICD-10 codes assigned by the hospital coder; these records cannot yield a percentage value and are excluded from this variable only.
** n=88: 7 of 95 records had zero ICD-9-CM procedure codes assigned; excluded from percentage calculation only.
All other variables use the full sample n=95.

Prior to descriptive analysis, the Shapiro-Wilk normality test was applied to six continuous numeric variables derived from the dataset (Table 1). The Shapiro-Wilk test was selected because it is recommended for samples of n ≤ 2,000 and demonstrates superior statistical power over the Kolmogorov-Smirnov test for detecting non-normality in small to medium samples [9]. Two of the six variables, diagnosis code accuracy percentage and procedure code accuracy percentage were calculated from the subset of records that had at least one code assigned; records with zero codes assigned were excluded from these percentage calculations only, as dividing by zero would produce undefined values. This is why the denominator for these two variables differs from the total sample: n=89 for diagnosis accuracy percentage (6 records had zero ICD-10 codes assigned) and n=88 for procedure code accuracy percentage (7 records had zero ICD-9-CM procedure codes assigned). These excluded records are retained in all other analyses.

All six variables showed Shapiro-Wilk W statistics substantially below 1.0, with $p < 0.001$ for all variables, leading to rejection of the null hypothesis of normal distribution at $\alpha = 0.05$. The non-normality of accuracy percentage variables is attributable to a ceiling effect: the distribution is negatively skewed, with a large proportion of records achieving 100% accuracy and a small subset of records exhibiting substantial inaccuracies. Consequently, median and mode are reported as primary measures of central tendency throughout this study, with mean and standard deviation reported additionally for comparative purposes. Any future inferential analyses comparing groups should employ nonparametric statistical tests.

Descriptive analysis was performed on three main groups of variables, namely: (a) the number of codes per medical record, (b) the number of accurate codes per medical record, and (c) the coding accuracy percentage per medical record. The descriptive statistics for diagnosis and procedure coding variables are presented in Table 2.

Table 2. Descriptive Statistics of Diagnosis and Procedure Coding Variables

Variable	n	Min	Max	Mean	Median	Mode	SD	Variance	IQR
A. Number of codes/medical records									
ICD-10 diagnosis codes	95	0	4	3.34	4.00	4.00	1.03	1.06	1.00
ICD-9-CM procedure code	95	0	2	0.99	1.00	1.00	0.37	0.14	0.00
B. Number of accurate codes/medical records									
Accurate diagnosis codes (ICD-10)	95	0	4	2.80	3.00	3.00	1.05	1.10	1.00
Accurate procedure codes (ICD-9-CM)	95	0	2	0.91	1.00	1.00	0.33	0.11	0.00
C. Accuracy percentage/medical records (%)									
% Accuracy of diagnosis codes	89*	0	100	85.21	100.00	100.00	20.94	438.37	25.00

Variable	n	Min	Max	Mean	Median	Mode	SD	Variance	IQR
(ICD-10)									
Procedure Code Accuracy (ICD-9-CM)	88**	0	100	93.75	100.00	100.00	21.19	448.99	0.00

* n=89 (records with ≥1 ICD-10 code assigned).
** n=88 (records with ≥1 ICD-9-CM procedure code assigned). All count variables use n=95.

The mean number of ICD-10 diagnosis codes per record was 3.34 (Median=4.00; Mode=4.00; SD=1.03; IQR=1.00), reflecting that most records contained four diagnosis codes: primary diagnosis, mode of delivery, concomitant diagnoses, and delivery outcome. The IQR of 1.00 indicates low variability in coding volume. For ICD-9-CM procedure codes, the mean was 0.99 (Median=1.00; Mode=1.00; SD=0.37; IQR=0.00), indicating that nearly all records received only one procedure code the cesarean section itself, consistent with the undercoding pattern identified in Table 4.

The mean number of accurate diagnosis codes per record was 2.80 (Median=3.00; Mode=3.00; SD=1.05), and for procedure codes was 0.91 (Median=1.00; Mode=1.00; SD=0.33). The difference between total diagnosis codes (3.34) and accurate diagnosis codes (2.80) yields approximately 0.54 inaccurate codes per record, aggregating to 114 inaccurate codes across 380 total diagnosis codes analyzed. The high variance in accuracy percentage (diagnosis: 438.37; procedure: 448.99) confirms substantial heterogeneity across records in coding quality.

Coding accuracy was evaluated using the following operational definition: a code was classified as accurate when it was identical to the standard code independently determined by the research team using ICD-10 Volume 1, 2, and 3 [10], including correct category selection, sufficient specificity (correct 4th and 5th characters where applicable), compliance with ICD-10 Chapter XV priority rules, and inclusion of all clinically documented conditions.

A code was classified as inaccurate when it differed in any of these respects, including incorrect category, insufficient specificity, chapter assignment error, or omission (undercoding). Standard codes were determined by the principal researcher and verified against a second assessor on a 21% subsample (n=20 records); inter-rater agreement was calculated using Cohen's Kappa prior to main analysis. The distribution of ICD-10 diagnosis code accuracy across coding components is presented in Table 3.

Table 3. Distribution of ICD-10 Diagnosis Code Accuracy Based on Coding Components

Diagnosis Code Component	Total Hospital Codes (n)	Accurate rate (n)	Accurate rate (%)	Inaccurate rate (n)	Inaccurate rate (%)	Accuracy Category
Primary diagnosis code (ICD-10 Chapter XV)	95	61	64.2	34	35.8	Moderate
Secondary diagnosis code for cesarean delivery (cesarean section)	95	86	90.5	9	9.5	Excellent
Concomitant/other secondary diagnosis codes	95	33	34.7	62	65.3	Low
Secondary code for delivery outcome (Z37.x)	95	86	90.5	9	9.5	Excellent
Total ICD-10 diagnosis codes	380*	266	83.9	114	16.1	Good

* Total of 380 codes = 4 components × 95 records. The sum 61+86+33+86 = 266 accurate codes; 34+9+62+9 = 114 inaccurate codes. Accuracy rate: 266/380 × 100 = 70.0% component-weighted; 266 accurate codes of 317 assigned codes = 83.9% of codes actually assigned by hospital coders (53 records had only 3 diagnosis codes, not 4, reducing the denominator from 380 to 317 assigned codes).

When considering all potential coding components (n = 380), the component-weighted accuracy was 70.0%. However, when considering only the 317 diagnosis codes actually assigned by hospital coders, 266 codes (83.9%) were accurate and 51 codes (16.1%) were inaccurat. When evaluated component-by-component across all 95 records per component, the total potential codes are 380 (4 components × 95), of which 266 were accurate (70.0% component-weighted accuracy).

Secondary diagnosis codes for mode of delivery and delivery outcome codes each achieved 90.5% accuracy (86/95). Primary diagnosis codes achieved 64.2% (61/95). Concomitant/other secondary diagnosis codes achieved the lowest accuracy at 34.7% (33/95), with 62 of 95 records (65.3%) containing inaccuracies in this component.

The analysis of ICD-9-CM procedure code accuracy covers two components: the procedure code for cesarean section (code 74.x) and the procedure code for non-cesarean section procedures accompanying the operative procedure. The analysis results are presented in Table 4.

Table 4. Distribution of ICD-9-CM Procedure Code Accuracy Based on Coding Components

Procedure Code Component	Total Hospital Codes (n)	Accurate (n)	Accurate (%)	Inaccurate (n)	Inaccurate (%)	Accuracy Category
SC procedure code (code 74.x)	95	81	85.3	14	14.7	Good
Non-SC procedure code (other procedures)	95	5	5.3	90	94.7	Very Low
Total ICD-9-CM procedure codes†	190†	86	45.3†	104	54.7	-

† Total potential procedure codes = 2 components × 95 records = 190. Of these, 86 were coded accurately and 104 inaccurately. The previously reported total of 94 codes reflected only the codes actually assigned by hospital coders (excluding unassigned slots); this figure is retained for comparison but the component-level analysis uses n=95 per row as the denominator, consistent with Table 3.

‡ The 45.3% figure reflects accuracy across all 190 potential procedure code slots. The 91.5% figure (86/94) reflects accuracy only among codes actually assigned — both figures are reported for transparency.

Cesarean section procedure codes (74.x) achieved 85.3% accuracy (81/95). The 14 inaccurate cases involved insufficient specificity: use of code 74.4 (cesarean section, not otherwise specified) in cases where the surgical report documented a specific incision type warranting code 74.1 (low cervical cesarean section) or 74.99 (other specified).

Non-cesarean section procedure codes achieved only 5.3% accuracy (5/95). However, this figure requires careful interpretation: in 90 of the 95 records, inaccuracy did not result from an incorrect code being assigned, but from the complete absence of codes for additional operative procedures documented in the surgical report including peritoneal adhesiolysis, blood transfusion, and urinary catheter placement. This constitutes systematic undercoding rather than coding error per se.

3.3. Research Discussion

The aggregate accuracy rate for ICD-10 diagnosis codes in this study was 83.9% (of codes assigned) and for ICD-9-CM procedure codes was 91.5% (of codes assigned). According to the Ministry of Health of the Republic of Indonesia standard which defines 'good' coding quality as accuracy exceeding 80% [11], the aggregate diagnosis coding falls in the good category and aggregate procedure coding in the very good category. However, these aggregate figures require careful contextualization.

First, the aggregate accuracy of 83.9% is computed only from codes that hospital coders chose to assign. It excludes the 90 records where coders did not assign any non-cesarean section procedure code despite documented procedures in the surgical report. Therefore, two complementary measures of procedure coding accuracy are reported in this study. The first is assigned-code accuracy (91.5%), which evaluates only the procedure codes that were actually assigned by hospital coders. The second is component-based accuracy (45.3%), which considers all potential coding opportunities, including documented procedures that were not coded. Reporting both measures provides a more comprehensive assessment of coding quality and highlights the impact of undercoding on overall coding performance. If the denominator is expanded to include all coding opportunities (190 potential procedure code slots), the component-weighted procedure code accuracy falls to 45.3%, a figure that more accurately reflects the full scope of the coding quality problem. Reporting only the 91.5% figure risks substantially overstating coding adequacy.

Second, the distribution of accuracy across components is highly uneven. As shown in Table 3 and Table 4, two components, concomitant diagnosis codes (34.7%) and non-cesarean section procedure codes (5.3%), fall far below any acceptable threshold, while two other components achieve excellent accuracy (90.5% each). This pattern is consistent with findings from RSU Premagana Bali, where accuracy across obstetric coding components ranged from 45.98% to 93.2% in the same institution [6], confirming that aggregate metrics can obscure critical component-level weaknesses.

The 34.7% accuracy rate for concomitant/other secondary diagnosis codes represents this study's most significant finding from a clinical data quality perspective. Three distinct error patterns were identified in this component, each with different implications.

The first pattern was chapter assignment errors: assigning codes from ICD-10 Chapter XI (K codes, digestive disorders) or Chapter III (D codes, blood disorders) to conditions that in the context of a pregnant patient must be coded using Chapter XV (O codes) with obstetric-specific modifiers. ICD-10 Volume 2 Section 4.4 specifies that conditions complicating pregnancy should be classified to Chapter XV wherever a code exists, with non-obstetric chapter codes used only when the condition is explicitly stated as not related to or not affecting the pregnancy. Previous studies suggest this error pattern is associated with limited familiarity with the priority coding rules of Chapter XV [12].

The second pattern was insufficient specificity: use of 3-character codes (e.g., O26) where the clinical documentation provided sufficient information to assign a 4th or 5th character (e.g., O26.4x). Research at RSU Aisyiyah found that 54.55% of coding inaccuracies in surgical cases involved incorrect 4th or 5th character selection [13], a finding consistent with the present study's observations.

The third and most consequential pattern was undercoding by omission: failure to list codes for comorbid conditions documented in the medical summary including anemia, hypertension complicating pregnancy, uterine fibroids, and prior uterine surgery, even when these conditions were explicitly documented as influencing clinical management. Previous literature suggests this pattern is linked to incomplete review of the full medical summary rather than the discharge diagnosis list alone [14].

The clinical data quality implications of 34.7% accuracy in this component are substantial. Concomitant diagnoses determine severity level assignment in INA-CBG grouping; their systematic undercoding or miscoding can shift cases to lower severity groups with reduced reimbursement rates. This dynamic has been documented empirically: at RSUD Pandan Arang Boyolali, coding-related claim holds totaled Rp286,930,600 in a single quarter [3], while at RSI Sultan Agung Semarang, 25 of 282 returned BPJS claim files in Q1 2024 were attributable to coding discrepancies [4]. Although the present study did not directly measure financial impact, the prevalence of concomitant diagnosis undercoding in 65.3% of records indicates that similar INA-CBG grouping distortions are a realistic risk at the study institution.

The 5.3% accuracy of non-cesarean section procedure codes is the most striking finding of this study numerically, but it must be interpreted with precision. In 90 of the 95 records where this component was inaccurate, the inaccuracy consisted entirely of coding omission: procedures documented in the surgical report were not coded at all, rather than coded incorrectly. Only 5 records contained an assigned non-cesarean section procedure code, of which all 5 were accurate. This means that the 94.7% 'inaccuracy' rate is entirely attributable to systematic undercoding, not to misidentification or miscategorization of procedures.

The documented but uncoded procedures included peritoneal adhesiolysis (ICD-9-CM 54.59), perioperative blood transfusion (99.03–99.04), urinary catheter placement (57.94), and intraoperative hemostasis procedures. These are clinically meaningful procedures with direct relevance to case complexity, resource utilization, and INA-CBG severity grouping.

This undercoding pattern is consistent with international evidence. A retrospective analysis of 34,104 records at UCLA Medical Center found that 32.9% contained clinical conditions with no corresponding ICD code, and 5.8% of these omissions would have resulted in a higher DRG modifier if coded correctly [14]. In the Indonesian INA-CBG context, procedure undercoding has an analogous impact: additional procedures documented in the surgical report but absent from the coding record cannot be factored into grouping algorithms, resulting in cases being assigned to lower-rate groups than their actual clinical complexity warrants.

Based on the patterns observed and consistent with the literature, the explanation for this systematic undercoding most likely lies in the coding workflow rather than coder incompetence per se. When coders work primarily from the discharge summary or medical resume rather than the full surgical report, additional operative procedures recorded only in the surgical report may not enter the coding workflow. Erlindai and Indriani reported a similar phenomenon at RSU Imelda Pekerja Indonesia Medan, where coders' limited attention to surgical report details was the primary contributing factor to procedure coding inaccuracies [15]. This interpretation should be understood as a hypothesis supported by literature patterns rather than a finding directly measured in this study.

Primary diagnosis code accuracy of 64.2% indicates that more than one-third of records contained errors in the determination of the principal condition — the condition established to be chiefly responsible for the patient's admission and the decision to perform a cesarean section. The dominant error pattern was selection of the first condition mentioned in the discharge summary rather than application of the principal diagnosis selection rules in ICD-10 Volume 2 Section 3, a practice the literature describes as 'coding by proximity' [14]. This is not a finding unique to this study: Albagmi's study at Najran Hospital Saudi Arabia found primary diagnosis inaccuracy in 26.8% of records [16], and Yulida et al. at RSUD Sleman Yogyakarta reported overall diagnostic accuracy of 55% in cesarean section cases [5], both suggesting that principal diagnosis selection is a cross-institutional challenge.

In contrast, secondary diagnosis codes for mode of delivery (90.5%) and delivery outcome codes (90.5%) achieved excellent accuracy. These components involve relatively standardized codes applied consistently in every cesarean section record, O82 for elective cesarean section, O82.1 for repeat cesarean section, and Z37.0 for single live birth. Their high accuracy underscores that coding staff are proficient with routine, protocol-driven coding tasks. The real challenge lies in the components requiring clinical judgment: principal condition selection and concomitant diagnosis identification.

The findings of this study have direct practical implications at three levels: coder competency development, clinical documentation improvement, and institutional audit systems.

At the coder competency level, the specific error patterns identified in this study, chapter assignment errors in concomitant diagnoses, specificity failures in procedure codes, and systematic omission of additional procedures, point to targeted training needs rather than general coding refreshers. Training programs should specifically address: (1) ICD-10 Chapter XV priority rules and the condition for using non-obstetric codes in pregnant patients; (2) 4th and 5th character selection in obstetric and surgical codes; and (3) procedural coding from surgical reports using ICD-9-CM, with practical exercises on common obstetric operative procedures.

At the clinical documentation level, the low accuracy of concomitant diagnosis codes and additional procedure codes is partly attributable to documentation practices. When physicians list diagnoses without specifying their relationship to the pregnancy, or when surgical reports use generic terms that do not describe the specific procedure performed, coders lack the information needed for accurate coding. Structured discharge summary templates that prompt explicit documentation of all conditions affecting management and all procedures performed with space for procedure-specific details could reduce this source of coding inaccuracy.

At the institutional audit level, the component-level analysis in this study provides a template for a targeted coding audit instrument. Rather than computing an aggregate accuracy rate, a coding quality audit for obstetric cesarean section cases should evaluate each of the six coding components separately, flag records with concomitant diagnosis codes from non-Chapter XV chapters, and verify procedure code completeness against surgical reports. Regular internal audits using this framework, combined with feedback to coding staff, represent the most efficient pathway to improving coding quality without requiring wholesale reorganization of coding workflows.

The potential impact on INA-CBG reimbursement accuracy is also direct. Systematic undercoding of concomitant diagnoses and additional procedures results in cases being grouped into lower severity levels than their actual clinical complexity warrants. Correcting this undercoding through the mechanisms described above would not only improve reimbursement accuracy but would also produce morbidity data that more accurately reflects the clinical burden of obstetric care at the institution, with downstream implications for resource planning and service quality evaluation.

This study has four principal limitations that readers should consider when interpreting the findings.

First, the study was conducted at a single hospital in Jakarta, limiting the generalizability of findings to other institutional contexts. Coding practices, documentation quality, and coder competency levels vary substantially across hospital types, regions, and ownership status in Indonesia; the patterns identified here may not be representative of the national picture.

Second, coding accuracy was determined through researcher re-verification rather than by an external credentialed clinical coding auditor. Although inter-rater reliability was assessed on a 21% subsample using Cohen's Kappa, and discrepancies were resolved through structured discussion referencing ICD-10 and ICD-9-CM guidelines, the verification process may still reflect the principal researcher's coding interpretation, which may differ from that of other equally qualified coders in ambiguous cases.

Third, this study did not investigate the factors contributing to coding inaccuracies through direct measurement, for example, through coder competency assessments, structured documentation completeness audits, or coder interviews. All statements regarding contributing factors in the Discussion section are literature-supported interpretations of the patterns observed, not empirically measured causal relationships. Future studies using mixed methods or analytical designs would be better positioned to establish these relationships.

Fourth, the study did not assess the actual financial impact of identified coding inaccuracies on INA-CBG grouping and claim values. While the literature reviewed suggests a plausible financial impact, quantifying this impact for the specific records analyzed would require integration with the hospital's claim management and grouper systems, which was outside the scope of this study.

4. Conclusion

Of the 95 cesarean section medical records analyzed, the overall accuracy rate was 83.9% for assigned ICD-10 diagnosis codes and 91.5% for assigned ICD-9-CM procedure codes. These values meet the Ministry of Health's thresholds for good and very good coding quality, respectively. However, component-level analysis revealed substantial variation that was masked by the aggregate results. Secondary diagnosis codes for mode of delivery and delivery outcome achieved the highest accuracy (90.5% each), followed by cesarean section procedure codes (85.3%). In contrast, primary diagnosis codes achieved 64.2% accuracy, concomitant or other secondary diagnosis codes only 34.7%, and non-cesarean section procedure codes only 5.3%.

The most critical findings are the 34.7% accuracy of concomitant diagnosis codes attributable to chapter assignment errors, specificity failures, and omission of documented comorbidities and the 5.3% accuracy of non-cesarean section procedure codes, which reflects near-complete systematic undercoding of additional operative procedures documented in surgical reports rather than coding misidentification. These two components represent the primary risk areas for INA-

CBG severity misclassification and reimbursement discrepancy.

These findings demonstrate that obstetric coding quality audit frameworks should evaluate coding components separately rather than relying on aggregate accuracy metrics. Aggregate metrics can indicate overall adequacy while obscuring clinically and financially significant failures in specific components. Future studies should investigate the factors contributing to component-level accuracy differences through analytical or mixed-methods approaches that directly assess coder competency, documentation completeness, and workflow characteristics. Future research should also quantify the financial impact of undercoding on INA-CBG claim reimbursement.

References

[1] World Health Organization, "Caesarean section rates continue to rise, amid growing inequalities in access," 2021. [Online]. Available: https://www.who.int/news/item/16-06-2021-caesarean-section-rates-continue-to-rise-amid-growing-inequalities-in-access. [Accessed: 17-May-2026].

[2] Kementrian Kesehatan RI, *Survey Kesehatan Indonesia (SKI) 2023*. Jakarta: Kementrian Kesehatan RI, 2023.

[3] Anton Susanto, Galih Maheswari Candraningtyas, and Agus Trijono, "Akurasi Pengkodean Diagnosis dan Prosedur Medis serta Implikasinya terhadap Klaim JKN di RSUD Pandan Arang," *J. Ilm. Kedokt. dan Kesehat.*, vol. 4, no. 3, pp. 306–317, Jun. 2025, doi: 10.55606/klinik.v4i3.4610.

[4] J. N. Rizki, A. A. Sari, Yektiningtyastuti, and E. Suprihatin, "Penyebab Pengembalian Klaim BPJS Rawat Inap Triwulan 1 Tahun 2024 di RSI Sultan Agung Semarang," *J-REMI J. Rekam Med. dan Inf. Kesehat.*, vol. 6, no. 2, pp. 156–167, Mar. 2025, doi: 10.25047/j-remi.v6i2.5853.

[5] R. Yulida, H. N. Seha, and H. M. Khafi, "Analisis Ketepatan Kode Diagnosa Pada Kasus Persalinan Dengan Sectio Caessarea Di RSUD Sleman Yogyakarta," *J. PERMATA Indones.*, vol. 12, no. 1, May 2021, doi: 10.59737/jpi.v12i1.9.

[6] I. W. G. Arimbawa, N. P. L. Yunawati, and I. A. P. F. Paramita, "Hubungan Kelengkapan Penulisan Diagnosis Terhadap Keakuratan Kode ICD-10 Kasus Obstetri Triwulan III Pasien Rawat Inap Di RSU Premagana," *J. Manaj. Inf. Kesehat. Indones.*, vol. 10, no. 1, pp. 31-35, Mar. 2022, doi: 10.33560/jmiki.v10i1.370..

[7] A. Anggraini, L. Widjaja, L. Indawati, and D. Rosmala Dewi, "Analisis Ketepatan Kode Diagnosis Kasus Persalinan Secara Sectio Caesarea Di Rumah Sakit Pelabuhan Jakarta," *Cerdika J. Ilm. Indones.*, vol. 3, no. 1, pp. 6–11, Jan. 2023, doi: 10.59141/cerdika.v3i1.505.

[8] F. F. Hidayah and Y. Yunengsih, "Analisis Ketepatan Kode Diagnosa Pada Kasus Persalinan Dengan Sectio Caesarea di PKU Muhammadiyah Kutowinangun," *J. Kesehat. Tambusai*, vol. 5, no. 3, 2024.

[9] N. M. Razali and Y. B. Wah, "Power comparisons of Shapiro-Wilk, Kolmogorov-Smirnov, Lilliefors and Anderson-Darling tests," *J. Stat. Model. Anal.*, vol. 2, no. 1, pp. 21–33, 2011.

[10] WHO, *International Classification of Diseases (ICD-10)*. World Health Organization (W.H.O), 2019.

[11] Kementrian Kesehatan RI, *Prosedur Penyelenggaraan & Prosedur Rekam Medis Rumah Sakit*. Jakarta: Kesehatan RI, Departemen, 2006.

[12] C. A. Puspaningtyas, H. S. Sangkot, P. S. Akbar, E. S. Dewi, and A. Wijaya, "Analisis Hubungan Ketepatan Penulisan Diagnosis dengan Keakuratan Kode Diagnosis pada Kasus Obstetri danGinekologi di Rumah Sakit Tk. IV DKT Kediri," *RAMMIK J. Rekam Med. dan Manaj. Inf. Kesehat.*, vol. 1, no. 2, pp. 94-100, 2022, doi: 10.47134/rammik.v1i2.22.

[13] N. F. Khumaira, A. Wahyuni, and S. Siska, "Ketidaktepatan Kode Diagnosis Kasus Bedah RSU 'Aisyiyah berdasarkan ICD-10," *J-REMI J. Rekam Med. dan Inf. Kesehat.*, vol. 5, no. 4, pp. 299–306, Aug. 2024, doi: 10.25047/j-remi.v5i4.5080.

[14] E. Gabel, J. Gal, T. Grogan, and I. Hofer, "A retrospective analysis using comorbidity detecting algorithmic software to determine the incidence of International Classification of Diseases (ICD) code omissions and appropriateness of Diagnosis-Related Group (DRG) code modifiers," *BMC Med. Inform. Decis. Mak.*, vol. 24, no. 1, p. 309, Oct. 2024, doi: 10.1186/s12911-024-02724-8.

[15] E. Purba and A. Indriani, "FAKTOR-FAKTOR YANG MEMPENGARUHI KETIDAKTEPATAN KODE PADA PERSALINAN SECTIO CAESAREA DI RUMAH SAKIT UMUM IMELDA PEKERJA INDONESIA MEDAN TAHUN 2018," *J. Ilm. Perekam dan Inf. Kesehat. Imelda*, vol. 3, no. 2, pp. 453–465, Dec. 2019.

[16] S. Albagmi, "Impact of Inaccurate Clinical Coding on Financial Outcome: A Study in a local hospital in Najran, Saudi Arabia.," *F1000Research*, vol. 13, p. 820, Jul. 2024, doi: 10.12688/f1000research.149154.1.