British Journal of Hospital Medicine

Br. J. Hosp. Med. (Lond) **2026**; 87(3): 52079
https://doi.org/10.31083/BJHM52079

Article

Improvement in Secondary Diagnosis-Related Coding Under DRG Reform: A Retrospective Dual-Center Study in China

Hubing Wu[1,*], Lili Zhao[2], Jianming Chen[1], Wenyi Fan[3]

[1]Medical Record Department, Peking University Third Hospital, 100191 Beijing, China
[2]Performance Office, Beijing Tongren Hospital, CMU, 100730 Beijing, China
[3]Medical Affairs Department, Peking University Third Hospital, 100191 Beijing, China
*Correspondence: wuhubing@126.com (Hubing Wu)
Academic Editor: Chris T Derk
Submitted: 28 September 2025 Revised: 7 December 2025 Accepted: 26 December 2025 Published: 20 March 2026

Abstract

Aims/Background: Accurate and standardized diagnostic coding is essential for hospital performance, reimbursement, and the secondary use of health data for management and research. Despite its significance, how diagnostic coding structure and complexity shift over time, and how these trends differ between institutions, remain poorly defined at the health system level. This study aimed to assess long-term trends and inter-hospital variability in diagnostic coding practice in Chinese public hospitals, focusing on the secondary diagnosis number (SDN), International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) chapter composition, and diagnostic diversity over a 12-year period. **Methods**: The retrospective study analyzed inpatient discharge data from two public tertiary hospitals (Hospital A and Hospital B) at three time points: 2012, 2019, and 2024. Key indicators included the mean SDN per admission, ICD-10 chapter-level diagnosis composition, the shannon diversity index, and heatmap-based profiling of coding structure. Non-parametric tests were applied to assess temporal trends and differences between hospitals. **Results**: Mean SDN increased significantly in both hospitals between 2012 and 2024 (Hospital A: 1.45 vs. 1.62; Hospital B: 1.37 vs. 1.61; $p < 0.01$), suggesting progressively deeper documentation of comorbidities and complications. ICD-10 chapter distributions also changed over time, with a gradual increase in chronic disease-related diagnoses and a substantial reduction in non-specific categories, such as symptoms and abnormal findings ($p < 0.001$). The shannon diversity index showed a modest downward trend, reflecting increasing concentration and standardization in coding patterns. Heatmap analysis further revealed a decrease in inter-hospital variation and a clear convergence toward standardized diagnostic structures over time. **Conclusion**: Over the past decade, diagnostic coding practices in Chinese public hospitals have become more comprehensive, more structured, and increasingly standardized. These trends are consistent with the combined effects of policy reforms, institutional learning, and the digitalization of health information systems. Furthermore, we observed that improvement in coding depth and structure is likely to contribute to greater accuracy and reliability of Diagnosis-Related Group (DRG)-based payment systems.

Keywords: diagnostic coding; DRGs; ICD-10; secondary diagnosis; data accuracy

1. Introduction

Accurate, standardized diagnostic coding is central to modern healthcare systems, which support clear diagnostic documentation and communication, and also drive hospital financing, quality reporting, epidemiologic tracking, and health policy formulation [1]. The International Statistical Classification of Diseases and Related Health Problems (ICD), developed by the World Health Organization (WHO), offers a globally accepted framework for encoding health conditions. The chapter-based structure of this framework systematically organizes diseases and health conditions, broadly aligned with organ systems, underlying etiologies, and clinical contexts [2]. The utility of the ICD coding system, however, depends heavily on how diagnoses are performed in routine clinical practice, including their accuracy, completeness, and level of specificity.

Among the parameters used to evaluate diagnostic data quality, the secondary diagnosis number (SDN), defined as the average number of coexisting conditions documented alongside the primary diagnosis, is commonly used as a valuable indicator of documentation depth, case complexity, and coding behavior [3]. In parallel, the distribution of ICD codes across chapters can reflect an institution's clinical focus, local disease burden, and the overall maturity of its coding practices [4]. These structural indicators are especially crucial under Diagnosis-Related Group (DRG) payment models, where reimbursement depends on diagnoses being coded accurately and justified by clinical documentation [5].

Globally, many health systems shift from fee-for-service to prospective, value-based payment, and the accuracy and consistency of diagnostic coding have become increasingly scrutinized [6]. Within DRG frameworks, hospitals have a clear incentive to document relevant comorbidities and complications because these directly influence case classification, reimbursement levels, and the apparent complexity of care [7]. Conversely, incomplete documentation or use of vague or non-specific codes (for example, ICD Chapter XVIII: symptoms, signs, and abnormal clinical and laboratory findings not elsewhere classified) as the principal diagnosis can result in financial penalties and misclassification [8]. In response, many countries have introduced national initiatives to enhance diagnostic coding quality through comprehensive training for professional coders, software-assisted code selection, and routine or real-time coding auditing [9].

In China, the move toward DRG-based payment reform is a key policy development. After pilot programs initiated in 30 cities in 2019, the reform was expanded nationwide by 2021, aiming to improve hospital cost-efficiency, reduce overtreatment, and standardize diagnostic documentation [10]. To support this reform, the National Healthcare Security Administration (NHSA) has recommended coding guidelines, introduced hospital performance benchmarking strategies, and developed diagnostic grouping catalogs, including the Chinese Healthcare Security Diagnosis-Related Group (CHS-DRG) and Diagnosis-Intervention Packet (DIP) systems, to enhance reimbursement equity and data comparability across regions [11]. These initiatives have effectively shifted diagnostic coding from a background administrative task to a core operational priority. In parallel, national reforms are also assessing broader organizational challenges, such as the National Health Commission Hospital Administration Institute project on the management status and development strategies of ‘one hospital, multiple campuses’ in Chinese public hospitals [Project No. 2024-106 (CON20241153281)], aiming to provide more systematic solutions for hospital governance.

Despite these advances, institution-level evidence on how diagnostic coding has evolved over time remains limited, especially with respect to multi-year trends in documentation depth and structural coding patterns. Much of the existing research has focused on macro-level DRG outcomes or single-year audits of coding accuracy. Consequently, several practical questions remain inadequately addressed: Has coding quality improved over time? Are hospitals moving toward more standardized and specific ICD coding? Do coding patterns vary across hospital groups or policy phases? And is there measurable evidence of structural refinement, meaning a shift away from generalized, non-specific coding towards patterns that are more clinically focused and aligned with policy frameworks?

Addressing these questions requires a more nuanced and granular approach that goes beyond counting diagnoses. It should also examine diagnostic compositional features, how it changes over time, and how they vary across institutions. The current study addresses this gap by assessing diagnostic coding from two major hospital groups at three time points spanning 12 years. Using SDN, ICD chapter distribution, entropy-based diversity indices, and hospital-level diagnostic profiles, the study aims to uncover system-level behavioral patterns shaped by policy reform, institutional capacity, and evolving health information systems. In doing so, the study contributes meaningful evidence relevant to hospital payment reform, diagnostic governance, and health data quality in the digital era.

2. Methods

2.1 Study Design and Setting

This is a retrospective, descriptive study evaluating inpatient discharge front page data from two tertiary hospitals in Beijing, China: Peking University Third Hospital (Hospital A) and Beijing Tongren Hospital, CMU (Hospital B). Because the study analyzed de-identified data and involved no direct patient contact, the ethics committees of the hospitals exempted the need for formal ethical review. Hospital A was classified as an early adopter of DRG reform, whereas Hospital B showed a more gradual response. The study included all inpatient discharges from three calendar years, 2012, 2019, and 2024, selected to represent three distinct phases of DRG policy implementation.

2.2 Data Collection and Cleaning

We extracted all secondary diagnosis (SD) entries from the inpatient front page (IFP) dataset for each hospital and study year. Duplicate records and entries or missing diagnosis fields were excluded. Each SD was then assigned to an International Statistical Classification of Diseases and Related Health Problems, 10th Revision (ICD-10) chapter (I–XXII) based on the first character of its ICD-10 code. Records without any SD were retained and coded with zero SD entries to enable consistent normalization across records. The specific metrics used to characterize coding behavior and structural patterns, including the SDN, coding density (CD), and the secondary diagnosis diversity index (SDDI), are detailed in Table 1 along with their definitions and intended interpretations.

2.3 Statistical Analysis

Statistical analyses were conducted in R (version 4.2.2, R Core Team, Vienna, Austria) and SPSS version 26.0 (IBM Corp., Armonk, NY, USA). Continuous variables, including the secondary diagnoses number (SDN) per admission, were summarized using the mean, median, and standard deviations, and their distributions across study years and hospitals were assessed for normality.

Normality across continuous variables was assessed using the Shapiro–Wilk test (α = 0.05) and by visually reviewing Q–Q plots and histograms. Where appropriate, the homogeneity of variances was assessed using Levene’s test. Given that most variables showed statistically significant deviations from normality or clear skewness on visual assessment, we applied non-parametric approaches for inferential comparisons. Specifically, SDN and other continuous variables (such as coding density) were compared across hospitals and years using the Mann–Whitney U test. Categorical variables were assessed using the chi-square test.

Table 1. Variable definitions used in assessing coding behavior.

Variable	Definition	Purpose
SDN (secondary diagnosis number)	Count of secondary diagnoses per record	Reflects coding density
CD (coding density)	Annual SD total / Annual number of discharges	Measures temporal coding intensity
Chapter proportion	Distribution of SDs by ICD-10 chapters	Tracks structural changes in coding
SDDI (secondary diagnosis diversity index)	Shannon entropy index of chapter distribution	Reflects coding variety
Strategy chapter growth rate	Proportional increase of non-core chapters (XXI, XVIII, V)	Detects strategic behavior

SD, secondary diagnosis; ICD-10, International Statistical Classification of Diseases and Related Health Problems, 10th Revision.

To evaluate structural changes in coding composition, we calculated the proportion of secondary diagnoses assigned to each ICD-10 chapter and used chi-square tests to compare these distributions across different timepoints using the chi-square test. Coding diversity of each year was assessed using the Shannon entropy index, which quantifies both the richness and balance of chapter-level coding; entropy values were plotted to illustrate temporal trends.

To investigate potential behavior suggestive of strategic coding, we developed a chapter contribution model based on the relative increase in diagnosis counts per chapter between 2012 and 2024. Where appropriate, linear trend fitting was applied to model continuous changes over time. All statistical tests were two-tailed, and p-values less than 0.05 were considered statistically significant.

3. Results

3.1 Trends in Secondary Diagnosis Number (SDN) Over Time

We evaluated how the SDN changed over time in two hospitals (Hospital A and Hospital B) across three years: 2012, 2019, and 2024. Descriptive analysis revealed a clear upward shift in mean and median SDN in both hospitals throughout the study period (Table 2). At Hospital A, SDN was considerably higher in both 2019 and 2024 compared to 2012 ($p < 0.01$), but there was no statistically significant difference between 2019 and 2024 ($p = 0.096$, Fig. 1). This finding indicates that the largest increase in SDN at Hospital A occurred between 2012 and 2019.

Hospital B showed a distinct pattern, with significant increases at every adjacent time point. SDN in 2019 was significantly higher than in 2012 ($p < 0.01$), and SDN in 2024 was higher than in 2019 ($p < 0.01$). These results suggest a consistent increase in the recording of secondary diagnoses over the entire observation period.

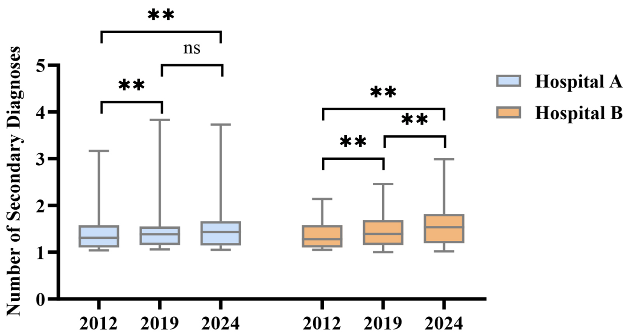

Fig. 1. Comparison of secondary diagnosis counts between two hospitals (Hospital A and Hospital B) across three years: 2012, 2019, and 2024. The box plot illustrates the distribution of SDN across two independent groups (Hospital A and Hospital B) over three distinct time points (2012, 2019, and 2024). ns, not significant; **, $p < 0.01$. Hospital A, Peking University Third Hospital; Hospital B, Beijing Tongren Hospital, CMU.

Table 2. Descriptive analysis of the secondary diagnosis number (SDN) at Hospital A and Hospital B across three years (2012, 2019, and 2024).

Year	Hospital	Average SDN	Median
2012	A	1.45	1.31
	B	1.37	1.28
2019	A	1.55	1.40
	B	1.46	1.39
2024	A	1.62	1.45
	B	1.61	1.52

3.2 ICD Chapter Composition Across Groups and Years

To explore how diagnostic coding structures changed over time, we compared the ICD-10 chapter-level composition at both Hospital A and Hospital B from 2012 to 2024. As illustrated in the stacked bar chart (Fig. 2) and detailed in Table 3, both hospitals demonstrated clear shifts in diagnostic structures. Chi-square tests showed significant differences in ICD chapter distributions between 2012 and 2024 in both Hospital A and Hospital B ($p < 0.001$).

In Hospital A, the most prominent increase occurred in Chapter XV (Pregnancy, childbirth and the puerperium), where the average number of secondary diagnoses per case rose from 2.28 to 3.43 (an increase of 1.15). Substantial increases were also observed in Chapter XVI (certain conditions originating in the perinatal period), Chapter IV (endocrine, nutritional and metabolic diseases), Chapter IX (diseases of the circulatory system), and Chapter III (diseases of the blood and blood-forming organs). Hospital B showed a similar pattern, with Chapter XV indicating the most significant increase (from 2.06 to 3.43, an increase of 1.37). Further increases were found in Chapter XII (diseases of the skin and subcutaneous tissue), Chapter X (dis-

Table 3. Top 5 ICD chapters with the highest increase in the number of secondary diagnoses (ranked by growth value from 2012 to 2024).

ICD chapter	Chapter range	Chapter name	2012	2019	2024
Hospital A					
Chapter XV	O00–O99	Pregnancy, childbirth and the puerperium	2.28	3.21	3.43
Chapter XVI	P00–P96	Certain conditions originating in the perinatal period	3.17	3.83	3.73
Chapter III	D50–D89	Diseases of the blood and blood-forming organs	1.15	1.22	1.52
Chapter IV	E00–E90	Endocrine, nutritional and metabolic diseases	1.30	1.41	1.70
Chapter IX	I00–I99	Diseases of the circulatory system	1.84	1.92	2.21
Hospital B					
Chapter XV	O00–O99	Pregnancy, childbirth and the puerperium	2.06	3.24	3.43
Chapter XII	L00–L99	Diseases of the skin and subcutaneous tissue	1.05	1.43	1.58
Chapter X	J00–J99	Diseases of the respiratory system	1.26	1.33	1.67
Chapter III	D50–D89	Diseases of the blood and blood-forming organs	1.10	1.39	1.46
Chapter XI	K00–K93	Diseases of the digestive system	1.49	1.64	1.94

ICD, International Statistical Classification of Diseases and Related Health Problems.

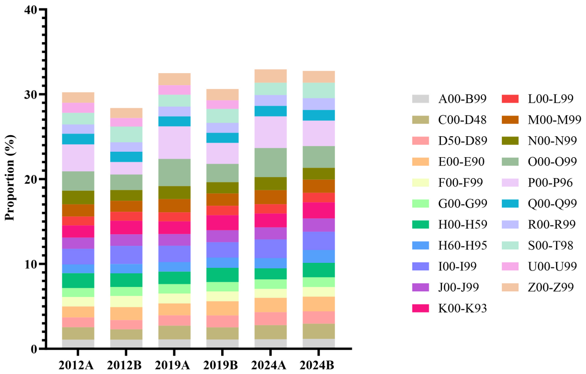

Fig. 2. Stacked bar chart of ICD-10 chapter-level diagnostic composition across hospital groups and years. Each bar represents the distribution of 21 ICD-10 chapters among all documented diagnoses for both Hospital A and Hospital B across the three years (2012, 2019, and 2024). The observable changes in chapter composition over time suggest a shifting diagnostic focus, with greater contribution from chronic disease-related chapters and a declining reliance on non-specific, symptom-based codes.

eases of the respiratory system), and Chapter III (diseases of the blood and blood-forming organs).

Conversely, we found an overall decline in the relative use of less specific codes, particularly those from Chapter XVIII (symptoms, signs and abnormal diagnostic and laboratory findings, not elsewhere classified), which is consistent with enhanced diagnostic precision. This structural refinement suggests that diagnostic coding gradually shifted away from vague symptom-based categories toward more clinically specific diagnoses over time. Such a pattern may indicate progressive alignment of coding practices with policy requirements and improved coding awareness under DRG reform. Longitudinal results for all ICD-10 chapters are given in **Supplementary Table 1**.

It is also noteworthy that chapters outside the core chronic disease categories, such as Chapter XXI (factors influencing health status and contact with healthcare services), remained less frequent overall but demonstrated a consistent upward trend in 2024 compared with 2012 (**Supplementary Table 1**). For example, the average SDN value for Chapter XXI in Hospital A increased from 1.23 to 1.58, suggesting an increasing trend towards behavior-driven or strategic coding.

3.3 Diagnostic Coding Structure and Concentration

The results indicate more than a simple increase in the number of recorded diagnoses. Instead, the findings reflect a shift in how diagnoses are coded, with trends moving from broad and sometimes ambiguous diagnostic patterns toward a more focused use of clinically definitive ICD chapters. This trend is particularly evident in the reduced reliance on non-specific categories such as "Chapter XVIII" (symptoms and abnormal findings), which are often used when a precise diagnosis is yet to be established or is recorded in a delayed or incomplete manner. Concurrently, chapters typically associated with higher diagnostic certainty, such as "Chapter C00-D48" (Neoplasms) and "Chapter IX" (diseases of the circulatory system), increased over time, suggesting improved diagnostic clarity and stronger documentation.

The shift towards a more consolidated coding structure was more pronounced in Hospital A, which may reflect earlier adoption of standardized coding protocols and, potentially, greater emphasis on DRG-linked reimbursement requirements. In contrast, Hospital B exhibited more heterogeneous coding profiles, indicating differences in caseload complexity, workforce capacity, or the maturity of health information systems. Overall, the trends illustrated in Fig. 3 suggest a gradual move away from broad, non-specific coding toward a more streamlined, diagnosis-driven approach that supports clinical accuracy, reimbursement integrity, and interoperable health data.

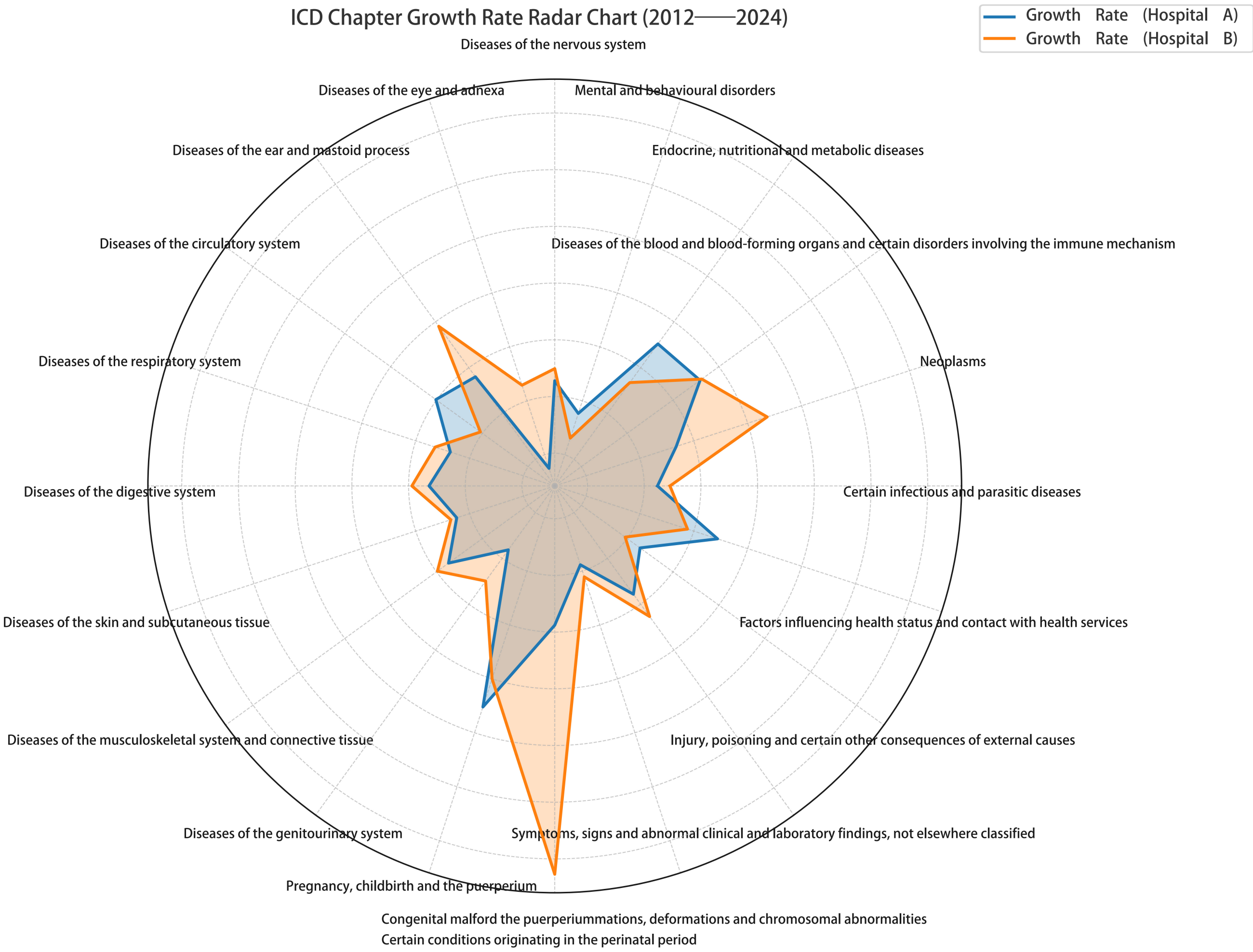

Fig. 3. Temporal changes in diagnostic structure highlight a clear realignment across ICD chapters. The visualization illustrates how emphasis on different ICD chapters has shifted over years and across hospital groups. The observed redistribution away from vague or generalized categories toward more definitive diagnostic chapters reflects increased coding specificity and a more optimized, structured approach to diagnostic documentation.

3.4 Temporal Convergence in Diagnostic Diversity and Structural Standardization

As detailed in Table 4, Hospital B showed a small but consistent reduction in the diversity index, decreasing from 2.997 in 2012 to 2.993 in 2019 and finally reaching 2.990 in 2024. Hospital A followed a similar trend initially, decreasing from 2.986 in 2012 to 2.977 in 2019, before showing a modest rebound to 2.980 in 2024. Despite this modest recovery, both hospitals showed lower diversity indices in 2024 than at baseline in 2012, reflecting an overall trend towards a more uniform coding pattern over the study period. Notably, the reduction in entropy does not suggest a reduced diagnostic detail. Instead, it is more likely to indicate increasing standardization in how diagnoses are distributed across ICD chapters (Fig. 4).

The box plots reveal that while the overall diversity index decreased, the chapter-level coding depth (average SDN) increased over time, especially in Hospital B. Median SDN increased across chapters, while the interquartile range gradually decreased, indicating more consistent recording of secondary diagnosis across different disease categories. Furthermore, the plots showed significant outliers (diamonds), particularly in Chapters XV (pregnancy, childbirth, and puerperium) and Chapter XVI (certain diseases originating in the perinatal period), where coding depth consistently exceeded the typical range.

Trend analysis confirmed a significant upward shift in coding depth in Hospital B ($p < 0.001$). Similarly, Hospital A showed a significant increase in coding depth ($p = 0.003$), even though its diversity index had a mild fluctuation in the later period. Collectively, these results suggest a trend towards more concise coding practices alongside deeper secondary diagnoses, particularly more evident in Hospital B.

3.5 Institutional-Level Diagnostic Signatures and Their Stabilization

To evaluate hospital-level variation in diagnostic coding patterns, we generated a heatmap of ICD chapter com-

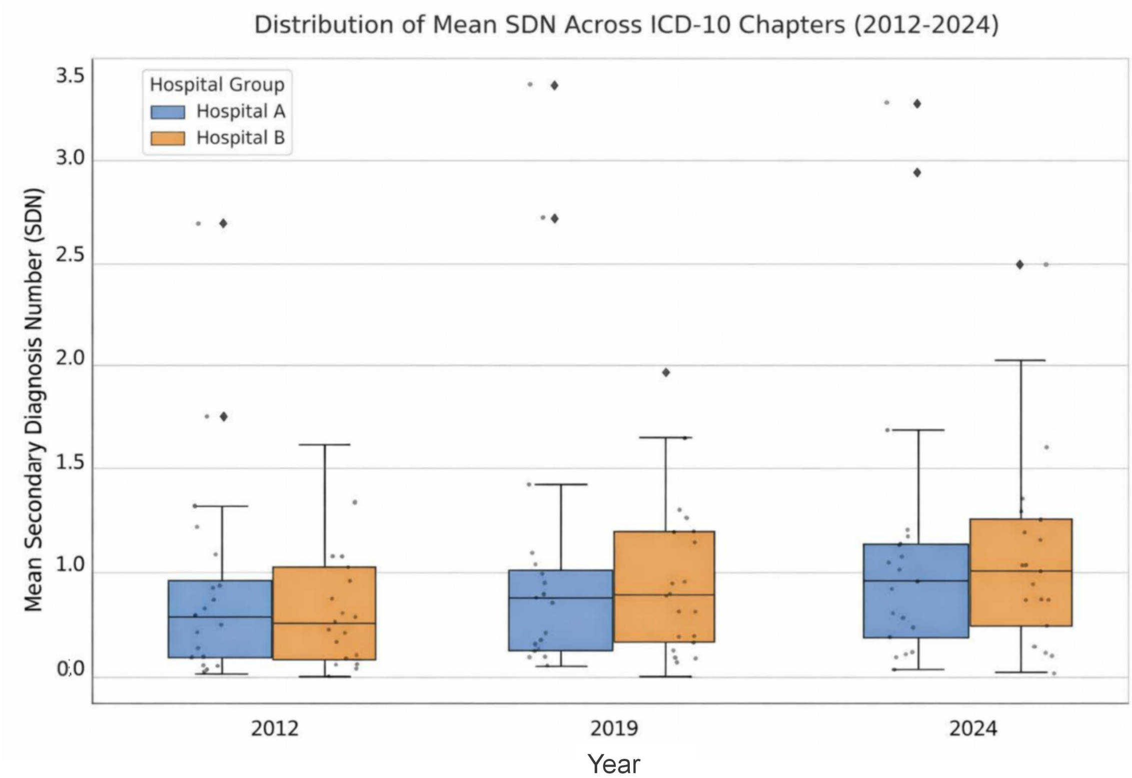

Fig. 4. Distribution of mean secondary diagnosis number (SDN) in ICD-10 chapters (2012–2024) across the 2012, 2019, and 2024. The box plots illustrate the distribution of mean SDN values across 21 ICD-10 chapters in Hospital A (blue) and Hospital B (orange) at three time points. Within each box, the central horizontal line represents the median SDN across all chapters, while the top and bottom boundaries indicate the interquartile range (IQR) (25th–75th percentiles). The whiskers extend to the highest and lowest values within 1.5 × IQR. ⧫ indicates outliers, which correspond to specific chapters (notably Chapter XV pregnancy, childbirth and the puerperium) and Chapter XVI (certain conditions in the perinatal period) that exhibited significantly higher secondary diagnosis counts than the overall chapter distribution. ⧫ indicates statistical outliers identified beyond 1.5 × IQR according to the boxplot algorithm. Because the boxplot summarizes chapter-level distributions within each year-hospital group, outlier symbols may appear without adjacent separately visible points when values overlap within the boxplot structure.

Table 4. Shannon diversity index (SDI) in both hospitals across the three years.

Hospital	2012	2019	2024
Hospital A	2.986	2.977	2.980
Hospital B	2.997	2.993	2.990

position across hospital-year combinations (Fig. 5). This provides a visual "diagnostic fingerprint" for each institution, offering insights into how consistently and how extensively specific ICD chapters are utilized over time. In the heatmap, color intensity represents the relative frequency of each ICD chapter, with darker shading indicating a higher proportion of diagnoses.

In 2012, the heatmap indicated significant heterogeneity, evidenced by mismatches in color patterns between the two hospitals. Specific chapters emerged as high frequency "hot spots" (darker cells) in Hospital A but appeared less prominent in Hospital B, reflecting meaningful differences in diagnostic focus and coding behavior at baseline.

By 2024, the inter-hospital variation reduced substantially. The 2024 heatmap rows for Hospital A and Hospital B exhibit a synchronized color distribution pattern, indicating that both institutions converged toward a more standardized diagnostic structure with similar high-frequency chapters, neoplasms (Chapter II), and circulatory system diseases (Chapter IX). The analysis showed consistently high representation of Chapter XXI (factors affecting health service contacts), underscoring its significance in routine documentations and its potential relevance to reimbursement systems. Conversely, chapters, such as congenital anomalies and perinatal conditions, remained consistently low, reflecting their specialized clinical scope in general tertiary care settings. This transition from clearly distinct 'diagnostic fingerprints' to a more uniform structure suggests system-level learning and alignment over time. This convergence may indicate the combined effects on national DRG policy implementation, wider adoption of best coding standards, diffusion of best practices, and gradual improvements in health information systems.

Overall, the results demonstrate a profound transformation in diagnostic coding over 12 years, with three consistent patterns. First, SDN increased steadily in both hospitals, reflecting more comprehensive documentation of comorbidities and complications. Second, the ICD chapter composition shifted structurally, with lower reliance on non-specific coding (particularly Chapter XVIII) and increased representation in more diagnostically specific, chronic disease-related categories, supporting improved coding precision. Third, these shifts occurred alongside systemic-wide convergence, as evidenced by a gradual reduction in the shannon diversity index and an increase in similarity of institutional coding signatures in the heatmap, suggesting greater homogeneity and policy-responsiveness in coding behavior across the two hospitals.

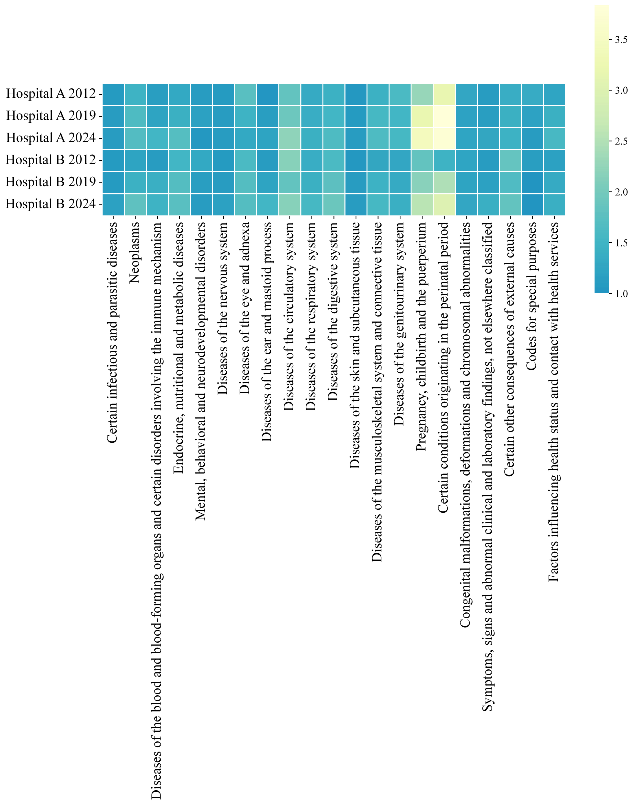

Fig. 5. Heatmap of normalized ICD chapter composition across hospitals and years. Each row represents the distribution of secondary diagnoses across 20 ICD-10 chapters for a given hospital–year combination (Hospital A and Hospital B, 2012, 2019, and 2024). Color intensity reflects the relative frequency or contribution of each chapter to the total number of SDs, with darker blue/green shades indicating higher relative frequency.

4. Discussion

With the continuous advances in DRG-based payment reform in China, hospital diagnostic coding has come under stronger policy oversight and operational pressure, with an increasing focus on standardization and greater coding precision [12]. This study analyzed 12 years of longitudinal data from two representative tertiary hospitals and provides empirical insights into how DRG reform reshapes diagnostic coding behavior. Particularly, the study assessed changes in secondary diagnosis documentation, shifts in ICD-10 chapter composition, trends in coding diversity, and the degree of convergence between institutions over time. Our findings are consistent with patterns reported internationally while also providing a China-specific perspective on how diagnostic governance is transforming in the context of the digital health era.

The significant upward trend in SDN across the three study years aligns with a previous study showing that DRG implementation tends to promote more comprehensive documentation of comorbidities and complications [4]. In payment systems where reimbursement is linked to case-mix severity, higher SDN can reflect better clinical documentation and improved diagnostic capture, but may also indicate coder responsiveness to financial incentives [13]. Our findings align with evidence from Germany and Korea, where DRG-based financing was linked to a measurable increase in SD coding per inpatient record, sometimes without a corresponding increase in true clinical complexity [14,15]. The consistent reduction in the SDN interquartile range across hospitals further suggests that differences in coding behavior between institutions are shrinking, likely due to national coding standards, more standardized digital recording systems, and audit-based feedback [16].

One of the most striking findings was the gradual reshaping of ICD-10 chapter composition over time. In both hospitals, the relative use of non-specific codes, particularly Chapter XVIII, reduced, while diagnostically more definitive categories such as Chapter II (Neoplasms), Chapter IV (Endocrine, nutritional and metabolic diseases), and Chapter IX (Diseases of the circulatory system) increased. This trend is consistent with experiences reported after DRG reforms in other settings, where symptom-based or provisional codes are less favored because they contribute less effectively to DRG classification and reimbursement [17]. In parallel, the considerable increase in the absolute mean number of Chapter XVIII secondary diagnoses per case, particularly in Hospital B, presents a nuanced picture. On the one hand, the declining proportion of Chapter XVIII indicates enhanced structural precision. On the other hand, the increase in absolute counts may reflect overall documentation expansion as SDN rises. Alternatively, it could reflect 'code creep', often presented as upcoding, in which vague codes are added to increase DRG weight even when they do not significantly enhance diagnostic specificity.

The increased use of Chapters XVII (Congenital malformations, deformations and chromosomal abnormalities), Chapter V (Mental and behavioral disorders), and Chapter XXI (factors influencing health status and contact with health services) can be interpreted in at least two ways. It may represent more holistic documentation of patient complexity and social or contextual factors. Equally, it could indicate more strategic coding behavior aimed at optimizing DRG grouping and reimbursement without necessarily altering clinical care. Given the policy emphasis on more "refined DRG grouping" under China's CHS-DRG and DIP systems, further research is needed to determine whether these changes reflect better capture of clinical reality or documentation shifts driven by incentive structures.

The gradual decline in the shannon diversity index suggests system-wide convergence in coding behavior, with secondary diagnoses becoming increasingly concentrated within a smaller subset of ICD chapters. While reduced diversity can sometimes indicate incomplete documentation or overly narrow diagnostic labeling, the trend found here more likely reflects improved coding consistency, closer alignment with DRG requirements, and overall institutional learning. Similar reduction in entropy has been observed in Organisation for Economic Co-operation and Development (OECD) settings after structured coding improvement initiatives [9].

Hospital A, as the earlier reform adopters, showed a more pronounced decline in diversity, which supports the hypothesis that coding behavior can shift in response to regulatory oversight and financial incentives. However, convergence raises its own concerns. The hospital may begin to rely on "safe" coding patterns that optimize reimbursement, potentially at the expense of capturing clinically meaningful diagnostic nuance. Prior evidence has cautioned that excessive standardization and resulting data homogeneity can hinder the granularity required for epidemiological analysis and personalized care planning [18]. Thus, the goal should be standardization that enhances comparability and integrity without compromising diagnostic fidelity.

This study indicates that DRG reforms are reshaping more than hospital financing. Reforms are also altering the underlying structure of clinical data through measurable shifts in diagnostic coding practices. The progressive changes in coding behavior observed here are consistent with models of institutional adaptation under performance-based incentives. For policymakers, this study highlights the dual role of diagnostic coding: it is both a clinical communication tool and a mechanism that directly influences payment. If payment frameworks become more sophisticated, coding quality must improve at some pace. This necessitates sustained investment in professional coder training, routine and automated quality control processes, and the incorporation of diagnostic decision-support systems within electronic health workflows.

Future studies should explore the causal links among coding behavior, DRG grouping and reimbursement accuracy, and downstream patient outcomes. Mixed-method approaches that integrate quantitative coding with coder and clinician interviews may help clarify the behavioral mechanisms behind observed patterns, including moral hazard, strategic documentation, and the administrative burden created by DRG requirements. Expanding similar analysis to county-level and non-tertiary hospitals will be essential to elucidate how China's DRG rollout affects equity and efficiency across different health settings.

We acknowledge several limitations in the present study. First, our analysis was restricted to inpatient discharge data from only two tertiary general hospitals in Beijing. While the two institutions provided useful comparative insight because of their distinct rates of DRG reform adoption, the results may not fully generalize to hospitals in other regions of China, such as lower-tier cities or rural areas, or to other hospital types, including specialized hospitals or secondary-level care facilities. Tertiary hospitals in Beijing typically operate with higher resource capacity and within a distinctive policy and information infrastructure environment, which may affect both coding depth and standardization. Future studies should include a wider range of hospital levels and geographical contexts to better capture national variation and assess the equity and efficiency implications of DRG reform rollout.

5. Conclusion

This study demonstrates that China's DRG reform is linked with measurable improvements in diagnostic coding depth, specificity, and structural consistency across hospitals. The observed increase in recorded secondary diagnoses, the reduced reliance on vague codes, and the increasing similarity in ICD chapter patterns together reflect stronger documentation practices and greater alignment with policy expectations. Furthermore, the findings suggest that coding behavior is becoming more standardized and may be influenced by DRG-linked incentives. These improvements should be interpreted alongside potential risks, including upcoding and the possibility that over-standardization could reduce clinically meaningful nuance. To maintain high-quality coding and support the long-term effectiveness of DRG-based payment reform, continued investment is required in coder training, well-integrated digital coding support systems, and audit systems that link coding quality not only to reimbursement accuracy but also to clinical outcomes.

Key Points

• Diagnostic coding in Chinese public hospitals became significantly more comprehensive over the 12-year period, as evidenced by a substantial increase in the mean secondary diagnosis number (SDN) and standardization in line with the objectives of DRG reform.

• The distribution of ICD chapter shifted toward greater diagnostic specificity, with a relative decline in vague codes (such as Chapter XVIII) and a corresponding increase in more definitive, chronic disease-related diagnoses.

• We found clear evidence of system-wide convergence, reflected by a reduction in the shannon diversity index and lower variation between hospitals, suggesting that coding practices are becoming more homogeneous and increasingly aligned with policy expectations across institutions.

• A nuanced dual trend in coding behavior was identified: although the relative proportion of vague codes decreased, their absolute volume increased, potentially reflecting the emergence of 'code creep' influenced by financial incentives.

• These trends indicate the likely beneficial effects of national DRG payment policies, institutional learning, and ongoing digitalization in improving diagnostic data quality and enhancing the accuracy and reliability of DRG-based reimbursement systems.

Abbreviations

DRG, Diagnosis-Related Group; SD, secondary diagnosis; SDN, secondary diagnosis number; CD, coding density; SDDI, secondary diagnosis diversity index; ICD, International Statistical Classification of Diseases and Related Health Problems; NHSA, National Healthcare Security Administration; IFP, inpatient front page; CHS-DRG, Chinese Healthcare Security Diagnosis-Related Group; DIP, Diagnosis-Intervention Packet.

Availability of Data and Materials

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Author Contributions

HBW: Study design, Data analysis, Methodology, Software, Writing—original draft. LLZ: Study design, Data analysis, Literature analysis, Software. JMC: Data collection, Literature analysis, Investigation, Validation. WYF: Data collection, Literature analysis, Software. All authors contributed to revising the manuscript critically for important intellectual content. All authors read and approved the final manuscript. All authors have participated sufficiently in the work and agreed to be accountable for all aspects of the work.

Ethics Approval and Consent to Participate

This study was conducted in accordance with the principles of the Declaration of Helsinki. Because this study did not involve human subjects or animal experiments, and analyzed de-identified data, the Ethics Committees of Peking University Third Hospital and Beijing Tongren Hospital, CMU exempted the need for formal ethical review and patient informed consent.

Acknowledgment

Not applicable.

Funding

This study was supported by the Research Project on the Management Status and Development Strategy of "One Hospital, Multiple Districts" in Public Hospitals in China funded by the National Health Commission of the People's Republic of China, Hospital Management Research Institute [Project No. 2024-106 (CON20241153281)].

Conflict of Interest

The authors declare no conflict of interest.

Supplementary Material

Supplementary material associated with this article can be found, in the online version, at https://doi.org/10.31083/BJHM52079.

References

[1] Dong H, Falis M, Whiteley W, Alex B, Matterson J, Ji S, *et al*. Automated clinical coding: what, why, and where we are? NPJ Digital Medicine. 2022; 5: 159. https://doi.org/10.1038/s41746-022-00705-7.

[2] Harrison JE, Weber S, Jakob R, Chute CG. ICD-11: an international classification of diseases for the twenty-first century. BMC Medical Informatics and Decision Making. 2021; 21: 206. https://doi.org/10.1186/s12911-021-01534-6.

[3] Souza RCD, Pinheiro RS, Coeli CM, Camargo KRD, Jr. The Charlson comorbidity index (CCI) for adjustment of hip fracture mortality in the elderly: analysis of the importance of recording secondary diagnoses. Cadernos De Saude Publica. 2008; 24: 315–322. https://doi.org/10.1590/s0102-311x2008000200010.

[4] Otero Varela L, Doktorchik C, Wiebe N, Quan H, Eastwood C. Exploring the differences in ICD and hospital morbidity data collection features across countries: an international survey. BMC Health Services Research. 2021; 21: 308. https://doi.org/10.1186/s12913-021-06302-w.

[5] Hu WY, Yeh CF, Shiao AS, Tu TY. Effects of diagnosis-related group payment on health-care provider behaviors: A consecutive three-period study. Journal of the Chinese Medical Association. 2015; 78: 678–685. https://doi.org/10.1016/j.jcma.2015.06.012.

[6] Campbell SE, Campbell MK, Grimshaw JM, Walker AE. A systematic review of discharge coding accuracy. Journal of Public Health Medicine. 2001; 23: 205–211. https://doi.org/10.1093/pubmed/23.3.205.

[7] Chen YJ, Zhang XY, Yan JQ, Xue-Tang, Qian MC, Ying XH. Impact of Diagnosis-Related Groups on Inpatient Quality of Health Care: A Systematic Review and Meta-Analysis. Inquiry. 2023; 60: 469580231167011. https://doi.org/10.1177/00469580231167011.

[8] Nicholls SG, Langan SM, Benchimol EI. Routinely collected data: the importance of high-quality diagnostic coding to research. Canadian Medical Association Journal. 2017; 189: E1054–E1055. https://doi.org/10.1503/cmaj.170807.

[9] Drösler SE, Romano PS, Tancredi DJ, Klazinga NS. International comparability of patient safety indicators in 15 OECD member countries: a methodological approach of adjustment by secondary diagnoses. Health Services Research. 2012; 47: 275–292. https://doi.org/10.1111/j.1475-6773.2011.01290.x.

[10] Li Q, Fan X, Jian W. Impact of Diagnosis-Related-Group (DRG) payment on variation in hospitalization expenditure: evidence from China. BMC Health Services Research. 2023; 23: 688. https://doi.org/10.1186/s12913-023-09686-z.

[11] Zhou W, Xu C, Zhang L, Fu H, Jian W. Behaviours and drivers of diagnosis-related group upcoding in China: A mixed-methods study. Social Science & Medicine (1982). 2025; 366: 117660. https://doi.org/10.1016/j.socscimed.2024.117660.

[12] Zhang H, Fu X, Wu Y, Tang Y, Jin H, Xie B. The Impact of DRG-Based Payment Reform on Inpatient Healthcare Utilization: Evidence from a Natural Experiment in China. Healthcare. 2025; 13: 2424. https://doi.org/10.3390/healthcare13192424.

[13] Anthun KS, Bjørngaard JH, Magnussen J. Economic incentives and diagnostic coding in a public health care system. International Journal of Health Economics and Management. 2017; 17: 83–101. https://doi.org/10.1007/s10754-016-9201-9.

[14] Böcking W, Ahrens U, Kirch W, Milakovic M. First results of the introduction of DRGs in Germany and overview of experience from other DRG countries. Journal of Public Health. 2005; 13: 128–137. https://doi.org/10.1007/s10389-005-0103-4.

[15] Jung YW, Pak H, Lee I, Kim EH. The Effect of Diagnosis-Related Group Payment System on Quality of Care in the Field of Obstetrics and Gynecology among Korean Tertiary Hospitals. Yonsei Medical Journal. 2018; 59: 539–545. https://doi.org/10.3349/ymj.2018.59.4.539.

[16] Liu Y, Wang G, Qin TG, Kobayashi S, Karako T, Song P. Comparison of diagnosis-related groups (DRG)-based hospital payment system design and implementation strategies in different countries: The case of ischemic stroke. Bioscience Trends. 2024; 18: 1–10. https://doi.org/10.5582/bst.2023.01027.

[17] Wang C, Yao C, Chen P, Shi J, Gu Z, Zhou Z. Artificial intelligence algorithm with ICD coding technology guided by embedded electronic medical record system in medical record information management. Microprocessors and Microsystems. 2023. https://doi.org/10.1016/j.micpro.2023.104962. (in press)

[18] Helgeland J, Kristoffersen DT, Skyrud KD, Lindman AS. Variation between Hospitals with Regard to Diagnostic Practice, Coding Accuracy, and Case-Mix. A Retrospective Validation Study of Administrative Data versus Medical Records for Estimating 30-Day Mortality after Hip Fracture. PLoS ONE. 2016; 11: e0156075. https://doi.org/10.1371/journal.pone.0156075.