Development of Diagnostic Terminology and Coding for General Practice and Empirical Study of Their Application: Postprint

Authors: Hong Yuchun, Wu Hua, Du Yishan, Li Shuran, Sun Wenmin, Ye Mingyu, Zhang Yongjian, Li Yang

Date: 2022-10-31T00:00:00+00:00

Abstract

Objective To construct a general practice diagnostic terminology and coding set for community health services, and validate its effectiveness, thereby promoting the standardization and normalization of community general practice diagnosis. Methods Through literature review and expert consultation, combined with the spectrum of common health problems and diseases in communities, a general practice diagnostic terminology and coding set applicable to community health services was constructed based on ICD-10; an empirical study was conducted by piloting the developed terminology and coding set in 259 community health service centers across two districts of Shenzhen from September 2021 to February 2022, using diagnostic data from the backend of the community health service information system and questionnaire surveys of community general practitioners as basic data to analyze the pilot implementation and validate the rationality and effectiveness of the diagnostic terminology and coding set. Results The compilation of the general practice diagnostic terminology and coding set was completed, including 41 general practice diagnostic categories with 6,629 standardized general practice diagnostic terms and codes. The pilot results from two districts in Shenzhen showed that the standardized composition of diagnostic names in the standardized intervention area was 86.13%-95.38% higher than in the non-pilot area; the diagnostic names involved in the pilot area during this period decreased by 96.60% year-on-year, with the standardization intervention greatly focusing commonly used general practice diagnoses, which were basically consistent with the community disease spectrum, demonstrating high aggregation of common disease diagnoses in the community. The survey showed that 87.3% (227/260) of respondents believed the coding set promoted standardization of community diagnosis, and 77.3% (201/260) supported its online application. Conclusion This study preliminarily established a standardized and normalized general practice diagnostic terminology and coding set, forming

a community general practice diagnostic terminology dictionary that solves the problem of multiple terms for one meaning during data analysis, while also accommodating upward and downward referrals and medical insurance integration, effectively improving community diagnosis and treatment efficiency and standardization levels, and positively promoting the development of primary healthcare.

Full Text

Development and Empirical Study of General Practice Diagnostic Terminology and Coding

Hong Yuchun[1], Wu Hua[2], Du Yishan[3], Li Shuran[4], Sun Wenmin[1], Ye Mingyu[5], Zhang Yongjian[6], Li Yang[1]*

1. Shenzhen Health Capacity Building and Continuing Education Center, Shenzhen 518000, China
2. Bao'an People's Hospital Community Health Service Management Center, Shenzhen 518000, China
3. International Centre for Higher Education Innovation under the auspices of UNESCO, Shenzhen 518000, China
4. Bao' an District Public Health Service Center, Shenzhen 518000, China
5. Futian District Community Health Service Management Center, Shenzhen 518000, China
6. Peking University Shenzhen Hospital, Shenzhen 518000, China

Co-first authors: Wu Hua, Hong Yuchun
Corresponding author: Li Yang, Assistant Researcher; E-mail: 15816883396@139.com

Abstract

Objective To construct a standardized diagnostic terminology and coding set for general practice in community health services and validate its effectiveness in promoting standardized and normalized community-based general practice diagnosis.

Methods Through literature review and expert consultation, we constructed a general practice diagnostic terminology and coding set based on ICD-10, incorporating the spectrum of common health problems and diseases in community settings. From September 2021 to February 2022, we conducted an empirical study by piloting this terminology and coding set across 259 community health service centers in two administrative districts of Shenzhen. Using diagnostic data from the community health service information system backend and questionnaire surveys of community general practitioners, we analyzed the pilot implementation to verify the rationality and effectiveness of the diagnostic terminology and coding set.

Results We compiled a comprehensive general practice diagnostic terminology and coding set comprising 41 diagnostic categories with 6,629 standardized terms and codes. The Shenzhen pilot demonstrated that the proportion of standardized diagnostic names in intervention areas reached 100%, representing an 86.13%-95.38% improvement over non-pilot areas. During the pilot period, the number of diagnostic names decreased by 96.60% year-over-year. Standardized intervention significantly focused common general practice diagnoses, aligning closely with the community disease spectrum and achieving high aggregation of common disease diagnoses. Survey results showed that 87.3% (227/260) of respondents believed the coding set promoted diagnostic standardization, and 77.3% (201/260) supported its system-wide implementation.

Conclusion This study establishes a preliminary standardized and normalized general practice diagnostic terminology and coding set, creating a community general practice diagnostic dictionary that resolves the challenge of multiple terms for single concepts in data analysis while accommodating two-way referrals and medical insurance integration. The system effectively improves community diagnostic efficiency and standardization, positively contributing to primary healthcare development.

Keywords: General Practice; Community Health Services; Diagnosis; Terminology Coding

Introduction

As gatekeepers of community health, primary healthcare institutions bear the responsibility of providing first-contact care and represent a critical component in advancing the tiered diagnosis and treatment system [1]. The deepening of China' s new healthcare reform and the shift of medical resources to community levels have led to rapid growth in patient volumes, while the establishment of two-way referral channels has correspondingly increased patient transfers between primary and higher-level institutions [2]. With the development of community health services, government departments and medical institutions have raised requirements for data accuracy. However, the long-standing absence of standardized diagnostic terminology and coding suitable for China' s general practice development context has resulted in non-standardized and unstandardized diagnoses in basic medical care across most regions, severely hindering the advancement of general practice in China [3,4].

Establishing a unified general practice diagnostic terminology and coding set (hereinafter referred to as "general practice terminology and coding") to enable interoperability and information sharing among community health service institutions and between primary and higher-level healthcare facilities can not only facilitate implementation of the tiered diagnosis and treatment system [4,5] but also enable statistical analysis of diseases in community general practice, laying the foundation for evaluating general practitioners' professional competen-

cies, conducting epidemiological surveys, and monitoring residents' health status. This represents an important pathway for enhancing community health service capacity and developing the general practice system. Therefore, constructing a unified, sustainable, and maintainable general practice terminology and coding applicable to China' s primary healthcare context is imperative [6].

1.1.1 Literature Review Method

This study conducted domestic and international literature reviews and examined municipal government documents to inform the principles and methods for developing general practice diagnostic coding. Based on ICD-10, we constructed a general practice diagnostic terminology and coding set. ICD, the International Classification of Diseases, is the internationally recognized standard classification for health information and the most widely adopted unified disease classification system globally. ICD-10, the tenth revision, is currently the most extensively used disease classification version worldwide [6], characterized by its comprehensiveness, consistency, integrity, and practicality. It enables easy retrieval and achieves one disease-one code, scientifically adhering to disease classification principles and evaluation criteria, making it a relatively complete international disease classification method [7]. China has conducted numerous practical explorations in ICD-10 implementation, including the National Health and Family Planning Commission' s ICD 2012 edition (6-digit extension codes), Beijing' s ICD-101 clinical version (6-digit extension codes), and Shanghai' s ICD-10 (7-digit detailed codes) [8-9]. However, ICD-10 application remains controversial and requires further localization research and empirical validation [3].

1.2.2 Expert Consultation Method

We assembled an expert panel comprising 36 experienced professionals, including general practice specialists, clinical specialists, and medical record management and coding experts. Six expert consultation meetings were held to determine experimental methods, standardized corpus selection principles, and research approaches. Based on the spectrum of common community health problems and diseases, we screened standardized general practice diagnostic terms and codes from ICD-10 (Chinese version), established corresponding relationships with ICD-10 specialist diagnoses, and compiled the general practice terminology and coding.

1.2.1 Pilot Implementation

Using cluster sampling, we selected 259 community health service centers across two administrative districts in Shenzhen for pilot application from September 2021 to February 2022 (six months). Community general practitioners in pilot

areas uniformly used the general practice terminology and coding in daily clinical practice and medical record writing, dynamically updating and maintaining the system based on actual diagnostic needs. During data collection, we ensured uniform reporting standards to improve data quality [6].

The pilot implementation included: (1) upgrading the Shenzhen Community Health Service Information System to embed standardized general practice diagnostic terminology and codes into the pilot area information platform to cover general practitioners' daily clinical activities; (2) printing and distributing the *Shenzhen General Practice Diagnostic Classification and Coding Manual (Trial)* and conducting related training; (3) establishing pilot safeguard mechanisms and feedback/revision mechanisms for new codes, forming a pilot work team to ensure smooth implementation; and (4) creating a pilot working group to facilitate problem reporting and resolution channels [5].

1.2.2 Applicability and Standardization Verification

We collected diagnostic data from the community general practice platform before and after implementation in pilot areas and compared it with concurrent data from community health service centers in other administrative districts to analyze and validate the applicability and standardization of the general practice terminology and coding. Chi-square analysis was used to test the significance of changes in diagnostic standardization after pilot implementation, using the frequency of standardized versus non-standardized disease terminology as indicators. Improved standardization was determined by increased use of standardized disease terminology and decreased use of non-standardized terms. Verification objects were categorized along spatial and temporal dimensions: pilot versus non-pilot area comparisons, and pre-pilot versus post-pilot comparisons.

1.2.3 Satisfaction Survey Analysis

Following the pilot period, we distributed questionnaires to general practitioners, institutional managers, and medical quality control personnel in pilot areas. The questionnaire covered general demographic variables (institution, position, professional title) and satisfaction measurement variables regarding the general practice diagnostic coding, including perceived degree of promotion of diagnostic standardization (5-point scale) and attitudes toward system implementation (5-point scale), concluding with questions about suggestions for information system improvements.

Table 1 describes the variables used in empirical analysis.

2.1 Composition of General Practice Terminology and Coding

Based on ICD-10, this study' s general practice terminology and coding comprises three components: (1) **Core General Practice Diagnoses**: 3,476 high-frequency diagnostic terms and codes corresponding to common community health problems and disease spectra; (2) **Comprehensive General Practice Diagnoses**: 6,629 terms and codes corresponding to non-common community health problems but retained to facilitate two-way referrals; and (3) **Specialist Diagnoses**: All 35,867 ICD-10 diagnostic terms and codes retained as specialist diagnoses. Both general practice and specialist diagnostic codes maintain corresponding relationships, preserving the one disease-one code principle (see **Table 2**).

Table 2 Introduction to General Practice Diagnostic Terminology and Coding Chapter

Chapter	Specialist Diagnosis Standardized Terms (items)	Core General Practice Diagnosis Standardized Terms	Comprehensive General Practice Diagnosis Standardized Terms (items)
Infectious Diseases			
Blood System Diseases			
Rheumatic and Immunological Diseases			
Endocrine System Diseases			

Chapter	Specialist Diagnosis Standardized Terms (items)	Core General Practice Diagnosis Standardized Terms	Comprehensive General Practice Diagnosis Standardized Terms (items)
Mental and Neurological Diseases			
Circulatory System Diseases			
Blood Circulation System Diseases			
Respiratory System Diseases			
Digestive System Diseases			
Abdominal and Pelvic Diseases			
Dermatological Diseases			

Chapter	Specialist Diagnosis Standardized Terms (items)	Core General Practice Diagnosis Standardized Terms	Comprehensive General Practice Diagnosis Standardized Terms (items)
Bone and Joint Diseases			
Immune System Diseases			
Muscle, Joint, and Nerve Lesions			
Urinary System Diseases			
Gynecological and Obstetric Diseases			
Abnormal Examination Results			
Injury and Poisoning			

Chapter	Specialist Diagnosis Standardized Terms (items)	Core General Practice Diagnosis Standardized Terms	Comprehensive General Practice Diagnosis Standardized Terms (items)
Injury Sequelae Contact, Exposure, and Pathogen Infestation Plastic Surgery and Allogeneic Implantation Potential Health Problems Problems Related to Medical Facilities and Other Healthcare			

2.2.1 Applicability and Standardization Analysis

We extracted diagnostic data from Shenzhen's community health service centers for the post-pilot period (October 2021-February 2022) and the same period in the previous year (October 2020-February 2021) for descriptive analysis and comparison. Results showed:

(1) **Diagnostic Volume and Standardization Across Shenzhen Districts**: Non-pilot districts involved up to 80,807 diagnostic names during the period, with standardized name composition ratios between 4.62%-13.87%. In contrast, the standardized intervention districts of Futian and Bao' an achieved 100% standardized diagnostic name composition, representing an 86.13%-95.38% improvement over non-pilot districts (see **Table 3**). Chi-square testing from a spatial sequence perspective (pilot vs. non-pilot areas) yielded a chi-square value of 197,180, indicating statistically significant differences in standardization between pilot and non-pilot areas ($p < 0.05$) (see **Figure 1 [Figure 1: see original paper]**).

Table 3 Diagnosis Data of Administrative Districts (New Districts) in Shenzhen from October 2021 to February 2022

District	Diagnostic Names Involved	ICD-10 Consistent Diagnostic Names (types)	Composition Ratio (%)
Pilot District 1			
Pilot District 2			
Non-pilot District 1			
Non-pilot District 2			
Non-pilot District 3			
Non-pilot District 4			

District	Diagnostic Names Involved	ICD-10 Consistent Diagnostic Names (types)	Composition Ratio (%)
Non-pilot District 5			
Non-pilot District 6			
Non-pilot District 7			
Non-pilot District 8			

Figure 1 [**Figure 1: see original paper**] Results of Normative Chi-square Test in Trial Area and Non-trial Area

(2) **Pre-Post Comparison in Two Pilot Districts**: General practice diagnostic volume decreased by 50.65% year-over-year, while involved diagnostic names decreased by 96.60%. The proportion of top 50 diagnoses in total diagnostic volume increased by 15.92%. The top 10 diagnoses during the pilot period were hypertension, COVID-19 screening, diabetes, hyperlipidemia, health examination, dyspepsia, sleep disorders, type 2 diabetes, chronic gastritis, and essential hypertension, accounting for 39.24% of total diagnoses (see **Table 4**). Standardization of common disease diagnoses improved substantially. For example, "hypertension" diagnoses were consolidated from 1,934 variations (e.g., "hypertension," "grade 2 hypertension," "grade 2 hypertension (very high risk)," "hypertension?," "hypertension medication refill") to 36 standardized terms, representing a 98.40% reduction (see **Table 5**). Chi-square testing from a temporal sequence perspective (pre-pilot vs. post-pilot) also yielded a chi-square value of 197,180, with $p < 0.05$, indicating statistically significant differences (see **Figure 2** [**Figure 2: see original paper**]).

Table 4 Comparison of Diagnostic Data Before and After the Pilot in the Two Administrative Regions

Indicator	Year-over-year Change (%)
Diagnostic Volume (entries)	

Indicator	Year-over-year Change (%)
Diagnostic Names Involved (types)	
Top 10 Diagnosis Frequency Total	
Top 10 Diagnosis Frequency Proportion (%)	
Top 50 Diagnosis Frequency Total	
Top 50 Diagnosis Frequency Proportion (%)	

Table 5 Changes in the Number of Common Diagnostic Names Before and After the Pilot

Common Diagnostic Main Fields	Diagnostic Names Involved (types)	Pre-pilot	Post-pilot	Year-over-year Decrease (%)
Upper Respiratory Infection				

Figure 2 [Figure 2: see original paper] Normative Chi-square Test Results Before and After the Pilot in the Pilot Area

2.2.2 Satisfaction Survey Results

We distributed questionnaires to general practitioners, institutional managers, and medical quality control personnel in pilot areas, receiving 260 valid responses. Analysis revealed:

(1) **General Demographics**: All 260 returned questionnaires were valid (100.0%). Respondents included 209 general practitioners (80.4%), 21 community health center directors and managers (9.2%), 17 medical quality control staff (6.5%), and 10 other related personnel (3.9%). Professional titles comprised 167 intermediate-level practitioners (64.2%), 50 junior practitioners (19.2%), 32 associate chief physicians (12.3%), 2 chief physicians (0.8%), and 9 others (3.5%) (see **Table 6**).

Table 6 General Statistics

Category	Percentage
Medical/Quality Control Related Business	6.54%
Community Health Management Center Administration	1.15%
Community Health Center Director	8.08%
Chief Physician (Senior)	3.85%
Associate Chief Physician (Deputy Senior)	0.77%
Attending Physician (Intermediate)	80.38%

Category	Percentage
Physician (Junior)	12.31%
Other	64.23%
	19.23%
	3.46%

(2) **Satisfaction with General Practice Diagnostic Terminology and Coding**: 87.3% (227/260) of respondents believed the terminology and coding promoted community diagnostic standardization, with 70% (182/260) reporting improved standardization. 85.0% (221/260) did not oppose system implementation, with 77.3% (201/260) expressing support (see **Figure 3 [Figure 3: see original paper]**).

Note: The left chart shows the degree to which respondents believed the diagnostic coding promoted standardization; the right chart shows attitudes toward system implementation.

Figure 3 [Figure 3: see original paper] Statistical Results of Satisfaction with General Practice Diagnostic Terminology and Coding

3.1 The Terminology and Coding System Promotes Community Diagnostic Standardization

The absence of standardized and efficient primary care terminology and coding has led many regions to allow physicians to manually write and input diagnostic names, resulting in varied disease descriptions, non-uniform diagnostic names for the same condition, and non-rigorous definitions. These inconsistencies cause clinical information loss and affect the reliability of disease classification, clinical evaluation, and statistical analysis [10]. As this study demonstrates, in Shenzhen' s community health service centers that allowed manual diagnostic entry (non-pilot areas), diagnostic name variations reached over 80,000 types, creating significant difficulties for primary care diagnostic data analysis and application.

Based on ICD-10 disease classification and coding, combined with community health service characteristics, this study developed general practice terminology and coding guided by practical work requirements. The system accommodates the comprehensive and broad nature of community health services with non-uniform diagnostic granularity—primarily moderately coarse while allowing granularity for residents' health problems, considering two-way referrals, generalist-specialist coordination, and medical insurance reimbursement, while minimizing burden on general practitioners. Pilot application results achieved intended objectives: standardized diagnostic name composition reached 100% in intervention areas, 86.13%-95.38% higher than non-pilot areas. During the pilot period, diagnostic volume decreased by 50.65% and involved diagnostic names

decreased by 96.60%, primarily because the standardized system required physicians to select standardized names, eliminating inaccurate or non-standard entries. The proportion of top 50 diagnoses increased by 15.92%, with the top 10 diagnoses accounting for 39.24% of total volume, indicating that standardized intervention focused common general practice diagnoses and aligned with the community disease spectrum. Standardization of common disease diagnoses improved substantially—for hypertension, diagnoses consolidated from 1,934 variations to 36 standardized terms (98.40% reduction). Satisfaction surveys showed 87.3% (227/260) of respondents believed the terminology promoted standardization.

3.2 ICD-10 Provides Theoretical and Practical Basis for Localized General Practice Terminology and Coding

Medical classification codes store, retrieve, and analyze disease-related data internationally for mortality, injury, epidemiological surveys, medical insurance payment, performance management, and other applications [11]. Currently, the main disease classification systems are the World Organization of Family Doctors' (WONCA) International Classification of Primary Care (ICPC-2) and ICD-10. ICD-10 emphasizes epidemiological and health management purposes and is internationally recognized as the standard health information classification and important foundational data for medical and health statistics [11].

Most Chinese medical institutions currently use ICD-10. However, ICD-10' s diagnostic classification and coding are primarily based on medical specialties with sufficiently detailed granularity, while general practice emphasizes comprehensiveness and breadth. Direct ICD-10 application in primary care results in overly specialized diagnostic entries, forcing general practitioners to select from excessive options and severely impacting efficiency. Additionally, general practice requires social, psychological, and non-disease health problem diagnoses beyond common community diseases. To address these issues, ICPC-2 has been adopted in some foreign primary care institutions and attempted in Shenzhen, but China' s general practice service scope differs from other countries, making it unsuitable for current Chinese community health service needs [12,13]. This study developed general practice terminology and coding by screening ICD-10 based on community health problem and disease spectra, establishing corresponding relationships and usage rules with ICD-10 specialist diagnoses. Pilot practice and empirical analysis demonstrated significant effectiveness and standardization improvements with high satisfaction among general practitioners, confirming that ICD-10 provides theoretical and practical basis for developing China' s localized general practice terminology and coding.

3.3 Information System Development Is the Next Improvement Direction for Local Implementation

This study demonstrates the necessity of information system support for general practice terminology and coding application. Further information system development represents the next improvement direction, focusing on enhanced personalization. First, logical judgments should be implemented to prevent unnecessary physician errors [14]. Second, personalized modifications should be made according to each institution' s and physician' s diagnostic characteristics and habits to accommodate differences in resource allocation and development directions across community health service institutions—for example, prioritizing frequently used individual or recent diagnoses to improve clinical efficiency and medical record quality.

This study preliminarily constructed a standardized general practice diagnostic terminology and coding database that unifies multi-source data through information standard alignment, connecting community health service institutions with higher-level hospitals and medical insurance departments while accommodating two-way referrals and insurance reimbursement. This facilitates understanding of primary care disease spectra, grasping actual community healthcare service demands, promoting primary care institution development and general practitioner capacity building, advancing tiered diagnosis and treatment and general practice system construction, and providing scientific references for public health planning and administrative management.

Limitations: First, the implementation period was relatively short, requiring further validation of generalizability and replicability through expanded application. Second, community Traditional Chinese Medicine diagnostics have not yet been standardized. The next step will involve developing standardized TCM diagnostics based on experiences from general practice diagnostic standardization.

Author Contributions: Wu Hua and Li Yang conceptualized and designed the study; Hong Yuchun collected and organized data and drafted the manuscript; Du Yishan conducted formal analysis; Wu Hua and Zhang Yongjian revised the manuscript; Sun Wenmin, Li Shuran, and Ye Mingyu provided project guidance and coordination; Li Yang performed quality control and final review, taking overall responsibility for the article.

Conflict of Interest: The authors declare no conflict of interest.

References [1] Shen SG, Zhang B. Tiered diagnosis and treatment, primary first-contact care, and primary healthcare institution construction [J]. Xuehai, 2016, (2): 48-57. [2] Chen MS, Lai W, Zhang ZC, et al. Survey on awareness and willingness of two-way referral in Songjiang District, Shanghai [J]. Journal of Shanghai Jiaotong University (Medical Science), 2010, 30(3): 340-342. [3] Wang M, Yu GJ, Liu HF, Liu YB, Sun JQ. Establishment of disease data quality management system based on ICD-10 coding application [J]. Chinese

Hospital Management, 2017, 37(05): 42-44. [4] Zhou JY, Bai X, Cui SN, Pang C, Liu AM. Systematic evaluation of ICD-10 disease classification coding quality in China [J]. Chinese Hospital Management, 2015, 35(12): 32-35. [5] Hao XN, Ma CY, Liu ZY, et al. Study on satisfaction of doctors and patients with primary healthcare information construction [J]. Health Economics Research, 2020, 37(7): 6-9. [6] Zhang M, Ci PW, Liu HM, et al. Design and construction of disease classification system for primary healthcare in China [J]. Chinese Medical Record, 2017, 18(2): 33-36. [7] Xu W, Tao TP. International Classification of Diseases (ICD10) and coding practice [J]. Chinese Health Statistics, 2003(05): 35-36. [8] Lu CL, Wang X, Cao HX. Reflections on localization research of ICD-10 [J]. Journal of Navy Medicine, 2016, 37(02): 190-191. [9] Yang XY, Wang HN, Yu LH, et al. Clinical application and practice of "Chinese Clinical Disease Diagnosis Standard Terminology Set" [J]. Chinese Health Economics, 2021, 40(4): 27-29. [10] Wen HL, Lu YN, Luo YJ, et al. Impact of disease classification on promoting tiered diagnosis and treatment system construction [J]. Modern Hospital, 2020, 20(11): 1593-1596. [11] Zhang XY, Xie M. Comparison and enlightenment of medical classification coding in health fields between China and the United States [J]. Chinese Medical Record, 2018, 19(9): 9-12. [12] Lu XQ, Liang WN, Jia QC, et al. International Classification of Primary Care (ICPC) and its application in general practice/family medicine [J]. Chinese General Practice, 2003, 6(1): 86-88. [13] Zhang M, Ci PW, Liu HM, et al. Design and construction of disease classification system for primary healthcare in China [J]. Chinese Medical Record, 2017, 18(2): 33-36. [14] Wang WJ, Zhou DP, Xu SM, et al. Analysis of factors influencing resident physicians' medical record homepage completion quality based on logistic model [J]. Chinese Health Statistics, 2016(4): 261-263.

Note: Figure translations are in progress. See original paper for figures.