BMJ Open

Validation of an algorithm to evaluate the appropriateness of outpatient antibiotic prescribing using big data of Chinese diagnosis text

Houyu Zhao [1], Jiaming Bian,[2] Li Wei,[3] Liuyi Li,[4] Yingqiu Ying,[5] Zeyu Zhang,[6] Xiaoying Yao,[1] Lin Zhuo [7], Bin Cao,[6] Mei Zhang,[2] Siyan Zhan [1]

To cite: Zhao H, Bian J, Wei L, *et al*. Validation of an algorithm to evaluate the appropriateness of outpatient antibiotic prescribing using big data of Chinese diagnosis text. *BMJ Open* 2020;**10**:e031191. doi:10.1136/bmjopen-2019-031191

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (http://dx.doi.org/10.1136/bmjopen-2019-031191).

Received 22 April 2019
Revised 10 December 2019
Accepted 02 March 2020

For numbered affiliations see end of article.

Correspondence to
Dr Siyan Zhan;
siyan-zhan@bjmu.edu.cn

ABSTRACT

Objective We aimed to evaluate the validity of an algorithm to classify diagnoses according to the appropriateness of outpatient antibiotic use in the context of Chinese free text.

Setting and participants A random sample of 10 000 outpatient visits was selected between January and April 2018 from a national database for monitoring rational use of drugs, which included data from 194 secondary and tertiary hospitals in China.

Research design Diagnoses for outpatient visits were classified as tier 1 if associated with at least one condition that 'always' justified antibiotic use; as tier 2 if associated with at least one condition that only 'sometimes' justified antibiotic use but no conditions that 'always' justified antibiotic use; or as tier 3 if associated with only conditions that never justified antibiotic use, using a tier-fashion method and regular expression (RE)-based algorithm.

Measures Sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) of the classification algorithm, using classification made by chart review as the standard reference, were calculated.

Results The sensitivities of the algorithm for classifying tier 1, tier 2 and tier 3 diagnoses were 98.2% (95% CI 96.4% to 99.3%), 98.4% (95% CI 97.6% to 99.1%) and 100.0% (95% CI 100.0% to 100.0%), respectively. The specificities were 100.0% (95% CI 100.0% to 100.0%), 100.0% (95% CI 99.9% to 100.0%) and 98.6% (95% CI 97.9% to 99.1%), respectively. The PPVs for classifying tier 1, tier 2 and tier 3 diagnoses were 100.0% (95% CI 99.1% to 100.0%), 99.7% (95% CI 99.2% to 99.9%) and 99.7% (95% CI 99.6% to 99.8%), respectively. The NPVs were 99.9% (95% CI 99.8% to 100.0%), 99.8% (95% CI 99.7% to 99.9%) and 100.0% (95% CI 99.8% to 100.0%), respectively.

Conclusions The RE-based classification algorithm in the context of Chinese free text had sufficiently high validity for further evaluating the appropriateness of outpatient antibiotic prescribing.

Strengths and limitations of this study

- This is the first study to establish rules for evaluating appropriateness of outpatient antibiotic prescriptions by using Chinese diagnosis text.
- The rule-based and tier-fashion algorithm for classifying diagnoses according to whether antibiotic was indicated or not had high validity.
- The algorithm provides a feasible method to use big electronic medical records or claim data to evaluate the appropriateness of antibiotic use in China.
- Certain synonyms, abbreviations or acronyms of medical terms may be omitted thus cannot be detected by the algorithm.

INTRODUCTION

Overuse of antibiotics and the consequential antimicrobial resistance (AMR) have been serious threats to public health worldwide.[1] China is one of the countries with the highest antibiotic consumption and hence, high prevalence of AMR in the world.[2 3] Reducing unnecessary and inappropriate use of antibiotics is essential to reduce both AMR and adverse drug reactions,[4 5] which cause a large number of deaths and economic losses every year.[4 6 7] In China, the evaluation of the appropriateness of antibiotic use has occurred mainly thorough manual prescription review, which is time consuming and, hence, it is not feasible for evaluating large-scale prescription data. The rapid implementation of the electronic medical records (EMRs) across the whole country has enabled the routine collection of medical data available for research purposes.[8 9] However, a huge amount of these data, especially diagnosis information, is in the form of Chinese free text, which makes it difficult for evaluating the appropriateness of antibiotic use.

Previous studies using big EMR or claim data for the evaluation of appropriate antibiotic prescribing applied an innovative tier-fashion method to classify and assign diagnosis to the outpatient visits based on whether antibiotic is indicated for treatment or not.[4 10 11] The

key aspect of this method is to classify all diseases into three tiers corresponding to whether the disease always, sometimes or never justifies antibiotic use[4 5 11]; that is, tier 1 diagnoses are diseases for which antibiotics are almost always indicated, such as pneumonia and specific bacterial infections; tier 2 diagnoses are diseases for which antibiotics are only sometimes indicated, such as sinusitis and pharyngitis; finally, tier 3 diagnoses are all the other diseases for which antibiotics are not indicated or the indication is unclear, such as influenza and cancer. However, most of these studies are conducted in the USA and the UK where 20%–30% of outpatient antibiotics are estimated to be used inappropriately,[4 10 11] with little reliable evidence from China and other developing countries where antibiotic use is high.[2 12]

The aim of this study was to establish and validate an regular expression (RE)-based algorithm for extracting and classifying outpatient diagnosis from the Chinese free text using the tier-fashion method mentioned above.

METHODS

Data sources

We used a national database for monitoring the rational use of drugs. Online supplementary appendix 1 describes the recruitment process of the hospitals and the representativeness of the database. In total, there were 194 hospitals from 128 cities of 31 provinces, autonomous regions or municipalities in China. The database consisted of both outpatients' and inpatients' information of demographic characteristics, prescriptions, item costs and diagnoses from EMR of sample hospitals between October 2014 and April 2018. The prescription for each outpatient visit consisted of three parts, which were recorded in different tables in the database. The preface mainly included basic information of the patients, such as gender and age, and diagnosis; the main body consisted of a list of drug information, including drug name, dosage and usage; and the postscript included other information such as doctor's signature. All information generated during the same visit could be linked by a unique identifier consisting of the hospital code, patient identification number and the date of visit. Chemical drugs including antibiotics were coded according to the Anatomical Therapeutic Chemical classification system.[13] Outpatient diagnosis in the database was in the form of Chinese narrative free text. Several diagnoses could be written together and separated by punctuation marks. We treated multiple prescriptions and diagnoses from the same patient on the same day in the same hospital as one visit. Thus, multiple diagnose and drugs could be linked to the same visit. At present, there were 239 million outpatient visits and 170 million diagnosis records. Fifty-five hospitals did not submit diagnosis records and 88% prescriptions from the remaining 139 hospitals could be linked to at least one valid diagnosis. This proportion was comparable to previous studies.[5 11] In this study, antibiotics for systemic use were evaluated (see online supplementry appendix 2 for the list of antibiotics in the database).

Diagnosis classification

Outpatient diagnoses were processed in three steps (see figures 1 and 2 in online supplementry appendix 3): standard tiers classifying, REs establishing, and dictionary and pattern mapping.

Standard tiers classifying

In the first step, based on the standard description and classification of the International Statistical Classification of Diseases and Related Health Problems 10th Revision (ICD-10) in 2016 Chinese version,[14] we first classified all diseases into three tiers according to whether antibiotics are indicated or not: tier 1 if the disease almost always justifies use of antibiotics, such as pneumonia; tier 2 if the disease only sometimes justifies use of antibiotics, such as sinusitis; or tier 3 if the disease almost never justifies use of antibiotics, such as cancer. Two groups of researchers worked on this independently and in parallel, the first group classified the standard diagnoses according to the list of categories established in previously published researches.[4 5 11] The other group consisting of one clinician and one pharmacist performed the same classification based on their clinical knowledge and experience. If there were any conflicts between the results of the two groups, recommendations for antibiotic use from the Guidelines for Clinical Application of Antibiotics[15] and UpToDate[16] were referenced and the final tiers of the diagnoses were determined. Online supplementry appendix 4 gives the basic rules that we used to classify the diagnoses. Then, the primary list of the tiers of diagnoses was reviewed by an expert of respiratory disease, an expert of infectious disease and an experienced pharmacist in three of the top hospitals in Beijing. Conflicts of the reviewers were discussed and the primary list was modified, if necessary. Finally, we classified 34 337 ICD codes, among which 2465 (7.2%) were classified as tier 1, 2608 (7.6%) were classified as tier 2 and 29 264 (85.2%) were classified as tier 3. The final list of the standard tiers of diagnoses (LiSToD) is available from https://www.researchgate.net/publication/336286590_LiSToD_The_list_of_the_standard_tiers_of_diagnoses_based_on_China-ICD-10. Online supplementry appendix 5 gives a comparison with the classification schemes from previous studies. Furthermore, we set up more diagnostic subcategories according the ICD-10 chapters and previous studies,[4 5 17 18] and finally all diagnoses were classified into 15, 21 and 13 different categories of tier 1, tier 2 and tier 3 diagnoses, respectively (see online supplementary appendix 6 for more details).

RE establishing

In the second step, we first established the dictionary and patterns of the clinical terms in the LiSToD. The clinical terms were derived from the following sources: (1) code string descriptions of Chinese version of ICD-10 and

corresponding synonyms or abbreviations; (2) pilot string searches for the diagnoses in the LiSToD; (3) clinicians' suggestions about the common abbreviations (in Chinese or in English) of infectious diseases. Similar patterns or key words of different descriptions of the same condition in the raw data were identified and used to establish the REs which can be used for information extraction. For example, *Helicobacter pylori* infection coded as A49.809, standard Chinese name in the ICD-10 system would be '幽门螺旋杆菌感染', but it may be written in the raw diagnosis text as '幽门螺旋杆菌', '幽门螺杆菌', '幽门螺旋菌', '幽门杆菌', '幽门螺菌', or even '幽门螺旋杆', '幽门螺杆' or in English abbreviation 'HP', 'Hp', 'hP' or 'hp'. Thus, the REs for matching the diagnoses of *H. pylori* infection after converting all the letters to the uppercase would be '(幽门螺?旋?杆?菌)|(幽门螺旋?杆)' and '([^A-Z]HP[^A-Z])|([^A-Z]HP$)|(^HP[^A-Z])|(^HP$)' (this subexpression means that the abbreviation of 'HP' cannot be prefixed or suffixed by any other letter). Finally, we established the lists of REs for diagnoses (REoD) in the LiSToD. Online supplementry appendix 7 gives more details of the rules we used for constructing REs.

Dictionary and pattern mapping

In the third step (figure 2 in online supplementry appendix 3), first, we did the preprocessing of the diagnosis text. All the punctuations (except question marks and short dashes which may indicate the uncertainty and negation) and blanks were converted to semicolons, and all the English letters were converted to single byte, uppercase ones. Then raw diagnosis text was cut into segments of single disease. Finally, the first five diagnoses were extracted and used for mapping the clinical terms in the LiSToD based on the REoD.

Initially, we tried to identify the tier 1 diagnoses from each of the first five diagnoses. For accurate extraction of the clinical information, modification information was further tried to detect from the identified tier 1 diagnoses, as the underlying meaning of the free diagnosis text is significantly affected by other co-occurring concepts.[19] We detected three kinds of modifiers: (1) negation modifiers, such as '排除' (except or exclude), '除外' (except or exclude), '阴性' (negative), '非' (non or not) and the short dash '-'; (2) temporal information, such as '复查' (re-examination), '复诊' (revisit); (3) uncertainty modifiers, which indicated that the event may not have actually occurred, such as '待排除' (to be excluded), '待查' (unknown origin or to be examined), '可能' (maybe or likely), '不能排除' (cannot be excluded), '咨询' (for consultation or consulting) and the question mark '?'. If negation modifiers were detected, the tier 1 diagnosis was changed to tier 3; while if the temporal or uncertainty modifiers were detected, the tier of the diagnosis remained unchanged with the addition of a marker indicating uncertainty to the diagnosis. Further, if no tier 1 diagnosis or only negative tier 1 diagnosis was identified, we tried to identify tier 2 diagnosis. Modification information was also further detected for classified tier 2 diagnosis. Negative tier 2 diagnosis was changed to tier 3 or uncertainty marker was added when the information of uncertainty was detected. If no tier 1 or tier 2 diagnosis, or only negative tier 1 or tier 2 diagnosis was identified, then the diagnosis was classified as tier 3.

After all the first five diagnoses were classified as tier 1, tier 2 or tier 3, with or without uncertainty, the tier-fashion method was applied to assign a single diagnosis to each visit. This means that, for multiple diagnoses in the same visit, priority was given to tier 1 diagnosis without uncertainty marker (tier 1A), followed by tier 1 diagnosis with uncertainty marker (tier 1B), then tier 2 diagnosis without uncertainty marker (tier 2A), then tier 2 diagnosis with uncertainty marker (tier 2B), and finally, tier 3 diagnosis. If multiple diagnoses from a single tier exist in the visit, the first-listed certain diagnosis was assigned.

All the procedures above have been encapsulated into a PL/SQL package, which is accessible in the online supplementry appendix 8.

Reference standard and validation

We selected a random sample of 10 000 outpatient visits that could be linked with diagnosis records during a 4-month period of 1 January 2018 to 30 April 2018 in the database. The sample data in this study had a good representativeness of the entire database (online supplementry appendix 9). Prescription review was used as the reference standard for classification validation against which the sensitivity, specificity, positive predictive value (PPV) and negative predictive value (NPV) for classifying different tiers of diagnosis were calculated. Two researchers who were trained with the classification scheme and blinded to drug exposure status independently reviewed the raw prescription and classified the diagnoses into tier 1, tier 2, or tier 3 according to the LiSToD, with or without uncertainty. Conflicting results of the two researchers were further discussed with a third clinician and the final tier of the diagnosis was determined.

Statistical analysis

We calculated the sensitivity, specificity, PPV and NPV for each tier of diagnosis and the Clopper-Pearson Exact method was used for calculating the 95% CIs. For tier X diagnosis (X was 1, 2, 3, or 1A, 1B, 2A, 2B), the sensitivity was the proportion of true tier X diagnosis that was classified as tier X by the RE algorithm; the specificity was the proportion of true non-tier X that was classified as non-tier X by the RE algorithm; the PPV was the proportion of tier X classified by the RE algorithm that was true tier X diagnosis; the NPV was the proportion of non-tier X classified by the RE algorithm that was true non-tier X diagnosis. The information extraction, and mapping the diagnosis text to the LiSToD using REoD were performed by using Oracle 11gR2 and PL/SQL developer V.11.0 (Oracle Corp., Redwood Shores, California, USA). Statistical analyses were performed by using SAS V.9.4 (SAS Institute).

Patient and public involvement

Patients and the public were not involved in the design or conduct of the study.

RESULTS

Out of the 10 000 outpatient visits, 91.1% occurred in outpatient clinics, and 92.9% visits occurred in tertiary-level hospitals. Further, 90.5% of the patients were adults and 51.0% were women. Most outpatient visits were from the Eastern and Western regions, which accounted for 54.8% and 30.2% of all the visits (table 1). One thousand and ninety-two visits (10.9%) ended with antibiotic prescription and 1417 individual antibiotics were prescribed. The most commonly used antibiotics in the dataset were cefdinir, azithromycin, levofloxacin, cefixime and moxifloxacin. Then, 9984 of the 10 000 visits could be linked to valid diagnoses, and the diagnoses of 16 visits were pure numbers or punctuation that did not contain useful information. Among the 10 000 visits, 66% had just one diagnosis, over 85% had less than 2 diagnoses, and only 2.2% had more than 5 diagnoses. Among all diagnosis records, the median of the length of diagnosis text was 6 (IQR 4–10) characters; 73.1% contained less than 10 characters.

Among the 9984 visits with valid diagnoses, 3.9% (n=390), 11.5% (n=1144) and 84.6% (n=8450) were, respectively, classified as tier 1, tier 2, and tier 3 diagnosis (table 2). The sensitivities of the RE algorithm were 98.2% (95% CI 96.4% to 99.3%), 98.4% (95% CI 97.6% to 99.1%) and 100.0% (95% CI 100.0% to 100.0%) for classifying tier 1, tier 2 and tier 3 diagnoses, respectively. The specificities were 100.0% (95% CI 100.0% to 100.0%), 100.0% (95% CI 99.9% to 100.0%) and 98.6% (95% CI 97.9% to 99.1%). The PPVs for classifying tier 1, tier 2 and tier 3 diagnoses were 100.0% (95% CI 99.1% to 100.0%), 99.7% (95% CI 99.2% to 99.9%) and 99.7% (95% CI 99.6% to 99.8%), respectively. The NPVs were 99.9% (95% CI 99.8% to 100.0%), 99.8% (95% CI 99.7% to 99.9%) and 100.0% (95% CI 99.8% to 100.0%), respectively, for tier 1, tier 2 and tier 3 diagnoses (table 3). In addition, the RE-based algorithm had sufficiently high accuracy to detect the diagnosis modifiers, with all sensitivities, specificities, PPVs and NPVs for classifying diagnoses with uncertainty (tier 1B and tier 2B diagnoses) approaching or exceeding 95%. Online supplementry appendix 10 gives more detailed results of all the 49 subcategories under different diagnosis tiers.

An analysis of the errors in precision was performed (table 4). In total, diagnoses from 30 visits were inaccurately classified, among these, 17 were because the infectious disease were written after the 5th diagnosis, 12 were due to inappropriate writing of the diagnoses (where single diagnosis was improperly split or multiple diagnoses were incorrectly concatenated together); the last inaccuracy was due to the use of traditional Chinese, which was not considered when constructing REs.

Table 1 Basic characteristics of the sample outpatient visits*

Subgroups	No of prescriptions	Percentage (%)
Type of patients		
Outpatient clinic	9105	91.1
Emergency department	895	9.0
Age group		
0–5	430	4.3
6–17	480	4.8
18–44	4180	41.8
45–64	3129	31.3
≥65	1742	17.4
Unknown	39	0.4
Gender of patients		
Female	5099	51.0
Male	4893	48.9
Unknown	8	0.1
Hospital level		
Second	711	7.1
Third	9289	92.9
Region of China†		
Eastern	5476	54.8
Central	542	5.4
Western	3021	30.2
North-eastern	961	9.6
Top five most used antibiotics		
Cefdinir	127	9.0
Azithromycin	126	8.9
Levofloxacin	111	7.8
Cefixime	103	7.3
Moxifloxacin	79	5.6
No of diagnoses		
Non-valid diagnosis‡	16	0.2
One diagnosis	6597	66.0
Two diagnoses	1921	19.2
Three diagnoses	753	7.5
Four diagnoses	330	3.3
Five diagnoses	163	1.6
>5 diagnoses	220	2.2
Length of diagnosis text§		
Non-valid diagnosis	16	0.2
1–4 characters	3645	36.5
5–9 characters	3662	36.6
10–14 characters	1477	14.8
15–19 characters	574	5.7

Continued

Table 1 Continued

Subgroups	No of prescriptions	Percentage (%)
≥20 characters	626	6.3

*For rounding reasons, the sum of the percentages of some subgroups may not be exactly equal to 100%.
†Regions of China were divided according the National Bureau of Statistics of China. http://www.stats.gov.cn/ztjc/zthd/sjtjr/dejtjkfr/tjkp/201106/t20110613_71947.htm
‡Diagnosis that contained only numbers, punctuation (eg, comma, semicolon, exclamation mark, etc), and other non-Chinese characters, told nothing about the indication for antibiotics and were defined as invalid diagnosis.
§Whitespace and punctuation were not counted.

DISCUSSION

We developed a rule-based approach with high validity which could be used for identifying whether the use of antibiotics in outpatient prescriptions is appropriate in the Chinese context. Our findings indicated that the sensitivities, specificities, PPVs and NPVs of the algorithm were all over 98% for classifying tier 1, tier 2 and tier 3 diagnosis according to whether antibiotics were indicated by using the previously proposed method.[4 5 11] To our knowledge, there is no current estimate of the appropriateness of outpatient antibiotic prescribing at the national level in China. This paper presents an approach to extract structured diagnosis information that represents the antibiotic usage to support subsequent pharmacoepidemiology studies, rather than depending on the manual review of prescriptions, which is very time consuming and involves high cost of labour. In addition, this approach can provide a method to evaluate antibiotic use for different categories of diseases by using large EMRs and administrative data.

The study is the first to use rule-based natural language processing (NLP) to establish the classification system for evaluating inappropriate prescribing of antibiotics using Chinese diagnosis text. NLP has been used for over 30 years to identify key clinical information from unstructured and semistructured text.[19–21] As the amount of EMRs and administrative data is increasing rapidly and most of the information is stored in unprocessed and heterogeneous textual formats, NLP plays a crucial role in this context to transform narrative medical text into structured data.[22 23] The main approaches to NLP are rule based,[24] machine learning[25] or hybrid approaches.[8 26] Machine learning can offer easier portability solutions, while rule-based methods tend to provide reliable results.[24] NLP has been used for extracting and standardising information on drugs,[9 27] for automatic detection of diseases,[28 29] and for deidentification of protected health information.[8] However, as most of the NLP researches have been performed in English, similar researches in Chinese are relatively limited.[30–32]

The primary reason for incorrect classification using our algorithm was that diagnoses of infectious diseases were not written or inputted in a priority order. In this study, 3 tier 1 and 11 tier 2 diagnoses were misclassified as tier 3 because the diagnoses of the infections were written after the fifth one. The Prescription Administrative Policy[33] and the Regulations on Prescription Review Management[34] issued by the Ministry of Health of China (now the National Health Commission) stipulate that no more than five kinds of drugs can be prescribed in the same prescription, and thus logically, if a clinician wants to prescribe antibiotics, he/she should mention the infectious diseases in the first five diagnoses; otherwise, the prescription will be considered as irrational. In addition, clinicians tend to write the primary diagnosis in a priority order and other diagnoses were written as complications or comorbidities. This was reflected in our research as nearly 90% of tier 1 and tier 2 diagnoses were in the first two diagnoses. Thus, these 14 misclassifications due to lack of prioritisation of infectious diagnosis writing may actually not be misclassified and if we assign single diagnosis to a visit based on more than 5 diagnoses, the risk of overdiagnosing may arise.

Another kind of errors that occurred was due to structural ambiguity of diagnosis writing. There were cases where the diagnosis text could be interpreted in contrary ways, since single diagnosis was improperly split, or multiple diagnoses belonging to different tiers were incorrectly concatenated together. This resulted in incorrect word segmentation. In the first case, parts of key information was cut into another independent diagnosis, which may contain the opposite meaning, resulting in mapping of incomplete information and misclassification. In the

Table 2 Confusion matrix for the regular expression (RE)-based diagnosis classification algorithm*

Classified by the RE algorithm	Classified by manual prescriptions review				
	Tier 1A	Tier 1B	Tier 2A	Tier 2B	Tier 3
Tier 1A	348	0	0	0	0
Tier 1B	1	41	0	0	0
Tier 2A	2	1	1074	0	0
Tier 2B	0	0	4	63	0
Tier 3	4	0	17	1	8428

*Tier 1A: tier 1 diagnoses without uncertainty. Tier 1B: tier 1 diagnoses with uncertainty. Tier 2A: tier 2 diagnoses without uncertainty. Tier 2B: tier 2 diagnoses with uncertainty.

Table 3 Validation of the regular expression-based classification algorithm

Diagnosis tiers*	Sensitivity (%, 95% CI)	Specificity (%, 95% CI)	PPV (%, 95% CI)	NPV (%, 95% CI)
Tier 1A	98.0 (96.0 to 99.2)	100.0 (100.0 to 100.0)	100.0 (98.9 to 100.0)	99.9 (99.9 to 100.0)
Tier 1B	97.6 (87.4 to 99.9)	100.0 (99.9 to 100.0)	97.6 (87.4 to 99.9)	100.0 (99.9 to 100.0)
Tier 2A	98.1 (97.1 to 98.8)	100.0 (99.9 to 100.0)	99.7 (99.2 to 99.9)	99.8 (99.6 to 99.9)
Tier 2B	98.4 (91.6 to 100.0)	100.0 (99.9 to 100.0)	94.0 (85.4 to 98.3)	100.0 (99.9 to 100.0)
Tier 1	98.2 (96.4 to 99.3)	100.0 (100.0 to 100.0)	100.0 (99.1 to 100.0)	99.9 (99.8 to 100.0)
Tier 2	98.4 (97.6 to 99.1)	100.0 (99.9 to 100.0)	99.7 (99.2 to 99.9)	99.8 (99.7 to 99.9)
Tier 3	100.0 (100.0 to 100.0)	98.6 (97.9 to 99.1)	99.7 (99.6 to 99.8)	100.0 (99.8 to 100.0)

*Tier 1A: tier 1 diagnoses without uncertainty. Tier 1B: tier 1 diagnoses with uncertainty. Tier 2A: tier 2 diagnoses without uncertainty. Tier 2B: tier 2 diagnoses with uncertainty.
NPV, negative predictive value; PPV, positive predictive value.

other case, since we performed the classification using a step-by-step process, in the first mapping process, the tier 1 or tier 2 diagnosis was detected, while in the second process involving detection of modifiers or lower tier diagnosis, negation was detected or another tier was mapped due to the coexistence of contrary information.

Typo and traditional Chinese created another category of errors that may led to misclassification but this was very rare in this study. As there are many homophonic and homomorphic words in Chinese language, especially homophonics, for example, '幽门' (Chinese pronunciation 'youmen'), which means pylorus in the diagnosis of *H. pylori*, may be inputted as '油门' (Chinese pronunciation 'youmen') which means accelerator and has nothing to do with disease. Thus, typos and traditional Chinese characters are likely to occur in a larger amount of diagnosis text, and it may depend on which type of input method is used by clinicians since Chinese characters can be inputted thorough pronunciation-based (the Chinese Pinyin) and glyph-based (the five stroke) input method, making this kind of error difficult for the rule-based method to detect.

Strengths and limitations

Our study had some strengths. The rule-based and tier-fashion algorithm that we used provided a feasible and validated method to evaluate the appropriateness of antibiotic prescribing by using big EMRs. It can process diagnosis extremely rapidly compared with manual prescription review, with less than 30 s needed to classify all the sample diagnoses. The algorithm was effective and had good extensibility and portability, as it is easy to add new REs or remove old ones from the REoD we developed. Since REs are well supported by most of the common database management system and statistical software, it is easy for other researchers to reuse our algorithm for conducting similar research using other kinds of data written in Chinese text.

However, there were limitations in our study. Building rule-based systems is often time consuming.[20] In this study, the first two steps for establishing the LiSToD and REoD took us several months. The validity of our method depended heavily on whether the LiSToD included all

Table 4 Reasons for inaccurate classifications of the regular expression (RE)-based algorithm

Classified by the RE algorithm*	Classified by manual prescription review*	Reasons for inaccurate classifications†
Tier 1B	Tier 1A	Multiple diagnoses incorrectly concatenated together (n=1).
Tier 2A	Tier 1A	The infectious disease written after the fifth diagnosis (n=2).
Tier 2A	Tier 1B	The infectious disease written after the fifth diagnosis (n=1).
Tier 2B	Tier 2A	Multiple diagnoses incorrectly concatenated together (n=4).
Tier 3	Tier 1A	Single diagnosis improperly split (n=1); The infectious disease written after the fifth diagnosis (n=3).
Tier 3	Tier 2A	Multiple diagnoses incorrectly concatenated together (n=5); Single diagnosis improperly split (n=2); The infectious disease written after the fifth diagnosis (n=10); Traditional Chinese used (n=1).
Tier 3	Tier 2B	The infectious disease written after the fifth diagnosis (n=1);

*Tier 1A: tier 1 diagnoses without uncertainty. Tier 1B: tier 1 diagnoses with uncertainty. Tier 2A: tier 2 diagnoses without uncertainty. Tier 2B: tier 2 diagnoses with uncertainty.
†There was a visit (classified as tier 3 by computer and tier 2A by manual review) in which one diagnosis was divided into two diagnoses and the second part was written together with another one; thus, the total number of incorrect classification for different reasons was 31 in this table.

the bacterial infections, and whether the REoD contained all possible patterns of diseases in the LiSToD. Medicine is a large and complex domain with rich synonyms and semantically similar and related concepts.[23] In addition, medical abbreviations and acronyms are common and can also be ambiguous, making it difficult to identify them.[20] Although we applied multiple strategies to make the list of synonyms, abbreviations and similar patterns of infectious disease as complete as possible, some variations of standard concepts in the ICD-10 may still be possible to not be found. Since we can only extract features through the manually constructed word list, some omissions may occur in a much larger diagnosis data.

CONCLUSIONS

In conclusion, to our knowledge, our study is the first to use the rule-based algorithm to establish the classification system for evaluating inappropriate prescribing of antibiotics using Chinese diagnosis text. Further studies focusing on antibiotics in China can apply this validated algorithm to evaluate the appropriateness of antibiotic use by using big EMRs or administrative data.

Author affiliations
[1]Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China
[2]Department of Pharmacology, Chinese PLA General Hospital, Beijing, China
[3]School of Pharmacy, University College London, London, UK
[4]Department of Healthcare-associated Infection Management and Disease Prevention and Control, Peking University First Hospital, Beijing, China
[5]Department of Pharmacy, Peking University Third Hospital, Beijing, China
[6]Department of Pulmonary and Critical Care Medicine, China-Japan Friendship Hospital, Beijing, China
[7]Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China

Contributors HZ and SZ developed the research question. HZ conducted the data analyses and drafted the manuscript. JB and MZ collected, cleaned and managed the data. SZ, JB and MZ are responsible for the integrity of the data. SZ, JB, LW and MZ reviewed and edited the manuscript. HZ, XY, YY and ZZ established the primary list of the standard tiers of diagnoses. JB, LL and BC reviewed the list of the standard tiers of diagnoses. HZ, XY and LZ constructed the regular expressions for mapping the diagnosis text. HZ, XY and MZ reviewed and evaluate the raw prescription for appropriateness of antibiotic prescribing. All authors read and approved the final manuscript.

Funding This study was supported by the National Natural Science Foundation of China (Grant numbers 81973146, 81473067 and 91646107).

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not required.

Ethics approval This study was approved by the Ethical Review Board of Peking University Health Science Centre (approval number: IRB00001052-18013-Exempt).

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement Data of prescriptions are available upon reasonable request. All other data relevant to the study are included in the article or uploaded as supplementary information.

ORCID iDs
Houyu Zhao http://orcid.org/0000-0002-4546-9342
Lin Zhuo http://orcid.org/0000-0001-9188-702X
Siyan Zhan http://orcid.org/0000-0001-7252-5349

REFERENCES

1 Cars O, Högberg LD, Murray M, *et al*. Meeting the challenge of antibiotic resistance. *BMJ* 2008;337:a1438.
2 Xiao Y. Antimicrobial stewardship in China: systems, actions and future strategies. *Clin Infect Dis* 2018;67:S135–41.
3 Zhang Q-Q, Ying G-G, Pan C-G, *et al*. Comprehensive evaluation of antibiotics emission and fate in the river basins of China: source analysis, multimedia modeling, and linkage to bacterial resistance. *Environ Sci Technol* 2015;49:6772–82.
4 Fleming-Dutra KE, Hersh AL, Shapiro DJ, *et al*. Prevalence of inappropriate antibiotic prescriptions among US ambulatory care visits, 2010-2011. *JAMA* 2016;315:1864.
5 Smith DRM, Dolk FCK, Pouwels KB, *et al*. Defining the appropriateness and inappropriateness of antibiotic prescribing in primary care. *J Antimicrob Chemother* 2018;73:ii11–18.
6 Roberts RR, Hota B, Ahmad I, *et al*. Hospital and societal costs of antimicrobial-resistant infections in a Chicago teaching hospital: implications for antibiotic stewardship. *Clin Infect Dis* 2009;49:1175–84.
7 European Centre for Disease Prevention and Control, European Medicines Agency. ECDC/EMEA joint technical report: the bacterial challenge: time to react, 2009. Available: https://www.ecdc.europa.eu/sites/portal/files/media/en/publications/Publications/0909_TER_The_Bacterial_Challenge_Time_to_React.pdf [Accessed 12 Jul 2018].
8 Du L, Xia C, Deng Z, *et al*. A machine learning based approach to identify protected health information in Chinese clinical text. *Int J Med Inform* 2018;116:24–32.
9 Zhang R, Chu F, Chen D, *et al*. A text structuring method for Chinese medical text based on temporal information. *Int J Environ Res Public Health* 2018;15:402.
10 Smieszek T, Pouwels KB, Dolk FCK, *et al*. Potential for reducing inappropriate antibiotic prescribing in English primary care. *J Antimicrob Chemother* 2018;73:ii36–43.
11 Chua K-P, Fischer MA, Linder JA. Appropriateness of outpatient antibiotic prescribing among privately insured us patients: ICD-10-CM based cross sectional study. *BMJ* 2019;364:k5092.
12 Holloway KA, Kotwani A, Batmanabane G, *et al*. Antibiotic use in South East Asia and policies to promote appropriate use: reports from country situational analyses. *BMJ* 2017;358:j2291.
13 WHO collaborating centre for drug statistics methodology. ATC/DDD index, 2017. Available: https://www.whocc.no/atc_ddd_index [Accessed 5 Jul 2017].
14 Classification and Code of Diseases. National Health and Family Planning Commission of the People's Republic of China, China National Standardization Administration. In: *GB/T 14396-2016*. Beijing: Standards Press of China, 2016. http://www.gb688.cn/bzgk/gb/newGbInfo?hcno=8127A7785CA677952F9DA062463CBC41
15 National Health and Family Planning Commission of the PRC. *Guidelines for clinical application of antibiotics (2015 edition*. Beijing: People's Medical Publishing House, 2015. http://www.nhc.gov.cn/ewebeditor/uploadfile/2015/09/20150928170007470.pdf
16 UpToDate. UpToDateSM clinical consultant. Available: https://www.uptodate.com/home [Accessed 12 Sep 2018].
17 Adriaenssens N, Coenen S, Tonkin-Crine S, *et al*. European surveillance of antimicrobial consumption (ESAC): disease-specific quality indicators for outpatient antibiotic prescribing. *BMJ Qual Saf* 2011;20:764–72.
18 Shapiro DJ, Hicks LA, Pavia AT, *et al*. Antibiotic prescribing for adults in ambulatory care in the USA, 2007-09. *J Antimicrob Chemother* 2014;69:234–40.
19 Friedman C, Shagina L, Lussier Y, *et al*. Automated encoding of clinical documents based on natural language processing. *J Am Med Inform Assoc* 2004;11:392–402.
20 Karystianis G, Sheppard T, Dixon WG, *et al*. Modelling and extraction of variability in free-text medication prescriptions from an anonymised primary care electronic medical record research database. *BMC Med Inform Decis Mak* 2016;16:18.
21 Hirschberg J, Manning CD. Advances in natural language processing. *Science* 2015;349:261–6.
22 Batet M, Sánchez D, Valls A. An ontology-based measure to compute semantic similarity in biomedicine. *J Biomed Inform* 2011;44:118–25.

23 Chen X, Xie H, Wang FL, *et al*. A bibliometric analysis of natural language processing in medical research. *BMC Med Inform Decis Mak* 2018;18:14.
24 Spasic I, Sarafraz F, Keane JA, *et al*. Medication information extraction with linguistic pattern matching and semantic rules. *J Am Med Inform Assoc* 2010;17:532–5.
25 Tang R, Ouyang L, Li C, *et al*. Machine learning to parse breast pathology reports in Chinese. *Breast Cancer Res Treat* 2018;169:243–50.
26 Pons E, Braun LMM, Hunink MGM, *et al*. Natural language processing in radiology: a systematic review. *Radiology* 2016;279:329–43.
27 Patrick J, Li M. High accuracy information extraction of medication information from clinical notes: 2009 i2b2 medication extraction challenge. *J Am Med Inform Assoc* 2010;17:524–7.
28 Zeng QT, Goryachev S, Weiss S, *et al*. Extracting principal diagnosis, co-morbidity and smoking status for asthma research: evaluation of a natural language processing system. *BMC Med Inform Decis Mak* 2006;6:30.
29 Wang Y, Luo J, Hao S, *et al*. Nlp based congestive heart failure case finding: a prospective analysis on statewide electronic medical records. *Int J Med Inform* 2015;84:1039–47.
30 Burger G, Abu-Hanna A, de Keizer N, *et al*. Natural language processing in pathology: a scoping review. *J Clin Pathol* 201610.1136/jclinpath-2016-203872. [Epub ahead of print: 22 Jul 2016].
31 Xu Y, Li N, Lu M, *et al*. Development and validation of method for defining conditions using Chinese electronic medical record. *BMC Med Inform Decis Mak* 2016;16:110.
32 Velupillai S, Mowery D, South BR, *et al*. Recent advances in clinical natural language processing in support of semantic analysis. *Yearb Med Inform* 2015;10:183–93.
33 Prescription Administrative Policy. Ministry of health of the PRC, 2007. Available: http://www.nhc.gov.cn/fzs/s3576/201808/d71d4735f6c842158d2757fbaa553b80.shtml
34 Regulations on prescription review management (TRAIL). Ministry of health of the PRC, 2010. Available: http://www.nhc.gov.cn/wjw/ywfw/201306/094ebc83dddc47b5a4a63ebde7224615.shtml