

# BMJ Open Quality appraisal of clinical practice guidelines for diabetes mellitus published in China between 2007 and 2017 using the AGREE II instrument

Yuting Gao,<sup>1,2</sup> Jinjing Wang,<sup>3</sup> Xufei Luo,<sup>1,4</sup> Xiaoyang Song,<sup>5</sup> Lian Liu,<sup>6</sup> Lixin Ke,<sup>6</sup> Zhihong Liao,<sup>2</sup> Dongke Wang,<sup>5</sup> Yiming Mu,<sup>7</sup> Yaolong Chen,<sup>1,8</sup> Janne Anton Markus Estill<sup>9,10</sup>

**To cite:** Gao Y, Wang J, Luo X, *et al.* Quality appraisal of clinical practice guidelines for diabetes mellitus published in China between 2007 and 2017 using the AGREE II instrument. *BMJ Open* 2019;**9**:e022392. doi:10.1136/bmjopen-2018-022392

► Prepublication history and additional material for this paper are available online. To view, please visit the journal (<http://dx.doi.org/10.1136/bmjopen-2018-022392>).

YG and JW contributed equally.

Received 8 March 2018

Revised 21 February 2019

Accepted 13 May 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

## Correspondence to

Dr. Yiming Mu;  
muyiming@301hospital.com.cn  
and Dr Yaolong Chen;  
chenyaolong21@163.com

## ABSTRACT

**Objective** The aim of this study was to systematically evaluate the quality of the clinical practice guidelines (CPGs) for diabetes mellitus published in China over the period of January 2007 to April 2017.

**Methods** We searched the China National Knowledge Infrastructure, Chinese Biomedical Literature database, VIP database and WanFang databases and guideline websites for CPGs for diabetes mellitus published between January 2007 and April 2017 in China. Two reviewers independently screened the literature according to the inclusion and exclusion criteria and extracted data. We used the the Appraisal of Guidelines for Research and Evaluation II (AGREE II) tool (Canadian Institutes of Health Research, Ottawa, Canada) to evaluate the quality of the included guidelines, calculated the scores of each domain and evaluated the consistency among the assessors via use of the intragroup correlation coefficient. And then we compared the results with Chinese CPGs and international CPGs. We conducted a subgroup analysis based on different classification criteria and compared scores of each domain subgroup analyses.

**Results** A total of 98 guidelines were identified. The correlation coefficient within the group was 0.93, suggesting that the consistency between the evaluators was good. The scores of the six domains of AGREE II were described in median (IQR) as follows: scope and purpose 53.7 (50.0–59.7), stakeholder involvement 31.5 (27.3–37.0), rigour of development 19.1 (15.3–22.2), clarity of presentation 59.3 (50.0–64.8), applicability 18.1 (13.9–25.7) and editorial independence 0.0 (0.0–0.0). The mean score in each domain of quality of Chinese diabetes CPGs was lower than that of CPGs published worldwide but higher than the mean score of Chinese guidelines of all topics. A funding source, the updated version, organisation and publishers of the guidelines and target fields are all the factors influencing the quality of CPGs to a certain degree.

**Conclusions** A large number of Chinese diabetes CPGs have been produced. Their quality remain unsatisfactorily low compared with CPGs worldwide, there is still room for improvement. Chinese guideline developers should pay more attention to the transparency of methodology, and use the AGREE II instrument to develop and report guidelines.

## Strengths and limitations of this study

- Covered all guidelines published on diabetes over the past one decade in China.
- It has the potential to contribute significantly in the upscaling guideline development
- Conducted a subgroup analysis based on different classification criteria and compared scores of each domain subgroup analyses.
- The AGREE II instrument only assesses the reporting of the different items and not the content validity of the recommendations.

## INTRODUCTION

Diabetes mellitus has become one of the leading causes of mortality and burden of disease worldwide, especially in China, which is now home to the largest number of diabetics worldwide.<sup>1</sup> The prevalence of diabetes mellitus among adults in China has increased substantially in recent decades, rising from 0.7% in 1980 to 10.9% in 2013.<sup>1–4</sup> During the recent years, the government and organisations have started to pay more attention to this chronic disease, and national clinical practice guidelines (CPGs) for diabetes mellitus are also increasingly produced and disseminated. CPGs are defined as ‘statements that include recommendations intended to optimise patient care, which are informed by a systematic review of evidence and an assessment of the benefits and harms of alternative care options’.<sup>5</sup> High-quality guidelines provide explicit recommendations for clinical practice, helping to manage health conditions and reduce the use of unnecessary, ineffective or harmful interventions.<sup>6</sup> CPGs for diabetes mellitus have been developed to help optimise the management of the condition and improve the quality, appropriateness and cost-effectiveness of patient care.<sup>7</sup>

However, the potential of CPGs to improve patient care and resource use largely depends on the rigour of their development as well as the dissemination and implementation strategies.<sup>89</sup> To the best of our knowledge, there has been no systematic evaluation of guidelines on diabetes mellitus in China. The growing number of guidelines may result in increasing variability and conflicts among guideline recommendations. China currently lacks capacity for evidence-based guideline development and coordination by a central agency. Most Chinese guideline users rely on recommendations developed by professional groups that lack demonstration of transparency (including conflicts of interest (COIs) management and evidence synthesis) and quality. In addition, misperceptions about the role of guidelines in assisting practitioners as opposed to providing rules requiring adherence, and a perception that traditional Chinese medicine (TCM) cannot be appropriately incorporated in guidelines are present.<sup>10</sup> Hence, we aimed to systematically review the existing Chinese guidelines for diabetes mellitus, assess their quality and eventually help Chinese guideline developers better follow methodological standards while developing guidelines in the future. However, the assessment of the content validity is not a part of this review as the instrument used in the review only assesses the reporting of the different items and not the clinical content. We will pay attention to the consistency of recommendations in different included guidelines loosely as well.

### Identification of guidelines

We conducted a computerised search of four major academic databases: Chinese Biomedical Literature database (<http://www.sinomed.ac.cn/>), WanFang database (Chinese Medicine Premier, <http://www.wanfangdata.com.cn/>), VIP (Chinese Journals Full-text database, <http://data.whlib.ac.cn/>) and China National Knowledge Infrastructure (<http://www.cnki.net/>). We used the same search strategy in all databases and restricted the search to guidelines published in China from January 2007 through April 2017. The search strategy included Chinese translations of terms such as 'guideline' and 'diabetes mellitus' (table 1). In addition, we searched guidelines for diabetes mellitus on several websites and search engines, including Google Scholar and Chinese Diabetes Mellitus Association (<http://www.zhtnbxh.org/>). If full text of guidelines could not be found, we contacted the author or the development agency. We also searched all articles published in *Chinese Journal of Evidence-Based Medicine*, *Chinese Journal of Diabetes Mellitus* and *Chinese Journal of Diabetes* before April 2017 manually. To identify guidelines published outside of indexed journals, our search of key databases was supplemented with a search of the grey literature. Grey literature were covered by systematically searching publications of all relevant diabetes societies and associations and other health organisations for CPGs that meet our inclusion criteria. We searched the PubMed specifically restricted to Chinese CPGs to retrieve the Chinese diabetes CPGs.

**Table 1** Search strategies

Search terms	Explanation
#1* 'zhinan' or 'zhiyin' or 'gongshi' or 'caoran'	'zhinan', 'zhiyin', 'gongshi' and 'caoran' mean 'guideline' in Chinese
#2 'tangniaobing' or 'xiaokezheng' or DM or NIDDM or IDDM or MODY or T2DM or T1DM	'tangniaobing' means diabetes mellitus in Chinese. 'xiaokezheng' is a term used by doctors of traditional Chinese medicine to refer to diabetes mellitus
#3 #1 and #2	

\*We used Zhinan as a subject heading in the Chinese Biomedical Literature database, and the terms 'zhinan', 'zhiyin', 'gongshi' and 'caoran' in the title of the paper in the other three full-text databases Wan Fang, VIP and China National Knowledge Infrastructure.

Old or variant versions of guidelines where a new version was available, and guidelines not originally developed in China (such as Chinese versions of foreign CPGs, and adapted versions of CPGs from other countries), were excluded.

### Screening and data extraction

All studies were independently reviewed for eligibility by two researchers (LK, LL). Disagreements were resolved by face-to-face discussion, or in case of persistent disagreement, by consultation with a third researcher (YG). We first screened the titles and abstracts, and in the next steps, the full texts of publications considered relevant. All included studies were imported in ENDNOTE. We first conducted one preliminary trial of data extraction. The extraction strategy was then discussed and agreed on by all researchers, and formal data collection was performed. Data extraction was done in Excel. For the validity and accuracy of data, the data collection was completed by two researchers independently. Disagreements were resolved by discussion.

### Evaluation of guidelines

AGREE II is an international, rigorously developed and validated instrument for evaluating CPGs.<sup>11 12</sup> It consists of 23 key items organised within 6 domains. Each item in a domain is given a score from 1 (strongly disagree) to 7 (strongly agree). The score for each domain is obtained by summing all the scores of the individual reviewers for all items in a domain and then standardising as follows: (obtained score–minimal possible score)/(maximal possible score–minimal possible score).<sup>13</sup> All authors who were involved in assessing the guidelines using AGREE II had taken a formal training on the AGREE Enterprise website. We initially conducted two rounds of pilot appraisals with a total of 10 guidelines and discussed discrepancies, ensuring that all reviewers came to an agreement in understanding each item of AGREE II. Each guideline was independently evaluated by four reviewers.

## Patient and public involvement

Patients were not involved in our study.

## Statistical analysis

We present the means and SD for AGREE II domain scores, and the number of cases and corresponding percentages for categorical variables. Data for each domain were analysed by Microsoft Excel 2013 and SPSS V.25.0. Intraclass correlation coefficients (ICCs) were used for testing inter-rater reliability among the four reviewers.<sup>14</sup> We started the formal appraisal only after ICC reached at least 0.8 in the pilot appraisals. We confirmed whether data were in accordance with normal distribution at the beginning with SPSS V.25.0. If it was normally distributed, we would use one-way analysis of variance to compute the p value between the differences of multi-groups. If it was not normally distributed, we described the data in median (IQR). And then we tested the statistical significance of differences among several subgroups ( $\geq 2$ ) using Kruskal-Wallis test. The independent sample t-test was used to compare the statistical significance of differences between two groups. All tests were two-sided, and p values  $< 0.05$  were considered to be statistically significant. Besides, we compared the scores of Chinese diabetes CPGs with those of Chinese CPGs<sup>15</sup> and CPGs published worldwide<sup>16</sup> of all topics.

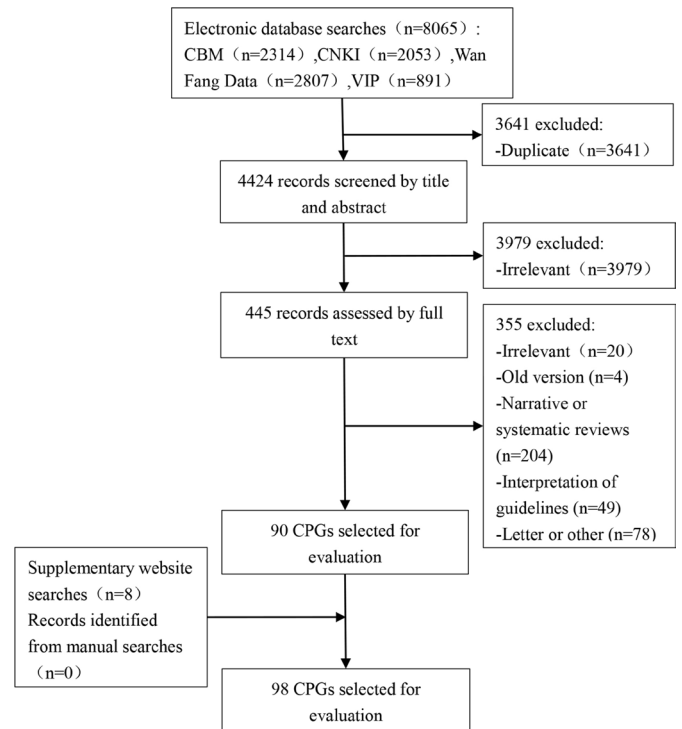
## RESULTS

### Literature search

Our systematic search yielded 8065 records, of which 3641 were excluded as duplicates. A total of 3979 records were excluded due to irrelevant topic after title and abstract screening, and 335 after reviewing full text. Twenty further guidelines were still excluded after evaluation. Eight guidelines were identified in the supplementary website search, one guideline was retrieved by PubMed search and no guidelines in the manual searches of the three journals. Finally, 98 eligible guidelines were included in our analysis (figure 1).

### Characteristics of the guidelines

The publication year of the included CPGs ranged between 2009 and 2017. The year with highest number of published diabetes CPGs (23) was 2011. Nine of the CPGs were updates of previous versions. The databases, exact terms and the time period of the search were accurately described in five guidelines. Seventeen guidelines targeted type 2 diabetes mellitus, two guidelines type 1 diabetes mellitus, two guidelines juvenile diabetes mellitus and one guideline latent autoimmune diabetes mellitus of adults. Some guidelines focused on complications only, including cardiovascular disease (five guidelines), diseases of the nervous system (five guidelines), kidney disease (five guidelines), hypertension (four guidelines), diabetic retinopathy (three guidelines) and diabetic foot (three guidelines). Seven guidelines reported that they received funding from governments, and two guidelines



**Figure 1** Flow chart of the literature search of Chinese clinical practice guidelines (CPGs) for diabetes published between 2007 and 2017. CBM, Chinese Biomedical Literature database; CNKI, China National Knowledge Infrastructure.

reported that they received funding from other academic organisations. Sixty guidelines were published in the Chinese Science Citation Database (CSCD), and one guideline online in Wiley Online Library (wileyonlinelibrary.com). Characteristics of the guidelines are displayed in online supplementary file 1.

### AGREE II results

The ICC for AGREE II assessment in the study was 0.929 (95% CI 0.910 to 0.942). It indicates that appraisers reached an agreement about the items in AGREE II. The scores of the included guidelines are summarised in table 2. The data were not a normal distribution by analysing in the test of normality with SPSS, so the data were described in median (IQR). However, the difference of mean score and median score was subtle, mean score could be used to analyse to some extent. Subgroup analysis across domains has been done.

### Scope and purpose

The median score in this domain was 53.7 (50.0–59.7). The median score was the second highest among the domains. Most guidelines performed well in this domain, with no guidelines scoring below 25%.

### Stakeholder involvement

The median score for the overall score for this domain was 31.5 (27.3–37.0). Twenty-eight guidelines referred to methodologists or evidence-based experts in the



**Table 2** AGREE II score of guidelines included

Domains	Content	Median score (IQR, %)	Mean score (X±SD, %)	Score segmentation (number of guidelines (%))			
				<25%	25%~50%	50%~75%	>75%
1	Scope and purpose	53.7 (50.0–59.7)	54.8±6.8	0 (0.0)	27 (27.6)	71 (72.4)	0 (0.0)
2	Stakeholder involvement	31.5 (27.3–37.0)	32.1±9.3	21 (21.5)	75 (76.5)	2 (2.0)	0 (0.0)
3	Rigour of development	19.1 (15.3–22.2)	20.2±10.8	81 (82.7)	15 (15.3)	2 (2.0)	0 (0.0)
4	Clarity of presentation	59.3 (50.0–64.8)	56.9±10.9	1 (1.0)	27 (27.6)	70 (71.4)	0 (0.0)
5	Applicability	18.1 (13.9–25.7)	19.2±8.7	74 (75.5)	24 (24.5)	0 (0.0)	0 (0.0)
6	Editorial independence	0.0 (0.0–0.0)	1.6±4.6	98 (100.0)	0 (0.0)	0 (0.0)	0 (0.0)

guideline development stage, and none involved patients in the development process. Only two CPGs scored at least 50%.

### Rigour of development

Two CPGs scored at least 50% in the domain ‘rigour of development’. The overall score in this domain was poor, with a median of 19.1 (15.3–22.2). Five CPGs described the systematic methods for searching and selecting evidence, 40 CPGs described the strengths and limitations of the body of evidence clearly, 29 CPGs used a nominal group technique or consensus-development conference, 81 CPGs considered the health benefits, side effects and risks in formulating the recommendations, 91 CPGs indicated a link between the supporting evidence and the recommendations and 4 CPGs were reviewed externally before publication. 9 guidelines were updated versions of older guidelines, but none of these described a procedure for updating the guideline.

### Clarity of presentation

Overall, the mean score for this domain was the highest among all domains, and only one guideline scored <25%. Recommendations were specific and unambiguous in all CPGs, although to various degrees. Four CPGs did not present the different options for management of the condition or health issue, and in one CPG key recommendations were not identifiable.

### Applicability

The median score for the overall quality score for this domain was 18.1 (13.9–25.7). 14 CPGs provided advice or tools on how the recommendations could be put into practice. Similarly, 14 CPGs described facilitators and barriers to its application. 14 CPGs considered the potential resource implications of applying the recommendations, and 87 CPGs presented monitoring or auditing criteria. 74 guidelines scored <25% for this domain.

### Editorial independence

The scores for this domain were the lowest among all domains, with all 98 guidelines scoring <25%. Only one guideline reported that there were no COIs. 9 CPGs reported that they received funding from governments (seven guidelines) or academic organisations (two guidelines). However, these 9 guidelines failed to report

whether the views of the funding body influenced the content of the guideline or not.

### Scores of AGREE II in each domain based on different classification criteria

We conducted a subgroup analysis based on different classification criteria and compared scores of each domain subgroup analyses. Table 3 shows the domain scores and p values for different subgroups of the guidelines. It is evident from table 3 that CPGs of integrated TCM and Western medicine performed clearly better than CPGs on Western medicine or TCM in all domains. Guidelines that reported external funding scored better than guidelines that did not report funding in editorial independence. Guidelines that involved methodologists scored higher than guidelines not reporting methodologists and the differences were statistically significant in ‘stakeholder involvement’ and ‘rigour of development’. To some extent, reporting methodologists could increase the scores in ‘stakeholder involvement’ and ‘rigour of development’. The difference between guidelines published in CSCD and non-CSCD journals was not statistically significant, developing organisation and type of guidelines as well. The results of the subgroup analysis showed that the updated version of the guidelines scored higher than the non-updated guidelines in all five fields (except for ‘field six: editorial independence’).

### Comparison of AGREE II scores between Chinese diabetes CPGs and CPGs published worldwide

Armstrong *et al*<sup>16</sup> used the AGREE II tool to evaluate 415 individual CPGs published between 1992 and 2014. The results showed that the average scores of the guidelines in the six AGREE II domains were 75.8%, 52.6%, 51.3%, 80.0%, 37.1% and 41.8%, respectively, as compared with the Chinese guidelines, which had a larger gap in all domains. Chen *et al*<sup>15</sup> used the AGREE II tools to evaluate 269 CPGs in China in all fields, with the results of 64%, 52%, 48%, 81%, 43% and 26%, respectively. The scores of Chinese CPGs on diabetes were clearly higher than those of Chinese CPGs of all topics on average. The scores of Chinese diabetes CPGs were substantially lower than the international average of CPGs on all topics.

**Table 3** Scores of AGREE II based on different classification criteria

Subanalysis	Number of guidelines (%)	Domain scores (mean±SD, %)					
		Scope and purpose	Stakeholder involvement	Rigour of development	Clarity of presentation	Applicability	Editorial independence
<b>Developing organisation</b>							
Chinese Medical Diabetes Association	60 (61.1%)	54.9±6.8	31.7±8.3	20.7±9.2	58.4±8.4	20.3±8.6	1.3±4.1
Specially formed group*	18 (18.4%)	55.1±8.0	32.8±13.9	19.8±15.8	56.2±12.1	17.9±10.9	2.6±6.5
National Health and Family Planning Commission	2 (2.1%)	45.4±9.2	27.8±2.6	13.5±17.2	41.7±32.7	13.9±0.0	0.0±0.0
Ministry of Public Health	1 (1.0%)	44.4	16.7	4.2	29.6	0.0	0.0
Hospitals	1 (1.0%)	59.3	31.5	21.5	59.3	16.7	0.0
Collaboration of two or more organisations	14 (14.3%)	56.1±5.5	33.5±5.3	19.7±9.8	56.0±12.0	18.2±7.2	0.8±3.0
Others	2 (2.1%)	52.8±3.9	39.8±11.8	25.0±7.9	53.7±10.5	18.8±6.9	8.3±11.8
P value		0.36	0.54	0.77	0.18	0.48	0.35
<b>Type of guidelines</b>							
Diagnosis and treatment	27 (27.6)	55.8±7.9	30.3±6.3	18.6±6.6	59.2±7.9	22.3±9.5	1.6±4.8
Treatment	22 (22.4%)	54.5±7.8	33.6±10.8	22.4±14.8	55.6±14.1	20.3±9.0	0.6±3.0
Management	14 (14.3%)	53.4±6.0	31.0±6.3	18.3±6.9	56.1±10.3	14.9±5.3	1.2±4.5
Prevention and treatment	9 (9.2%)	54.9±3.6	29.2±6.5	20.0±3.8	61.3±3.0	19.6±9.1	1.2±3.5
Technology†	6 (6.1%)	51.2±2.2	38.9±6.6	18.6±6.9	53.4±10.0	16.7±4.6	2.8±6.8
Medication‡	5 (5.1%)	56.7±7.9	34.8±7.6	22.5±7.5	58.1±11.2	19.2±9.1	0.0±0.0
Prevention	1 (1.0%)	68.5	37.0	24.3	64.8	29.2	0.0
Healthcare	1 (1.0%)	59.3	14.8	7.6	40.7	15.3	0.0
Comprehensive§	13 (13.3%)	54.4±6.6	33.0±14.8	22.2±17.9	54.1±14.6	15.8±8.9	4.1±6.5
P value		0.49	0.25	0.84	0.52	0.19	0.64
<b>Publication year</b>							
2009	3 (3.0%)	59.9±7.7	26.5±8.8	15.7±6.5	53.1±10.9	22.2±2.8	0.0±0.0
2010	7 (7.1%)	52.1±8.1	28.8±6.2	14.7±4.3	50.8±10.8	13.5±11.9	2.4±6.3
2011	23 (23.5%)	56.2±6.5	32.3±7.3	21.2±6.3	60.5±8.1	22.3±8.6	0.5±2.3
2012	13 (13.3%)	53.3±6.2	34.5±12.9	23.5±17.6	54.6±11.6	16.0±7.2	3.2±6.2
2013	10 (10.2%)	53.9±8.5	27.8±6.3	15.4±4.6	54.3±10.8	19.4±11.2	1.7±5.3
2014	12 (12.2%)	56.8±7.8	32.7±8.4	19.6±5.3	60.6±6.8	22.7±9.7	1.4±4.8
2015	9 (9.2%)	54.1±5.6	31.1±5.9	22.8±5.4	61.7±5.4	20.4±7.6	1.2±3.7
2016	14 (14.3%)	53.0±6.1	32.1±7.5	18.6±9.5	53.2±15.1	16.1±2.9	0.8±3.0
2017	7 (7.1%)	56.3±8.0	38.6±17.7	25.4±24.5	56.3±16.1	17.3±9.8	4.4±7.5
P value		0.58	0.38	0.42	0.20	0.13	0.59
<b>Western medicine or traditional Chinese medicine</b>							
Western medicine	78 (79.6%)	53.7±6.7	31.2±7.3	18.9±7.3	56.5±11.9	18.2±8.3	1.3±4.4
Traditional Chinese medicine	14 (14.3%)	58.6±5.9	31.1±7.5	19.9±4.8	58.7±5.6	21.8±9.7	1.6±4.0
Integrated medicine	6 (6.1%)	59.3±7.9	44.4±22.1	36.9±31.5	60.8±1.4	24.3±8.5	4.6±7.2
P value		0.01	0.002	0.001	0.53	0.11	0.23
<b>Participation of methodologists in guideline development</b>							
Methodologists involved	27 (27.6%)	55.6±7.2	36.9±11.4	25.2±16.8	58.3±10.0	19.8±6.6	1.4±4.2
Not reported	71 (72.4%)	54.6±6.9	30.2±7.6	18.3±6.6	56.4±11.3	19.0±9.4	1.6±4.7
P value		0.52	0.01	0.048	0.45	0.67	0.77
<b>External funding</b>							
No reported external funding	85 (86.7%)	55.2±6.6	32.3±8.3	19.7±9.6	56.6±11.4	19.7±8.3	0.5±2.6
External funding reported	13 (13.3%)	52.6±8.7	30.6±14.4	23.2±17.0	58.8±6.8	15.9±10.7	8.8±7.5

Continued

Table 3 Continued

Subanalysis	Number of guidelines (%)	Domain scores (mean±SD, %)					
		Scope and purpose	Stakeholder involvement	Rigour of development	Clarity of presentation	Applicability	Editorial independence
P value		0.19	0.54	0.25	0.31	0.13	0.001
Indexation in Chinese Sciences Citation Database (CSCD)							
Indexed in CSCD	60 (61.2%)	54.0±7.0	32.4±8.6	20.0±10.4	56.9±10.6	17.8±8.2	2.0±5.2
Not indexed in CSCD	37 (37.8%)	56.1±6.8	32.0±10.0	20.8±11.5	57.4±11.4	21.5±9.2	1.0±3.4
P value		0.14	0.65	0.82	0.98	0.46	0.27
Version							
Original	81 (82.7%)	54.8±7.3	31.9±9.3	19.9±11.5	56.2±0.1	18.9±0.1	1.9±5.0
Updated	17 (17.3%)	55.0±4.8	32.9±9.2	21.4±6.6	60.2±9.4	20.7±6.9	0.0±0.0
P value		0.14	0.43	0.38	0.30	0.20	0.00

\*Groups specifically formed for developing the guideline; members were staff from different hospitals.

†Refers to guidelines about the use of medical equipment, eg, insulin injection needle.

‡Refers to guidelines about the use of specific drugs, eg, insulin.

§Refers to guidelines covering more than two types of guidelines.

## DISCUSSION

### Chinese diabetes CPGs have an advantage in quantity this decade

It is astonishing that almost 100 domestic guidelines of diabetes mellitus have been published in China in just mere one decade. We found that annual number of CPGs for diabetes mellitus published in domestic Chinese journals increased dramatically up to 2011, after which the number has remained fairly stable during the last few years.

The increasing number indicates that more and more Chinese diabetes CPGs are being applied to clinical practice. Government agencies and the general and subspecialty medical societies in China promote the use of domestic guidelines in healthcare instead of merely adopting foreign guidelines, to account for the characteristics and needs of Chinese patients and clinicians. Meanwhile, the number of published randomised controlled trials and systematic reviews in China is increasing quickly, which may be associated with rapid increase in burden of diabetics. This has laid the groundwork for development of guidelines.

### Analysis of scores in each domain

An increasing amount of CPGs are being published. However, there is still much progress to be made as for quality of each domain. In the six major fields of the AGREE II tool scoring system, only the score of domain 1, 'scope and purpose' and domain 4 'clarity of presentation' were >50%; as such, the scores of the other four fields need to be improved upon. The low score of domain 2, 'stakeholder involvement', was because most of the guidelines included in the present study did not take into account the multidisciplinary intersections of the participants, as well as the preferences and values of the target population. The low score of domain 3, 'rigour of development', means that generally those guidelines

failed to show that they have conducted a systematic review on the best available evidences.<sup>17</sup> The low score was possibly due to the fact that the reporting on the approach to its development lacked transparency. Several of included guidelines did not report the methods of the systematic literature search, forming their recommendations and screening of evidence. Additionally, the low score of domain 5, 'applicability', suggested that most guidelines development agencies ignored the guidelines' application, and that most guidelines did not provide relevant supporting documents or recommendations or emphasise the promotion and hindrance factors in the application process. In fact, failure to address the issue of implementation and to provide clear, practical advice can have important consequences. The low score of domain 6, 'editorial independence', was likely due to the lack of disclosure of funding sources and related COIs. Only two of the included guidelines describes the methods by which potential COIs were sought and how COIs affected the recommendation development process. Failure to address the issue of implementation and to provide clear, practical advice can have important consequences. The Chinese diabetes guidelines' score was lower than that of international DR guidelines in this field, indicating that the Chinese Guidance Group has not paid enough attention to the interests of the members of the disclosure group and to declaring clearly the views of the sponsors. If a COI is not declared, the recommendations may be affected by multiple interests and biased. It may not be that the quality of the guidelines themselves are low but rather the quality of the non-standard report of the guidelines.<sup>13 18</sup> The low scores of these domains may be due to actual problems in guideline development, but because of the nature of the AGREE scoring, they may also simply be due to insufficient clarity and communication of the process used.<sup>19</sup>

### Quality of Chinese diabetes CPGs were comparatively low

The mean score in each domain of the quality of Chinese diabetes CPGs was tremendously lower than that of CPGs published worldwide. The difference can be at least partly explained by the selection criteria: the review by Armstrong *et al* included guidelines that are indexed in Medline and therefore likely to be of high quality, whereas our study tried to find all guidelines on a specific topic in one country. Besides, there was a huge gap between the Chinese diabetes CPGs and CPGs published worldwide in the domain 'editorial independence': this domain was also the one that had lowest score overall. COIs are a significant issue worldwide, as can introduce bias into almost every step of the guideline development process.<sup>20</sup> Our results show that there are serious reporting flaws for potential COIs for the members of the guideline development groups in Chinese CPGs. Most foreign and international organisations, such as WHO, have their own COI disclosure policies for members of guideline panels,<sup>21</sup> but until now, only few Chinese guideline-developing organisations have implemented such policies. Unlike in most other countries, in China the majority of guidelines are developed by professional committees, and despite reporting no external funding, many of these guidelines may be supported by pharmaceutical companies. It is critically important that guideline developers attend to declarations of potential COIs in a proactive, reasoned, transparent and defensible manner. An efficient way to decrease the influence of pharmaceutical companies could be to establish a government-controlled public foundation to develop guidelines. Armstrong *et al*<sup>16</sup> represented the quality of international CPGs, as compared with the Chinese guidelines, which had a larger gap in the 'rigour of development' and 'editorial independence' domains. The wide variability observed in these scores is worrying, since these domains directly reflect the reliability of the guidelines. Some of the included guidelines did not report the specific process of development and a COI as well. Holmer *et al*<sup>22</sup> used the AGREE II tools to evaluate 24 CPGs in the field of blood glucose management of type 2 diabetes mellitus, ultimately gaining results of 64%, 52%, 48%, 81%, 43% and 26%, respectively. Included guidelines in our study covers all contexts involved with diabetes mellitus, while Holmer *et al* only focused on blood glucose management of type 2 diabetes mellitus. That might be the reason why scores of guidelines in study of Holmer *et al* behaved much better than the Chinese diabetes CPGs. Anwer *et al*<sup>23</sup> selected seven evidence-based CPGs for management of type 2 diabetes mellitus in adults after a series of group discussion meetings, and then required results of 89.7%, 82.7%, 82.1%, 95.1%, 77.6% and 87.7%, respectively using the AGREE II tools. We can know that the quality of evidenced-based CPGs were much better to some extent. Chen *et al*<sup>15</sup> used the AGREE II tools to evaluate 269 CPGs in China in all fields. These results suggested that the Chinese guidelines still had some problems in terms of participation and application, mainly because of lacked

considerations about the participants in development of guidelines, and absence of the relevant report. Therefore, guidelines makers subsequently should consider the participation of multidisciplinary personnel and provide some guidance in application. Besides, several studies revealed that there are considerable variations and even conflicting recommendations concerning type 2 diabetes mellitus management from different guidelines.<sup>24 25</sup>

### Factors influencing the quality of CPGs

Many factors are capable of influencing the quality of CPGs. From the perspective of the funders, the quality of guidelines developed by organisations reporting a funding source was higher than those that did not report any funding. The scores of guidelines with methodologists taking part were higher than those not reporting methodologists. Evidence-based methods could improve the quality of CPGs, especially in stakeholder involvement and rigour of development. The guideline development groups involved methodological experts who could ensure that methodological checks were correctly applied and that the development process was fully documented.<sup>26</sup> Developing guidelines based on the methodology of evidence-based medicine could improve the quality of CPGs.<sup>27</sup> Yao *et al*<sup>28</sup> conducted a study comparing the quality of TCM CPGs with non-TCM CPGs, and the result showed that the quality of TCM CPGs was much better. The quality of TCM CPGs was higher than that of CPGs for Western medicine, which was compliant with the results of the study of Yao *et al*. It means that traditional medicine has drawn a lot of attraction and the relevant studies are becoming more scientific.

### Suggestions based on the results of the Chinese diabetes guidelines

First, evidence-based medicine emphasises focusing on the current optimal clinical medicine and clinicians' skills, as well as taking patients' preferences and values into consideration. Guideline development based on the methodology of evidence-based medicine could improve the quality of CPGs. The development of guidelines should incorporate multidisciplinary experts. Second, the guidelines can be registered during their development and should follow the contents of these to increase their transparency. Third, the promotion and impediment factors should be described in detail during the guidelines application process, while supporting tools for how the recommendations are to be applied in practice should be offered. Fourth, the development team members must also clearly state whether there are any potential COIs, and it is recommended that the editor of the medical journal ask the guidelines developers to provide a COI statement for all participating members as an attachment when delivering the guidelines.<sup>29</sup> Fifth, a central agency in China directing the development of evidence-based guideline is definitely required and a call for more rigorous training of guideline developers in China is warranted. As well as the quality of evidence like meta-analyses and randomized controlled trial needs



to be improved.<sup>30</sup> Sixth, a series of tools could be used to standardize the development of guidelines, such as that the guidelines report should follow the RIGHT report statement in order to ensure the transparency of the guidelines. Grading of Recommendations Assessment, Development and Evaluation curriculum vitae (GRADE CV) can help guideline developers assess a methodologist's expertise with methods and tasks.<sup>31</sup>

### Strengths and limitations

Our review has several strengths. First, the systematic review of CPGs for diabetes mellitus quality attempted to cover all guidelines published on diabetes over the past one decade in China. Our structured and explicit approach increases the validity of the findings. Second, we used the AGREE II instrument, which is a scientific and valid tool to assess the quality of CPGs. Furthermore, we conducted two rounds (for a total of 10 guidelines) of pilot appraisals and resolved disagreements, which further enhanced the confidence in our results. The extensive search strategy covering both indexed and grey literature, use of multiple appraisers who will complete training and calibration to assess the quality of CPGs and application of the AGREE II instrument, which has established validity and reliability, are all strengths of this review and may help the delivery of better care.

This review has also limitations. Our study was restricted to English and Chinese language guidelines, which excluded a small number of additional guidelines. Furthermore, CPG developers did not report all the details in developing guidelines even if they included some of the items listed in AGREE in process. Therefore, the score of AGREE II may underestimate the methodological quality of guidelines. Lastly, the AGREE II instrument only assesses the reporting of the different items and not the content validity of the recommendations. Currently, there is an ongoing research project, the AGREE-REX: Recommendation Excellence project that aims at developing and validating a new tool that will complement the AGREE II by assessing the clinical credibility and implementation of CPGs.<sup>32</sup> Current review covers the period from 2007 to 2017, however comparison was made, with 2010 and 2012 publication data. Clinical guideline development is a very dynamic process that progresses every year, so the comparison should just be a reference.

### CONCLUSION

A large number of CPGs for diabetes mellitus have been produced in China. Generally speaking, the quality of Chinese diabetes CPGs remain unsatisfactorily in comparison with other guidelines according to the evaluation by the AGREE II instrument. However, there is still room for improvement, especially in the aspect of editorial independence. Reporting the full texts of CPGs and COIs according to AGREE II checklists and abiding to the principles of evidence-based medicine can further improve the quality of guidelines.

### Author affiliations

<sup>1</sup>Evidence-Based Medicine Center, School of Basic Medical Sciences, Lanzhou University, Lanzhou, China

<sup>2</sup>The First Affiliated Hospital of Sun Yat-sen University, Department of Endocrinology and Metabolism, Guangzhou, Guangdong, China

<sup>3</sup>Department of Endocrinology, the Fifth Medical Centre, Chinese PLA General Hospital, 100071 (Former 307th Hospital of PLA), Beijing, China

<sup>4</sup>School of Public Health, Lanzhou University, Lanzhou, China

<sup>5</sup>The First Medical School of Lanzhou University, Lanzhou, China

<sup>6</sup>The Second Medical School of Lanzhou University, Lanzhou, China

<sup>7</sup>Department of Endocrinology, Chinese PLA General Hospital, Beijing, China

<sup>8</sup>Chinese GRADE Center, Lanzhou University, Lanzhou, China

<sup>9</sup>Institute of Global Health, University of Geneva, Geneva, Switzerland

<sup>10</sup>Institute of Mathematical Statistics and Actuarial Science, University of Bern, Bern, Switzerland

**Acknowledgements** The authors would like to thank Xiao YJ, Wang ZJ, Wang H and Tong YJ for their support in the screening, data extraction and appraisal of guidelines.

**Contributors** YC conceived the study idea and is the guarantor. JW, XL, YM and YG devised the study methodology. LL, LK and YG developed the search strategy and screened the guidelines, extracted the data and appraised the guidelines. XS and DW provided the methodological support. YG wrote the first draft. ZL, YM, YC and Janne read the drafts, provided the comments and agreed on the final version of the manuscript. XL, XS, ZL, YC and Janne contributed in improving the language of the article.

**Funding** The work was supported by the National Key R&D Program of China (2018YFC1705500); Beijing Municipal Science & Technology Commission (Project No. D141107005314004).

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data sharing statement** Supplementary data are available online at [https://pan.baidu.com/s/1KxyaEEqcSm\\_5Y0IWccuUIQ](https://pan.baidu.com/s/1KxyaEEqcSm_5Y0IWccuUIQ).

**Open access** This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

### REFERENCES

- Xu Y, Wang L, He J, *et al*. Prevalence and control of diabetes in Chinese adults. *JAMA* 2013;310:948.
- Martinovic AB, Zdravkovic A. [A mass survey of diabetes mellitus in a population of 300,000 in 14 provinces and municipalities in China (author's transl)]. *Zhonghua Nei Ke Za Zhi* 1981;20:678–83.
- Yang W, Lu J, Weng J, *et al*. Prevalence of diabetes among men and women in China. *N Engl J Med* 2010;362:1090–101.
- Wang L, Gao P, Zhang M, *et al*. Prevalence and ethnic pattern of diabetes and prediabetes in China in 2013. *JAMA* 2017;317:2515–23.
- Institute of Medicine (US) Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. Clinical practice guidelines we can trust[J]. *Port J Nephrol Hypert* 2013.
- Brouwers MC, Kho ME, Browman GP, *et al*. AGREE II: advancing guideline development, reporting and evaluation in health care. *J Clin Epidemiol* 2010;63:1308–11.
- Institute of Medicine (US) Committee to Advise the Public Health Service on Clinical Practice Guidelines. Clinical practice guidelines: directions for a new program. *Annals of Oncology* 1990;60:343–8.
- Grimshaw JM, Russell IT. Effect of clinical guidelines on medical practice: a systematic review of rigorous evaluations. *Lancet* 1993;342:1317–22.
- McAlister FA, van Diepen S, Padwal RS, *et al*. How evidence-based are the recommendations in evidence-based guidelines? *PLoS Med* 2007;4:e250.
- Yang K, Chen Y, Li Y, *et al*. Editorial: can China master the guideline challenge? *Health Res Policy Syst* 2013;11:1–3.
- Brouwers MC, Kho ME, Browman GP, *et al*. AGREE II: advancing guideline development, reporting and evaluation in health care. *CMAJ* 2010;18:E839–E842.



12. Burls A. AGREE II—improving the quality of clinical care. *Lancet* 2010;9747:1128–9.
13. AGREE Next Steps Consortium. *The AGREE II Instrument (Electronic version)*: AGREE Enterprise Website, 2009.
14. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979;86:420–8.
15. Chen YL, Yao L, Xiao XJ, *et al*. Quality assessment of clinical guidelines in China: 1993 - 2010. *Chin Med J* 2012;125:3660–4.
16. Armstrong JJ, Goldfarb AM, Instrum RS, *et al*. Improvement evident but still necessary in clinical practice guideline quality: a systematic review. *J Clin Epidemiol* 2017;81:13–21.
17. Graham R, Mancher M, Wolman D, *et al*. *Clinical practice guidelines we can trust*. Washington DC: Institute of Medicine, 2011.
18. Hu J, Chen R, Wu S, *et al*. The quality of clinical practice guidelines in China: a systematic assessment. *J Eval Clin Pract* 2013;19:961–7.
19. Vanommeslaeghe F, De Mulder E, Van de Bruaene C, *et al*. Selecting a strategy for prevention of contrast-induced nephropathy in clinical practice: an evaluation of different clinical practice guidelines using the AGREE tool. *Nephrol Dial Transplant* 2015;30:1300–6.
20. GPAC. *GPAC handbook-clinical practice guidelines and protocols by the guidelines and protocols advisory committee*. Vancouver: British Columbia, 2012:20–2.
21. World Health Organization. *WHO handbook for guideline development*: World Health Organization, 2014.
22. Holmer HK, Ogden LA, Burda BU, *et al*. Quality of clinical practice guidelines for glycemic control in type 2 diabetes mellitus. *PLoS One* 2013;8:e58625.
23. Anwer MA, Al-Fahed OB, Arif SI, *et al*. Quality assessment of recent evidence-based clinical practice guidelines for management of type 2 diabetes mellitus in adults using the AGREE II instrument. *J Eval Clin Pract* 2018;24:166–72.
24. Bennett WL, Odelola OA, Wilson LM, *et al*. Evaluation of guideline recommendations on oral medications for type 2 diabetes mellitus: a systematic review. *Ann Intern Med* 2012;156:27–36.
25. Stone MA, Wilkinson JC, Charpentier G, *et al*. Evaluation and comparison of guidelines for the management of people with type 2 diabetes from eight European countries. *Diabetes Res Clin Pract* 2010;87:252–60.
26. Robinson N, Liu J, Lee MS. Clinical guidelines: The way for best practice. *Eur J Integr Med* 2014;6:133–4.
27. Burgers JS, Grol R, Klazinga NS, *et al*. Towards evidence-based clinical practice: an international survey of 18 clinical guideline programs. *Int J Qual Health Care* 2003;15:31–45.
28. Yao L, Chen Y, Wang X, *et al*. Appraising the quality of clinical practice guidelines in traditional Chinese medicine using AGREE II instrument: A systematic review. *Int J Clin Pract* 2017;71:e12931.
29. Wang X, Chen Y, Yao L, *et al*. Reporting of declarations and conflicts of interest in WHO guidelines can be further improved. *J Clin Epidemiol* 2018;98:1–8.
30. Yao L, Sun R, Chen YL, *et al*. The quality of evidence in Chinese meta-analyses needs to be improved. *J Clin Epidemiol* 2016;74:73–9.
31. Norris SL, Meerpohl JJ, Akl EA, *et al*. Grading of Recommendations Assessment, Development and Evaluation (GRADE) Working Group. The skills and experience of GRADE methodologists can be assessed with a simple tool. *J Clin Epidemiol* 2016;79:150–8.
32. AGREE Research Projects. AGREE enterprise website. 2016 <https://www.agreetrust.org/> (Accessed 22 Nov 2016).