OPEN

# Methodological quality of clinical practice guidelines for genetic testing in children
## A systematic assessment using the appraisal of guidelines for research and evaluation II instrument

Xue-Feng Jiao, MD[a,b,c,d], Hai-Long Li, MD[a,b,c], Liang Cheng, MD[e], Chuan Zhang, MD[a,b,c], Chun-Song Yang, MD[a,b,c], Jonathan Han, BA[f], Qiu-Sha Yi, MS[a,b,c,d], Zhe Chen, MS[a,b,c], Li-Nan Zeng, MD[a,b,c], Ling-Li Zhang, MD[a,b,c,*]

## Abstract

Genetic testing of children is faced with numerous problems. High-quality clinical practice guidelines (CPGs) are needed to ensure its safe, and appropriate use. This study aimed to systematically identify the current CPGs for genetic testing in children, and to assess the methodological quality of these CPGs.

We searched 6 databases, 3 guideline clearinghouses, and 9 web sites of relevant academic agencies from inception to February 2019. CPGs focused on genetic testing in children were included. Four reviewers independently appraised the quality of the eligible CPGs using the appraisal of guidelines for research, and evaluation (AGREE) II instrument.

Seventeen CPGs meeting our inclusion criteria were included. Among them, 16 CPGs were focused on the genetic diagnosis/ evaluation of diseases, while only 1 CPG was focused on pharmacogenetics. The median domain scores from highest to lowest were: scope and purpose 80.56% (range: 56.95%–87.50%), clarity of presentation 72.22% (range: 45.83%–88.89%), stakeholder involvement 45.83% (range: 27.78%–55.56%), applicability 31.25% (range: 19.79%–54.17%), rigor of development 21.88%, (range: 13.02%–71.88%), and editorial independence 18.75% (range: 0%–83.33%). According to the overall quality, 6 (35%) CPGs were "not recommended," 8 (47%) CPGs were "recommended with modifications," and only 3 (18%) CPGs were "recommended." The clinical topics of the "recommended" CPGs were warfarin, familial Mediterranean fever, and pediatric pulmonary arterial hypertension.

The quality of CPGs for genetic testing in children was generally low, and variable across different CPGs and different AGREE II domains. In future guideline development, more attention should be paid to the aspects of stakeholder involvement, rigor of development, applicability, and editorial independence. Not only will guideline users benefit from our results when determining whether to adopt related CPGs to guide genetic testing in children, but guideline developers could also take into account our results to improve the quality of future CPGs.

**Abbreviations:** ACMG = the American College of Medical Genetics and Genomics, AGREE = appraisal of guidelines for research and evaluation, CPGs = clinical practice guidelines, FMF = familial Mediterranean fever, ICC = intraclass correlation coefficient, PAH = pediatric pulmonary arterial hypertension.

**Keywords:** AGREE II, children, clinical practice guidelines, genetic testing

# 1. Introduction

Genetic testing is a diagnostic technique that analyzes an individual's chromosomes, genes, or DNA to identify heritable disease-related mutations, genotypes, or karyotypes.[1] There are many kinds of genetic testing, including diagnostic testing, predictive, and presymptomatic testing, carrier testing, pharmacogenomics, prenatal diagnosis, newborn screening, and so on.[2] Since the mapping of the human genome in 2003, genetic testing and spending on the testing have increased rapidly. So far, there have been more than 75,000 genetic tests on the market, with approximately 10 new genetic tests entering the market daily.[3]

Although widely used, genetic testing of children is faced with numerous problems. One of the most critical issues is the informed consent to genetic testing. Due to the lack of decision-making ability in children, decisions about testing are often made by the parents, but must be driven by the child's best interest.[4,5] However, when faced with specific clinical conditions, the process of defining a child's "best interest" is usually complicated and controversial. Furthermore, there is limited evidence about the current benefits of genetic testing in children.[6] Due to developmental changes in gene expression, the genotype–phenotype relationships established in adults may not apply to children.[7]

Thus issues concerning the choice and context of children who should be tested as well as the choice of genes need to be explicitly addressed by high-quality and trustworthy clinical practice guidelines (CPGs). So far, a number of CPGs for pediatric genetic testing have been published, involving a variety of clinical topics. However, no study to date has evaluated the quality of these CPGs, which is a crucial emerging consideration in clinical practice. The implementation of poor-quality guideline recommendations may not only result in little or no medical benefit to children, but also cause a series of ethical, legal, and psychosocial problems.[4] For example, inappropriate genetic testing may result in psychological harms such as stigmatization, diffidence, confusion, guilt, and anxiety.[8] In addition, family relationships, and parental expectations of a child may also be influenced.[9]

The appraisal of guidelines for research, and evaluation (AGREE) II instrument is a widely validated and accepted tool used to assess the quality of CPGs for methodological rigor and transparency.[10] The instrument has been used to appraise the quality of CPGs in almost every clinical field, including CPGs for genetic testing.[11,12]

This study aimed to:

(1) systematically assess the quality of CPGs for genetic testing in children using AGREE II instrument and identify their quality to further improve future guideline development; and
(2) answer the following questions:

(1) how many CPGs are available for genetic testing in children?
(2) which CPGs are high-quality, and could be recommended?
(3) what is the content of the high-quality CPGs with regard to target disease/drug and recommendations?

# 2. Methods

## 2.1. Inclusion and exclusion criteria

We included CPGs focused on genetic testing in children (0–18 years old). We excluded CPGs that were:

(1) old versions of CPGs;
(2) duplications;
(3) not available in full text; or
(4) CPGs that did not make a specific recommendation for or against genetic testing.

Ethical approval and informed consent were not necessary, as no human beings were involved.

## 2.2. Data sources

We searched Pubmed, Embase (Ovid), PharmGKB, Guidelines International Network, U.S. National Guideline Clearinghouse, United Kingdom's National Institute for Health and Clinical Excellence, and 3 Chinese databases: China Knowledge Resource Integrated Database, VIP database, and Wanfang database for CPGs (until February 2019). The search terms included: genetic, genomic, pharmacogenetics, pharmacogenomics, pediatrics, pediatrics, newborn, infant, child, children, adolescent, minors, guideline, guidance, recommendation, consensus, and statement. We also searched CPGs at web sites of academic agencies, including the American Academy of Pediatrics, American Society of Human Genetics, the National Society of Genetic Counselors, Clinical Pharmacogenetics Implementation Consortium, and 5 Human Genetics Societies of different nations listed on the website of the International Federation of Human Genetics Societies. The references cited in published CPGs were considered if they met our inclusion criteria. The searches were limited to CPGs published in English or Chinese. Search results were selected for further review based on the inclusion and exclusion criteria.

## 2.3. Data extraction

Two reviewers selected CPGs independently and extracted: titles, publication years, countries, institutions, target populations, target diseases/drugs, detailed recommendations, quality of evidence, and strength of recommendations.

## 2.4. Guideline quality assessment

Four independent reviewers (XF Jiao, HL Li, C Zhang, and CS Yang) appraised the quality of each CPG using the AGREE II instrument. Among the reviewers, C Zhang and CS Yang have published studies about using AGREE II to appraise other types of CPGs and accumulated rich experiences in this field.[13,14] XF Jiao and HL Li were trained to use AGREE II through the online tutorial before this appraisal.[15]

AGREE II consists of 23 items organized into 6 domains: scope and purpose (domain 1), stakeholder involvement (domain 2), rigor of development (domain 3), clarity of presentation (domain 4), applicability (domain 5), and editorial independence (domain 6). The items were rated on a 7-point scale (1 = strongly disagree to 7 = strongly agree). The user's manual describes each item and helps users to determine a score for that item. Domain scores were calculated by summing all items scores proposed by all the 4 reviewers in a domain, then scaling the total as a percentage of the maximum possible score for that domain. The specific calculating formula was: (obtained score - minimum possible score)/ (maximum possible score - minimum possible score).[10]

The AGREE II user's manual does not provide cut-off scores for high/low quality CPGs. According to previous studies,[16] a CPG was "recommended" if no less than 3 domains (including domain 3 [rigor of development]) scored =60%. A CPG was "not recommended" if no less than 3 domains (including domain 3

[rigor of development]) scored ≤30%. A CPG was "recommended with modifications" in other cases.

### 2.5. Statistical analyses

Descriptive statistical analyses were performed for the scores of each domain and the overall quality of each CPG. Descriptive values included median, minimum, and maximum values. Agreement among the 4 reviewers was assessed by intraclass correlation coefficient (ICC). The ICC was calculated according to the scores from each reviewer. All analyses were performed by using MS Excel and SPSS Version 16.0.

## 3. Results

### 3.1. Guideline search and review process

A total of 4097 references were identified by the initial search. After selection, 17 CPGs[17–33] meeting our inclusion criteria were included, covering a period from 2001 to 2018 (Fig. 1).

### 3.2. Characteristics of the included CPGs (Table 1)

The 17 included CPGs covered a range of topics. Among them, 16 CPGs were focused on the genetic diagnosis/evaluation of diseases, while only 1 CPG[26] was focused on pharmacogenetics. Clinical topics included monogenic diabetes, uniparental disomy (n=2), hearing loss (n=2), developmental delay and mental retardation, fragile X syndrome, osteogenesis imperfecta, short stature, autism spectrum disorders, maturity-onset diabetes of the young, ataxias and spastic paraplegias, familial Mediterranean fever (FMF), pediatric pulmonary arterial hypertension (PAH), primary aldosteronism, cardiomyopathies, and warfarin. All the CPGs were developed by academic associations or societies, and more than half were from America (n=9). Six CPGs were developed specifically for children, while the others were for both children and adults. Eleven CPGs reported conflicts of interest. Only 5 (29%) CPGs[26,28–31] were evidence-based guidelines. Furthermore, there was a vast variation in the grading systems of evidence quality and recommendation strength. The number of cited references in each CPG ranged from 7 to 216 (median 41).
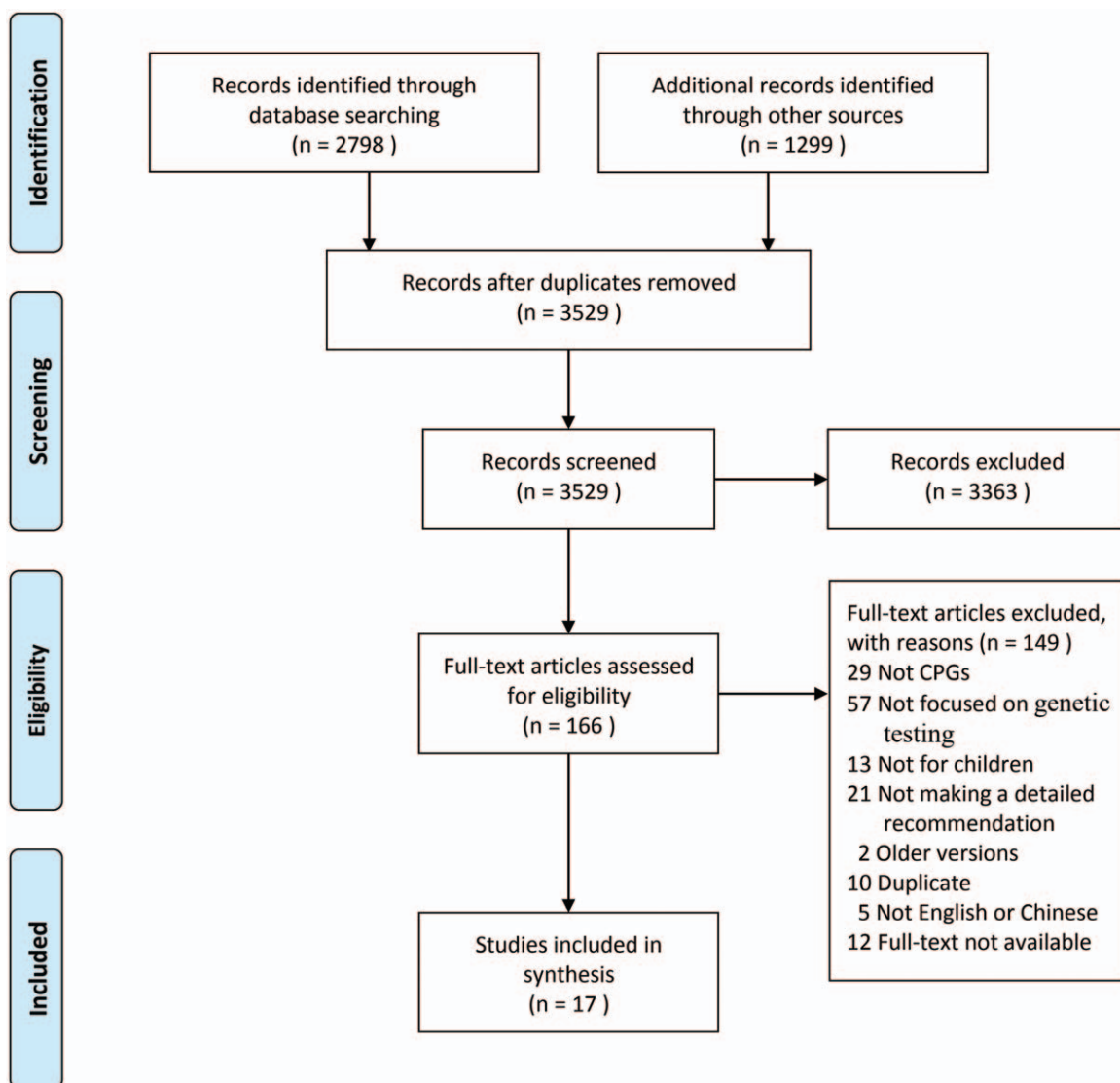


**Figure 1.** Flow diagram of guideline selection (PRISMA format).

**Table 1**

Characteristics of the 17 included clinical practice guidelines (CPGs).

| CPGs | Country | Institution | Target disease/drug | Target population | Conflicts of interest | Evidence-based guideline | Quality of evidence | Strength of recommendations | Numbers of references |
|---|---|---|---|---|---|---|---|---|---|
| Hattersley 2018[17] | International | ISPAD | Monogenic diabetes | Children | SCI | No | NA | NA | 216 |
| Shaffer 2001[18] | America | ACMG | Uniparental disomy | Fetuses, children, and adults | NA | No | NA | NA | 44 |
| Expert Panel 2002[19] | America | ACMG | Congenital hearing loss | Children | FPO | No | NA | NA | 31 |
| Shaffer 2005[20] | America | ACMG | Developmental delay, and mental retardation | Children | NA | No | NA | NA | 52 |
| Sherman 2005[21] | America | ACMG | Fragile X syndrome | Fetuses, children, and adults | NA | No | NA | NA | 19 |
| Byers 2006[22] | America | ACMG | Osteogenesis imperfecta | Fetuses, children, and adults | NA | No | NA | NA | 22 |
| Seaver 2009[23] | America | ACMG | Short stature | Children | NA | No | NA | NA | 7 |
| Schaefer 2013[24] | America | ACMG | Autism spectrum disorders | Children and adults | SCI | No | NA | NA | 76 |
| Alford 2014[25] | America | ACMG | Hearing loss | Children and adults | SCI | No | NA | NA | 114 |
| Johnson 2017[26] | America | CPIC | Warfarin | Children and adults | FPO | Yes | Self designed grading system | Self designed grading system | 46 |
| Ellard 2008[27] | European | EMQN | Maturity-onset diabetes of the young | Children and adults | FPO | No | NA | NA | 46 |
| Gasser 2010[28] | European | EFNS | Ataxias and spastic paraplegias | Children and adults | SCI | Yes | Grading system from ANN | Grading system from ANN | 36 |
| Giancane 2015[29] | European | SHARE | Familial mediterranean fever | Children and adults | FPO | Yes | Grading system from EULAR | Grading system from EULAR | 46 |
| Pattathu 2016[30] | European | EPPVDN | Pediatric pulmonary arterial hypertension | Children | FPO | Yes | Grading system from ESC and AHA | Grading system from ESC and AHA | 33 |
| Zennaro 2016[31] | France | SFE, SHFTA and AFCE | Primary aldosteronism | Children and adults | SCI | Yes | GRADE | GRADE | 27 |
| Dawson 2011[32] | Canada | CCMG | Uniparental disomy | Fetuses, children and adults | EI | No | NA | NA | 41 |
| CMA of pediatrics branch 2013[33] | China | CMA of pediatrics branch | Cardiomyopathies | Children | NA | No | NA | NA | 22 |

ACMG = the American College of Medical Genetics and Genomics, AFCE = Francophone Endocrine Surgery Association, AHA = the American Heart Association, ANN = the American Academy of Neurology, CMA = China Medical Association, CPGs = clinical practice guidelines; CCMG, the Canadian College of Medical Geneticists, CPIC = Clinical Pharmacogenetics Implementation Consortium, EFNS = the European Federation of the Neurological Societies, EI = editorial independence declare, EMQN = European Molecular Genetics Quality Network, EPPVDN = the European Pediatric Pulmonary Vascular Disease Network, ESC = the European Society of Cardiology, EULAR = the European League against Rheumatism, FPO = funding by external public organization reported, GRADE = grading of recommendations assessment, development, and evaluation, ISPAD = International Society for Pediatric and Adolescent Diabetes, NA = not available, SCI = statement about conflicts of interest of group members present, SFE = the French Endocrinology Society, SFHTA = the French Hypertension Society, SHARE = single hub and access point for pediatric rheumatology in Europe.

### 3.3. Comparison of the grading systems used in the 5 evidence-based CPGs (See Supplemental Digital Content, Table 1, http://links.lww.com/MD/D522. Comparison of the categorization of evidence, and recommendations in evidence-based CPGs)

Each of the 5 evidence-based CPGs adopted a different grading system from the others. There were vast differences among the 5 grading systems concerning the categorization of evidence, and recommendations.

### 3.4. Appraisal of the AGREE II domains (Table 2)

#### 3.4.1. Scope and purpose.
This domain evaluates the overall objectives, the health questions, and the target populations of CPGs. The median score for this domain was 80.56% (range: 56.95%–87.50%), which was the highest among the 6 domains. The overall objectives and health questions were well-described in all CPGs. However, the populations to whom the CPG was meant to apply were sometimes less detailed. For example, 2 CPGs[18,20] described the target population simply as "patients."

#### 3.4.2. Stakeholder involvement.
This domain evaluates the extent of professional group involvement, whether the views of the target populations are considered, and whether the target users are clearly defined. The overall score in this domain was low, with a median of 45.83% (range: 27.78%–55.56%). Most CPGs described the names, disciplines, institutions, and locations of guideline development group members, but only 1 CPG[30] described the member's role in guideline development. The extents of professional group involvement were not enough in all CPGs. For example, none of the CPGs included a methodology expert. No CPG stated that the views or preferences of the target populations were considered. The majority of CPGs offered clear descriptions of target users, for example, type of practitioner, specialty, while 3 CPGs[17,27,30] offered few details about target users.

#### 3.4.3. Rigor of development.
This domain relates to the methods of searching, grading, and synthesizing evidence, the process for formulating recommendations, and the procedure for updating them. The overall score in this domain was low, with a median of 21.88%, with great variation ranging from 13.02% to 71.88%. Only 3 CPGs[26,29,30] scored =60%. Only 5 CPGs[26,28–31] used systematic methods to search for evidence, and among them, 2 CPGs did not describe the search terms and search strategies.[28,31] Only 2 CPGs[26,29] explicitly described the criteria for including/excluding evidence, whereas 5 CPGs[18,21,22,24,25] offered no information about the criteria. Only 5 CPGs[26,28–31] used a system to grade the quality of evidence and the strength of recommendations. The methods used to formulate the recommendations varied: while 6 CPGs offered a great deal of details on how final decisions were arrived at,[26–31] the others offered minimal information if not none at all. Most CPGs considered both health benefits and harms when formulating their recommendations, whereas the method of balancing harms and benefits and how recommendations reflected this balance were usually not clearly reported. All CPGs provided a link between the recommendations and supporting evidence, but in some CPGs[19,24,32] the links were not easy to find. Only 5 CPGs[18,20,28,30,32] stated that they were externally reviewed before publication, while they just simply described the external reviewers with or without the methods taken to conduct the external review. Only 2 CPGs[26,28] declared that they would be updated periodically, and among them, 1 CPG[28] provided a time interval for updating.

#### 3.4.4. Clarity of presentation.
This domain assesses whether the recommendations are specific and unambiguous, whether the different management options are clearly presented, and whether key recommendations are easily identifiable. The median score for this domain was 72.22% (range: 45.83%–88.89%), which suggested that most CPGs met the criteria of this domain. All CPGs provided specific and precise recommendations. However, many did not describe the conditions or patients for whom the recommendations would not apply. Most CPGs clearly described different possible options for the management of a disease or condition. The key recommendations were easy to find in all CPGs except 2.[22,23]

#### 3.4.5. Applicability.
This domain assesses the consideration of facilitators or barriers to its application, as well as monitoring, and auditing criteria. The overall score in this domain was consistently low, with a median of 31.25% (range: 19.79%–54.17%). Most CPGs described the types of facilitators or barriers which would impact the implementation of guideline recommendations. Despite this, how these facilitators and barriers were sought and how they influenced guideline recommendations were often not described. Two CPGs[17,26] scored highly in providing advice and/or tools to facilitate application of the recommendations. Seven CPGs[17,24,26,27,29,30,33] considered the costs of genetic testing, but the information was less detailed. Most CPGs offered limited information about the monitoring and auditing criteria of guideline recommendations; for example, how the criteria should be measured was often not described.

#### 3.4.6. Editorial independence.
This domain addresses potential influences of the funding bodies and competing interests of the development members. The overall score in this domain was the lowest of all, with a median of 18.75%, with great variation ranging from 0% to 83.33%. Only 2 CPGs[26,30] scored ≥60%. Six CPGs[19,26–30] declared the names of funding bodies or sources of funding, while among them, only 2 CPGs[26,30] stated that the funding bodies did not influence the contents of CPGs. Six CPGs did not state the potential competing interests,[18–22,33] while remaining CPGs did so, none of them described how the competing interests were sought, or how they influenced the process of guideline development.

#### 3.4.7. Agreement among reviewers.
The ICC values for guideline appraisal using the AGREE II instrument ranged from 0.84 to 0.95, which indicated that overall agreement among the 4 reviewers was excellent for all CPGs.

### 3.5. Overall recommendation for use (Table 2)

The numbers of domains scoring ≥60% or scoring ≤30% were listed in Table 2. According to the recommended standard described previously, 6 (35%) CPGs[18–23] were "not recommended," 8 (47%) were "recommended with modifications," and only 3 (18%)[26,29,30] were "recommended." The 6 "not recommended" CPGs were all developed by The American College of Medical Genetics, and Genomics (ACMG), and their clinical topics were uniparental disomy, congential hearing loss, developmental delay and mental retardation, fragile X syndrome,

## Table 2

**Quality assessment of the 17 included clinical practice guidelines (CPGs) using the AGREE II instrument.**

| CPGs | Domain scores (%) | | | | | | ICC | Domains of scores ≥60% | Domains of scores ≤30% | Overall assessment |
|---|---|---|---|---|---|---|---|---|---|---|
| | Scope and purpose | Stakeholder involvement | Rigor of development | Clarity and presentation | Applicability | Editorial independence | | | | |
| Hattersley 2018 | 84.72 | 27.78 | 15.63 | 72.22 | 53.13 | 35.42 | 0.9 | 2 | 2 | RM |
| Shaffer 2001 | 56.94 | 45.83 | 13.02 | 63.89 | 19.79 | 12.50 | 0.9 | 1 | 3 | NR |
| Expert Panel 2002 | 69.44 | 38.89 | 19.27 | 68.06 | 28.13 | 20.83 | 0.9 | 2 | 3 | NR |
| Shaffer 2005 | 61.11 | 45.83 | 21.88 | 76.39 | 27.08 | 10.42 | 0.89 | 2 | 3 | NR |
| Sherman 2005 | 83.33 | 41.67 | 18.23 | 66.67 | 28.13 | 6.25 | 0.9 | 2 | 3 | NR |
| Byers 2006 | 59.72 | 38.89 | 14.06 | 45.83 | 20.83 | 0.00 | 0.95 | 0 | 3 | NR |
| Seaver 2009 | 75.00 | 45.83 | 13.54 | 54.17 | 27.08 | 18.75 | 0.94 | 1 | 3 | NR |
| Schaefer 2013 | 70.83 | 48.61 | 20.31 | 69.44 | 38.54 | 18.75 | 0.87 | 2 | 2 | RM |
| Alford 2014 | 87.50 | 44.44 | 16.67 | 63.89 | 31.25 | 37.50 | 0.91 | 2 | 1 | RM |
| Johnson 2017 | 77.78 | 48.61 | 71.88 | 80.56 | 54.17 | 83.33 | 0.84 | 4 | 0 | R |
| Elard 2008 | 83.33 | 37.50 | 22.92 | 72.22 | 40.63 | 45.83 | 0.9 | 2 | 1 | RM |
| Gasser 2010 | 83.33 | 55.56 | 57.81 | 72.22 | 26.04 | 41.67 | 0.85 | 2 | 1 | RM |
| Giancane 2015 | 86.11 | 50.00 | 63.54 | 87.50 | 50.00 | 43.75 | 0.88 | 3 | 0 | R |
| Pattathu 2016 | 66.67 | 36.11 | 61.98 | 84.72 | 35.42 | 77.08 | 0.89 | 4 | 0 | R |
| Zennaro 2016 | 83.33 | 50.00 | 45.83 | 88.89 | 33.33 | 18.75 | 0.91 | 2 | 1 | RM |
| Dawson 2011 | 80.56 | 52.78 | 30.73 | 77.78 | 29.17 | 16.67 | 0.9 | 2 | 2 | RM |
| CMA of pediatrics branch 2013 | 86.11 | 37.50 | 21.88 | 72.22 | 45.83 | 0.00 | 0.92 | 2 | 2 | RM |
| Median (range) | 80.56 (56.95–87.50) | 45.83 (27.78–55.56) | 21.88 (13.02–71.88) | 72.22 (45.83–88.89) | 31.25 (19.79–54.17) | 18.75 (0–83.33) | / | / | / | / |

CPGs = clinical practice guidelines, ICC = intraclass correlation coefficient, NR = not recommended, R = recommended, RM = recommended with modifications.

osteogenesis imperfecta, and short stature. The clinical topics of the 3 "recommended" CPGs were warfarin, FMF, and pediatric PAH. Their detailed recommendations are as follows:

### 3.5.1. Recommendations for warfarin dosing.

Warfarin is an extensively used oral anticoagulant with a narrow therapeutic range and large interindividual variability in its dose. To predict personalized warfarin dose more accurately, Clinical Pharmacogenetics Implementation Consortium (CPIC) made recommendations for warfarin dosing based on genetic information. For children of European ancestry, genetic testing of CYP2C9*2 and *3 and VKORC1-1639G>A genotype is recommended to guide warfarin dosing. For children of other ethnicities, genetic testing is not recommended due to lack of evidence.[26] Validated published pharmacogenetic algorithms for children are recommended to calculate warfarin dose.[34,35] Likewise, a pediatric warfarin dose calculator[33] is available at http://www.warfarindoserevision.com. The above recommendations were all based on pediatric data. The types of evidence underlying the above recommendations were cohort studies, case-control studies, cross-sectional studies, and case series. Moreover, the number, quality, and consistency of the individual studies were considered in grading the level of evidence.

### 3.5.2. Recommendations for FMF (see Supplemental Digital Content, Table 2, http://links.lww.com/MD/D523. Detailed recommendations for FMF).

FMF is a common monogenic autoinflammatory disease and generally has a childhood onset. To facilitate the diagnosis of children and young adults with FMF, single hub, and access point for pediatric rheumatology in Europe developed consensus recommendations for the genetic diagnosis of FMF. Genetic testing of MEFV mutations can support the clinical diagnosis of FMF. Among the known sequence variants of MEFV, M694V mutation or mutations at position 680 to 694 on exon 10 support the diagnosis, while the E148Q variant in exon 2 does not support. What's more, consultation with an autoinflammatory specialist is recommended in the indication and interpretation of genetic testing.[29] The types of evidence underlying the above recommendations were meta-analysis of cohort studies, cohort studies, case-control studies, and noncomparative descriptive studies.

### 3.5.3. Recommendations for pediatric PAH (See Supplemental Digital Content, Table 3, http://links.lww.com/MD/D524. Detailed recommendations for pediatric PAH).

PAH is a complex and multifactorial disease, with poor information about the natural history of the disease. To optimise the diagnosis, treatment, and prognosis of pediatric patients, the European pediatric pulmonary vascular disease network developed consensus recommendations for the genetic diagnosis of pediatric PAH. Genetic testing of PAH-associated genes such as ACVRL1, BMPR2, CAV1, KCNK3, and ENG is recommended for children with PAH (including hereditary PAH, idiopathic PAH, asymptomatic PAH, and "out of proportion" PAH) and their first-degree relatives. Moreover, for children with suspicion of pulmonary veno-occlusive disease, genetic testing of EIF2AK4 gene is recommended. For the genetic testing technologies, comprehensive next generation sequencing panels targeting all known PAH genes is recommended first. If this is not available, testing should move to PAH-associated genes with gene-specific direct sequencing technologies.[30] The above recommendations were all based on pediatric data. The types of evidence were large nonrandomized studies, cohort studies, case-control studies, cross-sectional studies, case series, and consensus of expert opinions.

## 4. Discussions

Genetic testing in children is a topic full of disputes. Over the past 3 decades, numerous publications have discussed its medical benefits and potential harms.[36,37] However, little attention has been paid to the CPGs in this field. This study is, to our knowledge, the first to systematically evaluate the quality of CPGs for genetic testing in children using the AGREE II instrument.

Our study identified only 17 CPGs published in 2001 or later which were focused on genetic testing in children. In many fields, such as cancer genetics, there are no specific CPGs for children, although in practice the indications are usually determined in individual level.[38] The small number of available CPGs is a reflection of the lack of genetic testing studies in children, which may be due to unsureness of ethical and legal issues of involving children in scientific studies.[39] The included CPGs paid more attention to the genetic diagnosis/evaluation of hereditary diseases. There were few commonalities among these CPGs because they focused on different clinical topics. Moreover, these CPGs commonly had enormous limitations, which was reflected by the fact that only 5 CPGs were considered to be evidence-based CPGs. All these evidence-based CPGs used different grading systems to evaluate the quality of evidence and strength of recommendations. The variation in terms of grading system may confuse both the readers and the future CPG developers. Furthermore, there existed some drawbacks in some of these grading systems. For example, the grading system from the European league against rheumatism failed to consider the consistency of results among studies, and the grading system from the European league against rheumatism, and the European society of cardiology (American heart association) lacked a strong correlation between quality of evidence, and strength of recommendations. Thus, we suggest future pediatric genetic testing CPGs use a uniform grading system with little drawbacks to evaluate the quality of evidence and strength of recommendations.

The quality of current CPGs indicated that there was a great scoring variability among different CPGs and across different domains of AGREE II. The median domain scores from highest to lowest were scope and purpose (80.56%), clarity of presentation (72.22%), stakeholder involvement (45.83%), applicability (31.25%), rigor of development (21.88%) and, editorial independence (18.75%). Previous research had adopted a score of 60% as a criterion for high quality.[40,41] In our study, the "scope and purpose" and "clarity of presentation" domains met this criterion, which were similar to previous assessments of CPGs in the field of genetic testing.[11,12] High scores for "scope and purpose" indicated that the objectives, health questions, and target populations were clearly defined in the current pediatric genetic testing CPGs. Likewise, high scores for "clarity of presentation" implied that the recommendations in these CPGs were clearly presented, which was especially important when considering the target users were often healthcare providers with little or no training experience in genetic testing. On the other hand, the domains "stakeholder involvement," "rigor of development," "applicability," and "editorial independence" had quite low scores, which were not completely consistent with

previous evaluations of genetic testing CPGs. For example, in a prior review of pharmacogenomics CPGs, both the "rigor of development" and "editorial independence" domains scored higher than 60% and rated as high-quality.[11] As identified in our evaluation, the main reasons for the low-quality scores were as follows:

(1) The most serious problem was the failure to implement the methods of evidence-based practice into guideline development. To ensure recommendations are based on the best available evidence, systematic methods should be used to search for evidence, and a consolidated and validated grading system should be adopted to evaluate quality of evidence and strength of recommendations. What's more, the details of search strategy, the criteria for including/excluding evidence, as well as the strengths and limitations of the evidence should be clearly described in CPGs.[10] However, most included CPGs did not perform well in the above aspects.

(2) The methods for formulating the recommendations were not described in the majority of the included CPGs. As the evidence of genetic testing in children is limited, it is important to provide how final recommendations are arrived at while ensuring minimum bias.

(3) Some CPGs ignored the ethical, policy, and psychosocial issues in genetic testing of children when formulating the recommendations. A series of ethical, policy, and psychosocial problems exist in genetic testing of children. For example, genetic testing may label a child as "at-risk" and cause stigmatization and discrimination, particularly in the insurance sector.[4] Even worse, genetic testing may cause children's psychological harms such as loss of self-esteem, confusion, guilt, or anxiety.[8] Moreover, genetic testing may also impact the children's families, such as causing parents' anxiety and guilt, or influencing family relationships.[9,36] Thus, the guideline recommendations should reflect the balance between the medical benefits and potential harms.

(4) Most CPGs had not been externally reviewed by experts before their publication. The external review of a CPG could improve its quality, evaluate its applicability and feasibility, and promote its dissemination.[10]

(5) Most CPGs failed to provide procedures for updating the CPG. Since new evidence may alter guideline recommendations, it is generally recommended to update the CPG at least every 3 years.[42]

(6) None of the CPGs included methodological experts in the guideline development groups. Methodological experts could ensure the methodological tools are correctly used and the development process is rigorous.

(7) Almost all CPGs failed to consider patients'/public's preferences and views. There are many methods to seek patients'/public's opinions. For instance, participation of patients/public in the guideline development group or the external review panel, formal interviews with patients/public, or literature reviews of patients'/public's expectations.[10] However, these strategies were often not implemented or not described in the included CPGs.

(8) Most CPGs did not pay enough attention to their applicability. The scarce feasibility or difficult implementation of recommendations are serious problems that could hinder the maximal use of CPGs.[43] The barriers to implementing recommendations should be identified and possibilities for overcoming them should be considered in the development of

CPGs. Moreover, CPGs should be implemented and disseminated with effective strategies such as guideline summary documents, educational tools, quick reference guides, or pilot testing among users. The monitoring criteria for evaluating guideline implementation is also necessary to consider. It is important to note that there are various genetic testing technologies with respectively different advantages and disadvantages,[20,29] and the costs of genetic testing are relatively high.[11,12] Thus, developers should also consider the technology availability and cost factors of genetic testing when developing CPGs. However, although important, these factors were often performed poorly in the included CPGs.

(9) The last problem was editorial independence. Editorial independence contains 2 aspects: first, the contents of CPGs should not be influenced by the funding body; second, the potential competing interests of guideline development team members should be clearly stated.[10] The low score for "editorial independence" did not necessarily mean that most CPGs failed to consider editorial independence, but rather that the description of this subject was not explicit and comprehensive. For example, although some CPGs stated competing interests, how the competing interests were sought and how they influenced the process of guideline development were often not described. Editorial independence has also been reported as opportunities for improvement in the previous assessments of other types of CPGs.[14]

So far, the AGREE II instrument has not provided a clear distinction between high-quality and low-quality CPGs. Thus, the criterion for the overall guideline quality was often self-defined and varied in different studies.[16] For example, some studies calculated the overall guideline quality as the mean of the 6 domain scores, and set specific cut-offs to differentiate high-quality from low-quality.[44] However, with this criterion, each domain had an equally impact on the overall guideline quality, which was not scientific enough.[45] According to Hoffmann-Eßer et al (2017), domain 3 (rigor of development) has the strongest influence on the overall guideline quality.[16] High score for this domain translates to an evidence-based guideline development with minimum bias,[45] while low score for this domain indicates that serious methodological problems existing.[46] Therefore, in this study we set the following criterion: if no less than 3 domains (including domain 3 [rigor of development]) scored ≥60%, then the CPG was rated as "high-quality" and was "recommended"; inversely, if no less than 3 domains (including domain 3 [rigor of development]) scored ≤30%, the CPG was rated as "low-quality" and was "not recommended." This criterion for the overall guideline quality is similar to those used in some previous studies.[40,45]

According to this criterion, the overall quality of pediatric genetic testing CPGs was suboptimal. Six CPGs were rated as "low quality" and were "not recommended," whereas only 3 CPGs were rated as "high-quality" and were "recommended." The 6 "not recommended" CPGs were all developed by ACMG and were all published in 2009 or prior. As the AGREE Instrument was first published in 2003 and was refined into the AGREE II version in 2009,[10] 1 explanation for the low-quality might be that the AGREE Instrument was not published or was not widely accepted when these CPGs were developed. Since low-quality CPGs may be harmful for children, it is necessary for ACMG to update these CPGs using the AGREE II Instrument. The 3 "recommended" CPGs were focused on warfarin, FMF,

and pediatric PAH respectively. None of the recommendations for children in these CPGs were deduced from adult data, which indicated that relatively sufficient pediatric evidence existed in these areas. The types of evidence in these CPGs were usually observational studies, such as cohort studies, case control studies, cross-sectional studies, and case series. This would be acceptable, because a randomized clinical trial is always infeasible to answer etiological questions or diagnostic questions.[47] However, when evaluating the quality of evidence, 2 of the 3 "recommended" CPGs[29,30] failed to consider the study methodology limitations (such as sampling, blinding, primary and secondary outcomes, or analytical methods), which should be enhanced in future guideline development. What's more, the genetic testing CPG for warfarin provided different recommendations for children of different races.[26] This strategy should be referenced to future genetic testing CPGs, because the distributions of many polymorphic genes are influenced by ethnicity.[48] In addition, the genetic testing CPG for pediatric PAH recommended specific genetic testing technologies. This is also important because the choice of suitable genetic testing technologies often confuses the guideline users.[30]

### 4.1. Implications for future guideline development

Our research found that there existed some methodological flaws in the current pediatric genetic testing CPGs. First, future CPG developers should direct more attention to develop evidence-based CPGs. Second, future CPG developers should pay more attention to external review, guideline updating, strengthening cooperation with methodological experts, seeking patients'/public's' views, considering the applicability of CPGs, funding issues, and conflicts of interest. Third, specific issues regarding genetic testing of children should be fully considered in the future CPGs, which includes ethical, policy and psychosocial problems, the technology available, costs, and the racial difference in gene polymorphisms. Finally, future CPG developers should improve compliance with the AGREE II Instrument in the guideline development process.

### 4.2. Limitations

Our research also has some limitations. First, we only included English or Chinese language CPGs, which might have led to the exclusion of relevant CPGs published in other languages. Second, we only included CPGs which were mainly focused on genetic testing in children. Thus the diagnosis and treatment CPGs which contained only a small amount of pediatric genetic testing information were excluded in our study, meaning we may have not identified all CPGs involving genetic testing in children. Third, the AGREE II instrument could only assess the methodological and reporting quality of CPGs,[10] and could not assess the content validity of guideline recommendations, which was also an important influencing factor of the overall guideline quality.

## 5. Conclusions

The quality of CPGs for genetic testing in children was generally low, and variable across different CPGs and different AGREE II domains. The quality of current CPGs was acceptable in the aspects of scope and purpose, and clarity of presentation. However, future guideline developers should pay more attention to the aspects of stakeholder involvement, rigor of development, applicability, and editorial independence. High-quality CPGs in this field were scarce, with only 3 CPGs were recommended for use. Six CPGs were not recommended due to their low-quality. There is scope, in numerous aspects, for improving the quality of current CPGs. Not only will guideline users benefit from our results when determining whether to adopt related CPGs to guide genetic testing in children, but guideline developers could also take into account our results to improve the quality of future CPGs in this field.

## Author contributions

**Conceptualization:** Xue-Feng Jiao, Ling-Li Zhang.
**Data curation:** Xue-Feng Jiao, Hai-Long Li.
**Formal analysis:** Xue-Feng Jiao, Hai-Long Li, Chuan Zhang, Chun-Song Yang.
**Funding acquisition:** Ling-Li Zhang.
**Investigation:** Xue-Feng Jiao, Hai-Long Li, Chuan Zhang, Chun-Song Yang.
**Methodology:** Xue-Feng Jiao, Hai-Long Li, Chun-Song Yang, Li-Nan Zeng.
**Project administration:** Liang Cheng, Qiu-Sha Yi, Zhe Chen.
**Resources:** Li-Nan Zeng, Ling-Li Zhang.
**Supervision:** Ling-Li Zhang.
**Validation:** Xue-Feng Jiao, Liang Cheng, Qiu-Sha Yi.
**Writing – original draft:** Xue-Feng Jiao.
**Writing – review and editing:** Hai-Long Li, Jonathan Han, Ling-Li Zhang.
Ling-Li Zhang orcid: 0000-0001-7339-2197.

## References

[1] National Academies of Sciences, Engineering, and MedicineAn Evidence Framework for Genetic Testing. Washington, DC: The National Academies Press; 2017.
[2] Committee on Genetics, American Academy of PediatricsMolecular genetic testing in pediatric practice: a subject review. Pediatrics 2000;106:1494–7.
[3] Phillips KA, Deverka PA, Hooker GW, et al. Genetic test availability, and spending: where are we now? Where are we going? Health Aff (Millwood) 2018;37:710–6.
[4] Botkin JR, Belmont JW, Berg JS, et al. Points to consider: ethical, legal, and psychosocial implications of genetic testing in children, and adolescents. Am J Hum Genet 2015;97:6–21.
[5] Geneticists, health professionals suggest recasting requests to test children for adult onset diseases: new study explores parents' reasons for seeking predictive genetic, testing. Am J Med Genet A 2017;173:8–9.
[6] Duncan RE, Savulescu J, Gillam L, et al. An international survey of predictive genetic testing in children for adult onset conditions. Genet Med 2005;7:390–6.
[7] Adam de Beaumais T, Jacqz-Aigrain E. Pharmacogenetics: applications to pediatric patients. Adv Pharmacol 2018;83:191–215.
[8] Malpas PJ. Predictive genetic testing of children for adult-onset diseases and psychological harm. J Med Ethics 2008;34:275–8.
[9] Lim Q, McGill BC, Quinn VF, et al. Parents attitudes toward genetic testing of children for health conditions: a systematic review. Clin Genet 2017;92:569–78.

[10] AGREE Next Steps Consortium (2009). The AGREE II Instrument [electronic version]. (2017) www.agreetrust.org

[11] Beckett RD, Kisor DF, Smith T, et al. Systematic evaluation of clinical practice guidelines for pharmacogenomics. Pharmacogenomics 2018;19: 693–700.

[12] Simone B, De Feo E, Nicolotti N, et al. Methodological quality of English language genetic guidelines on hereditary breast-cancer screening and management: an evaluation using the AGREE instrument. BMC Med 2012;10:143.

[13] Zhang LL, Li YP, Zhang C, et al. Analysis on status of clinical guidelines and evaluation on evidence-based guidelines of children in China (Chinese). Chin J Evid-Based Med 2011;11:991–9.

[14] Yang C, Zhang Z, Zhang L, et al. Quality assessment of clinical practice guidelines on tic disorders with AGREE II instrument. Psychiatry Res 2018;259:385–91.

[15] AGREE Enterprise. AGREE II training tools 2017. Available at: https://www.agreetrust.org/resourcecentre/agree-ii-training-tools/. Accessed 16 January 2019

[16] Hoffmann-Esser W, Siering U, Neugebauer EAM, et al. Guideline appraisal with AGREE II: systematic review of the current evidence on how users handle the 2 overall assessments. PLoS One 2017;12: e0174831.

[17] Hattersley AT, Greeley SAW, Polak M, et al. ISPAD clinical practice consensus guidelines 2018: the diagnosis, and management of monogenic diabetes in children, and adolescents. Pediatr Diabetes 2018;19(Suppl 27):47–63.

[18] Shaffer LG, Agan N, Goldberg JD, et al. American College of Medical Genetics statement of diagnostic testing for uniparental disomy. Genet Med 2001;3:206–11.

[19] Genetic Evaluation of Congenital Hearing Loss Expert PanelGenetics evaluation guidelines for the etiologic diagnosis of congenital hearing loss. Genet Med 2002;4:162–71.

[20] Shaffer LG. American college of medical genetics guideline on the cytogenetic evaluation of the individual with developmental delay or mental retardation. Genet Med 2005;7:650–4.

[21] Sherman S, Pletcher BA, Driscoll DA. Fragile X syndrome: diagnostic and carrier testing. Genet Med 2005;7:584–7.

[22] Byers PH, Krakow D, Nunes ME, et al. Genetic evaluation of suspected osteogenesis imperfecta (OI). Genet Med 2006;8:383–8.

[23] Seaver LH, Irons M. ACMG practice guideline: genetic evaluation of short stature. Genet Med 2009;11:465–70.

[24] Schaefer GB, Mendelsohn NJ. Clinical genetics evaluation in identifying the etiology of autism spectrum disorders: 2013 guideline revisions. Genet Med 2013;15:399–407.

[25] Alford RL, Arnos KS, Fox M, et al. American College of Medical Genetics and Genomics guideline for the clinical evaluation and etiologic diagnosis of hearing loss. Genet Med 2014;16:347–55.

[26] Johnson JA, Caudle KE, Gong L, et al. Clinical pharmacogenetics implementation consortium (CPIC) guideline for pharmacogenetics guided warfarin dosing: 2017 update. Clin Pharmacol Ther 2017;102: 397–404.

[27] Ellard S, Bellanne-Chantelot C, Hattersley AT. European Molecular Genetics Quality Network (EMQN) MODY groupBest practice guidelines for the molecular genetic diagnosis of maturity-onset diabetes of the young. Diabetologia 2008;51:546–53.

[28] Gasser T, Finsterer J, Baets J, et al. EFNS guidelines on the molecular diagnosis of ataxias and spastic paraplegias. Eur J Neurol 2010;17: 179–88.

[29] Giancane G, Ter Haar NM, Wulffraat N, et al. Evidence based recommendations for genetic diagnosis of familial Mediterranean fever. Ann Rheum Dis 2015;74:635–41.

[30] Pattathu J, Gorenflo M, Hilgendorff A, et al. Genetic testing and blood biomarkers in paediatric pulmonary hypertension. Expert consensus statement on the diagnosis and treatment of paediatric pulmonary hypertension. The European paediatric pulmonary vascular disease network, endorsed by ISHLT, and DGPK. Heart 2016;102(Suppl 2): ii36–41.

[31] Zennaro MC, Jeunemaitre X. SFE/SFHTA/AFCE consensus on primary aldosteronism, part 5: genetic diagnosis of primary aldosteronism. Ann Endocrinol (Paris) 2016;77:214–9.

[32] Dawson AJ, Chernos J, McGowan-Jordan J, et al. CCMG guidelines: prenatal and postnatal diagnostic testing for uniparental disomy. Clin Genet 2011;79:118–24.

[33] Subspecialty Group of Cardiology, the Society of Pediatrics Chinese Medical Association, the Editorial Board Chinese Journal of Pediatrics (Chinese)Recommendations for gene assays for cardiomyopathy in children. Chin J Pediatr 2013;51:595–7.

[34] Biss T, Avery P, Brandao L, et al. VKORC1 and CYP2C9 genotype, and patient characteristics explain a large proportion of the variability in warfarin dose requirement among children. Blood 2012;119:868–73.

[35] Hamberg AK, Wadelius M, Friberg LE, et al. Characterizing variability in warfarin dose requirements in children using modelling and simulation. Br J Clin Pharmacol 2014;78:158–69.

[36] Duncan RE, Delatycki MB. Predictive genetic testing in young people for adult-onset conditions: where is the empirical evidence? Clin Genet 2006;69:8–16.

[37] Garrett JR, Lantos JD, Biesecker LG, et al. Rethinking the "open future" argument against predictive genetic testing of children. Genet Med 2019.

[38] Dean SJ, Farmer M. Pediatric cancer genetics. Curr Opin Pediatr 2017;29:629–33.

[39] Kshirsagar N, Swaminathan S, Jog P, et al. Regulatory and ethical issues in pediatric clinical research: recommendations from a panel discussion. J Clin Pharmacol 2017;57:943–6.

[40] Rabassa M, Garcia-Ribera Ruiz S, Sola I, et al. Nutrition guidelines vary widely in methodological quality: an overview of reviews. J Clin Epidemiol 2018;104:62–72.

[41] Talagala IA, Samarakoon Y, Senanayake S, et al. Sri Lankan clinical practice guidelines: a methodological quality assessment utilizing the AGREE II instrument. J Eval Clin Pract 2019;25:630–6.

[42] Shekelle PG, Ortiz E, Rhodes S, et al. Validity of the agency for healthcare research and quality clinical practice guidelines: how quickly do guidelines become outdated? JAMA 2001;286:1461–7.

[43] Grol R, Cluzeau FA, Burgers JS. Clinical practice guidelines: towards better quality guidelines and increased international collaboration. Br J Cancer 2003;89(Suppl 1):S4–8.

[44] Chua ME, Mendoza J, See IVM, et al. A critical review of recent clinical practice guidelines on the diagnosis and treatment of non-neurogenic male lower urinary tract symptoms. Can Urol Assoc J 2015;9:E463–70.

[45] Brosseau L, Rahman P, Toupin-April K, et al. A systematic critical appraisal for non-pharmacological management of osteoarthritis using the appraisal of guidelines research and evaluation II instrument. PLoS One 2014;9:e82986.

[46] Yao L, Chen YL, Wang XQ, 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.

[47] Jadad AR, Moore RA, Carroll D, et al. Assessing the quality of reports of randomized clinical trials: is blinding necessary? Control Clin Trials 1996;17:1–2.

[48] Wang Z, Manichukal A, Goff DCJr, et al. Genetic associations with lipoprotein subfraction measures differ by ethnicity in the multi ethnic study of atherosclerosis (MESA). Hum Genet 2017;136:715–26.