Chinese General Practice Journal 1 (2024) 100035

Contents lists available at ScienceDirect

Chinese General Practice Journal

journal homepage: https://www.keaipublishing.com/en/journals/chinese-general-practice-journal/

# Report on methodological quality assessment of primary care and general practice research in China in 2021☆,☆☆ Quantitative Research, Systematic Review and Guidelines/Consensus Section

Quality Assessment Group for Quantitative Research, Systematic Review and Guidelines/Consensus of Chinese General Practice

ARTICLE INFO

*Keywords:*
General practice
Quantitative research
Systematic review
Methodological quality
Research design
Guidebooks

ABSTRACT

**Background:** As China advances healthcare reforms, research output in general practice and primary care has rapidly expanded in recent years. However, the methodological quality of this literature remains unclear.

**Objective:** This study evaluates the methodological quality of representative quantitative studies, systematic reviews, and guidelines published by Chinese researchers in the field of general practice and primary care in 2021, with the aim of providing an overview of methodological standards in this field.

**Methods:** From a pool of 3,122 papers collected in the Primary Care and General Practice Research Paper Productivity Report in China in 2021, 449 representative studies were sampled. A team of 22 researchers specializing in public health and general practice, from various institutions, assessed the methodological quality of 320 papers (71.3 %) using six design-specific tools (for cross-sectional studies, cohort studies, pre- and post-intervention studies, randomized controlled trials, systematic reviews, and clinical guidelines). Researchers worked in pairs under the supervision of an expert in evidence-based methodology, and descriptive statistics were used to present quality assessment results.

**Results:** In 114 cross-sectional studies, common methodological issues were identified in "whether the source population was representative of the study's target population" (41.2 %), "whether the reliability and validity of the survey instrument could be conclusively demonstrated" (32.5 %), and "whether the survey is clinically meaningful" (26.3 %). Among 25 cohort studies, quality issues were more concentrated in the areas of "whether the cohort was adequately followed up" (44.0 %) and "whether the co-intervention was similar among groups" (56.0 %). Of the 34 pre and post-intervention studies, quality issues were mostly found in the areas of "whether the target outcome was measured multiple times before and after the intervention" (97.1 %), "whether the sample size was large enough to generate confidence in the study results" (82.4 %), and "whether the study participants were representative of the eligible population" (61.8 %). Of the 122 randomized controlled trials, quality concerns were mostly in the areas of "blinding of different stakeholders" (25.4 %-61.5 %), "adequate concealment of random allocation" (41.8 %), and "other risks of bias" (72.1 %). Among the 19 systematic reviews, quality issues were mostly found in the areas "is the source of funding for the included studies reported" (100.0 %), "were the methods of the review developed before the start of the review" (94.7 %), "was heterogeneity reasonably discussed and explained" (84.2 %), and "was the risk of bias of individual studies considered" (84.2 %). Finally, the quality of all six clinical guidelines/consensus was rated low.

**Conclusion:** Overall, methodological quality remains limited in recent Chinese research in general practice and primary care, especially in cross-sectional studies, pre- and post-intervention studies, randomized controlled trials, and clinical guidelines. This highlights an urgent need for comprehensive research training, a stronger emphasis on evidence-based reporting standards, and the development of pragmatic guidelines in this field.

☆ The Chinese version of this paper was published in Chinese General Practice on [2023-12-19] (DOI:10.12114/j.issn.1007-9572.2023.0751). The current English paper is a compliant secondary publication by Chinese General Practice Journal after obtaining copyright permission from both the authors and Chinese General Practice.

☆☆ Quality Assessment Group for Quantitative Research, Systematic Review and Guidelines/Consensus of Chinese General Practice.

https://doi.org/10.1016/j.cgpj.2024.11.004
Received 1 November 2024; Received in revised form 13 November 2024; Accepted 23 November 2024
Available online 12 December 2024

**Table 1**
The number of three types of articles sampled in this study.

| Journal type | Cross-Sectional survey | Interventional study (Non-RCTs) | Randomized controlled trials (RCTs) |
|---|---|---|---|
| Chinese Core Journal | 33 | 8 | 8 |
| SCI/SSCI Journal | 12 | 1 | 3 |
| Other Non-Core Journals | 75 | 100 | 114 |

As one of the 14 strategic "levers" proposed by WHO to strengthen primary care, research in primary care is essential for generating and disseminating knowledge, practical experiences, and evidence. This supports the effective implementation of the other 13 "levers"—such as healthcare workforce supply, payment reform, and infrastructure development—ultimately fostering advancements in healthcare systems and improving population health outcomes.[1] The Basic Healthcare and Health Promotion Law of the People's Republic of China, enacted in 2019,[2] defines primary care as encompassing both essential medical services and public health services, primarily delivered by family doctor teams led by general practitioners(GPs). Thus, in the context of China's current healthcare reform, advancing research in primary care and general practice is critically important for the growth and empowerment of this discipline, supporting the broader development of an accessible and effective primary care system in China.

A bibliometric study indicates that in 2021, researchers in China published 3122 original research articles in primary care and general medicine—approximately 1.5 times the number produced in the United States, three times that of the United Kingdom, and over six times that of Australia—making China the leading country in terms of original research output in this field.[3] However, despite this rapid growth in research output, several barriers limit the effectiveness and quality of research in China. Common issues include isolated and fragmented collaborations among researchers, disparities in research proficiency across regions, and inadequate research infrastructure.[4–6] These obstacles hinder the production of objective, precise, reliable, comprehensive, and unbiased data, limiting the ability to generate robust, reproducible, and generalizable findings.[6]

This study adopts a disciplinary development perspective to conduct a cross-sectional methodological quality assessment of representative quantitative studies, systematic reviews, and guidelines published in 2021. By evaluating the methodological quality of these publications, this study aims to characterize the general methodology used in this field within China. Presenting these insights and analyzing related challenges can help researchers in this field reduce design and reporting errors in future studies, thus improving both research practices and outcomes.

## Methods

### *Literature search and sampling*

This study evaluated articles from the dataset compiled in Primary Care and General Practice Research Paper Productivity Report in China in 2021.[3] A systematic search was conducted in databases including CNKI, Wanfang Data, PubMed, and Web of Science to identify papers published in 2021 that focused on primary care and general practice within China. Two trained researchers with experience in literature analysis screened these articles manually using EndNote 20.4.1 (Clarivate Analytics, Philadelphia, USA, 2020) and Rayyan (https://www.rayyan.ai). The specific search strategy and screening process followed the methodology outlined in reference.[3]

Given the large volume of quantitative studies—comprising cross-sectional studies (1146 papers), non-randomized controlled trials (497 papers), and randomized controlled trials (1276 papers)—a random sampling approach was adopted to ensure study feasibility. Using a 5 % margin of error, a 95 % confidence level, and a 10 % expected incidence rate, the required sample size for each study type was determined (see Table 1). To reflect the relative academic recognition, stratified random sampling was applied. Papers were selected from non-core journals, core Chinese journals, and SCI/SSCI journals based on their proportional representation, resulting in a total sample of 320 papers. For prospective and retrospective longitudinal studies, systematic reviews, and guidelines/consensus articles—which each had fewer than 50 papers in the dataset (20, 48, 20, and 7 papers, respectively)—all available papers in these categories were included in the study.

### *Recruitment and composition of methodological quality assessment group*

To ensure objectivity and rigor in the assessment of methodological quality, 22 researchers who had published high-quality original papers in 2021 were invited nationwide to form the methodological quality assessment group for the quantitative research section. This group was organized through the journal of Chinese General Practice. Among the researchers, 10 were from various public health research institutions, and 12 were from clinical and research institutions in general practice.

Two scientific editors from the journal of Chinese General Practice managed coordination, communication, data cleaning, and visualization tasks. To prevent any potential conflicts of interest associated with the journal, the scientific editors did not participate in the methodological quality assessment and had no influence on the evaluation outcomes.

### *Implementation of methodological quality assessment*

Before starting the formal quality assessment, the following preparatory steps were undertaken.

(1) Training: The assessment group members participated in training sessions led by a methodological expert affiliated with a prominent evidence-based medicine research institution abroad. The month-long training covered assessment techniques and included Q&A sessions to clarify methods. Afterward, members engaged in peer discussions to align on the assessment criteria. They also conducted pilot evaluations on a few unrelated articles to standardize the use of assessment tools and ensure consistent quality evaluations.

(2) Evaluation tool discussion: The methodological expert held multiple discussions with the researchers involved in literature screening to confirm the choice and application of the evaluation tools. These discussions also addressed the specific applicability of each tool to different types of studies included in the dataset.

From August to December 2022, the methodological quality assessment team evaluated the selected papers in pairs. Following the methodological classifications recommended by the Research Agenda for General Practice/Family Medicine and Primary Health Care in Europe,[7] papers were reclassified into six categories (Table 2) to align with specific evidence-based assessment tools.[8–13] The number of assessors allocated to each category was based on the volume of literature in each classification. When discrepancies arose within paired evaluations, consensus was reached through discussions with experts in evidence-based medicine and methodology.

### *Result generation and analysis*

For the evaluation of research papers, results were categorized as "yes," "no," or "unclear." In the evaluation of guidelines or consensus

**Table 2**
Number of papers, number of experts and assessment tools for different categories of methodological research literature.

| Research method | Number of papers (Selected Papers) | Number of experts for evaluation | Evaluation time | Evaluation tool | Tool developer |
|---|---|---|---|---|---|
| Cross-Sectional Study | 114 (120) | 6 | 3 weeks | Risk of Bias Instrument for Cross-Sectional Surveys of Attitudes and Practices[8] | CLARITY Group at McMaster University |
| Cohort Study | 25 (68) | 2 | 2 weeks | Tool to Assess Risk of Bias in Cohort Studies McMaster University[9] | CLARITY Group at McMaster University |
| Pre-Post Intervention Study | 34 (109) | 2 | 1 week | Quality Assessment Tool for Before-After (Pre-Post) Studies With No Control Group[10] | National Heart, Lung, and Blood Institute |
| RCTs | 122 (125) | 10 | 9 weeks | Tool to Assess Risk of Bias in Randomized Controlled Trials[11] | CLARITY Group at McMaster University |
| Clinical Guidelines (Consensus) | 6 (7) | 2 | 3 weeks | National Guideline Clearinghouse Extent of Adherence to Trustworthy Standards Instrument[12] | Institute of Medicine |
| Systematic Review | 6 (7) | 2 | 3 weeks | AMSTAR 2 (A MeaSurement Tool to Assess systematic Reviews 2) [13] | AMSTAR 2 Research Group |

Note: The cohort study category was derived by combining prospective and retrospective longitudinal studies from the original extracted literature. However, since many retrospective longitudinal studies are difficult to assess using existing quality assessment tools (e.g., repeated surveys, secondary data analysis based on panel surveys), the evaluation team ultimately decided not to assess the quality of these studies. For practical reasons, the cohort study quality assessment tool developed by the McMaster University CLARITY group was used to assess both types of studies together.
The pre-post intervention study category includes approximately one-third of the intervention studies (non-RCTs) from the original extracted literature. Due to the large number of studies in the non-RCT intervention category (without a control group) that confuse the design elements of observational and intervention studies, the methodology quality evaluation team decided to assess only the methodology of intervention-before-and-after studies (without a control group) based on the feedback from the evaluation team. Only a subset of the studies that met the criteria were included in the quality assessment.
For the clinical guidelines/consensus and systematic reviews, one study from each category was excluded from the quality evaluation process because it was unsuitable for assessment with the quality assessment tool.

documents, some items were rated not only with “yes,” “ no,” or “unclear,” but also using a 5-point Likert scale to represent differences in quality. Quality assessment tables were created using Microsoft Excel 2019 to document the evaluation data, which were then imported into Stata 17.0 SE (StataCorp) for descriptive statistical analysis and visualization.

## Results

### *Cross-sectional studies*

As shown in Fig. 1, a total of 114 cross-sectional studies were included, comprising 12 papers from SCI/SSCI journals, 33 from core Chinese journals, and 69 from other non-core journals. Quality assessment of these studies involved five criteria, with three areas showing the most frequent negative evaluations: “whether the source population was representative of the study’s target population” (47 studies, 41.2 %), “whether the reliability and validity of the survey instrument could be conclusively demonstrated” (37 studies, 32.5 %), “whether the survey is clinically meaningful” (30 studies, 26.3 %).

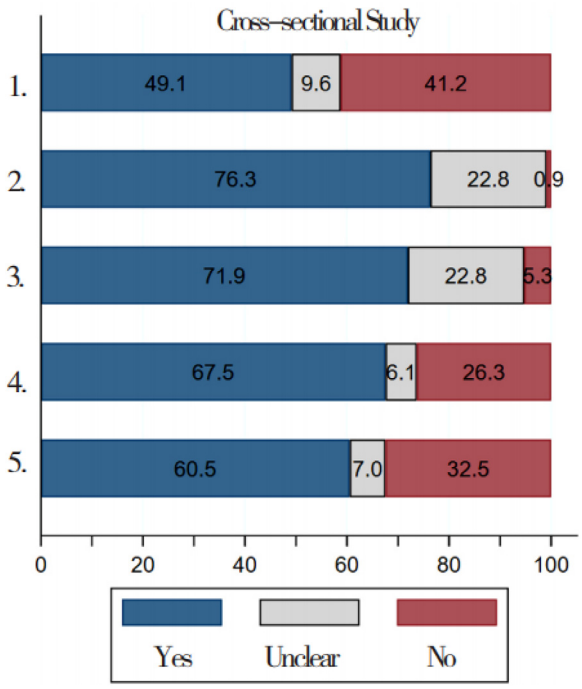


**Fig. 1.** Quality evaluation results of the cross-sectional studies
Notes: 1. Is the source population representative of the population of interest? 2. Is the response rate adequate? 3. Is there little missing data? 4. Is the survey clinically sensible? 5. Is there any evidence for the reliability and validity of the survey instrument?

### *Cohort studies*

Among the 25 cohort studies, 15 were published in SCI/SSCI journals, 7 in core Chinese journals, and 3 in other non-core journals. The quality assessment of these studies included eight criteria, with the following two items receiving the most frequent negative evaluations: “whether the cohort was adequately followed up” (11 studies, 44.0 %) and “whether the co-intervention was similar among groups” (14 studies, 56.0 %). On the other hand, the criteria receiving the most frequent positive evaluations were: “whether selection of exposed and non-exposed cohorts drawn from the same population” (21 studies, 84.0 %), “confidence in the assessment of exposure” (17 studies, 68.0 %), and “confidence in the assessment of outcome” (20 studies, 80.0 %). These findings, summarized in Fig. 2, highlight both strengths and common weaknesses in cohort study methodology.

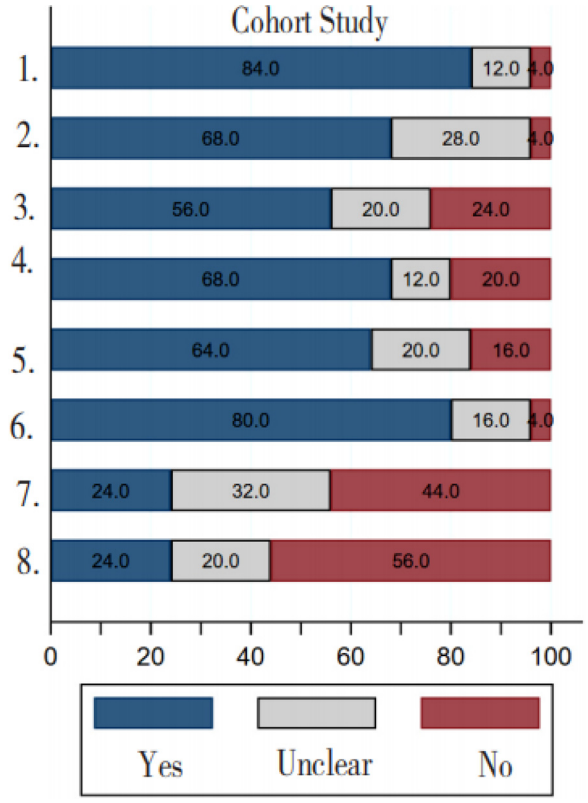


**Fig. 2.** Quality evaluation results of the cohort studies
Note: 1. Was selection of exposed and non-exposed cohorts drawn from the same population? 2. Can we be confident in the assessment of exposure? 3. Can we be confident that the outcome of interest was not present at start of study? 4. Did the study match exposed and unexposed for all variables that are associated with the outcome of interest or did the statistical analysis adjust for these prognostic variables? 5. Can we be confident in the assessment of the presence or absence of prognostic factors? 6. Can we be confident in the assessment of outcome? 7. Was the follow up of cohorts adequate? 8. Were co-interventions similar between groups?

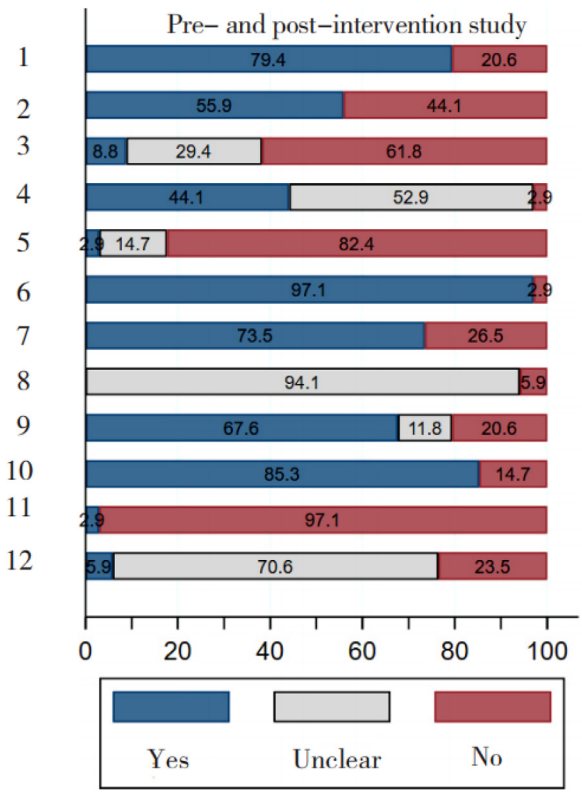


**Fig. 3.** Quality evaluation results of the pre- and post-intervention studies
Note: 1. Was the study question or objective clearly stated? 2. Were eligibility/selection criteria for the study population prespecified and clearly described? 3. Were the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest? 4. Were all eligible participants that met the prespecified entry criteria enrolled? 5. Was the sample size sufficiently large to provide confidence in the findings? 6. Was the test/service/intervention clearly described and delivered consistently across the study population? 7. Were the outcome measures prespecified, clearly defined, valid, reliable, and assessed consistently across all study participants? 8. Were the people assessing the outcomes blinded to the participants' exposures/interventions? 9. Was the loss to follow-up after baseline 20 % or less? Were those lost to follow-up accounted for in the analysis? 10. Did the statistical methods examine changes in outcome measures from before to after the intervention? Were statistical tests done that provided p values for the pre-to-post changes? 11. Were outcome measures of interest taken multiple times before the intervention and multiple times after the intervention (i.e., did they use an interrupted time-series design)? 12. If the intervention was conducted at a group level (e.g., a whole hospital, a community, etc.) did the statistical analysis take into account the use of individual-level data to determine effects at the group level?

*Pre-post intervention studies*

A total of 34 pre-post intervention studies were included in this analysis, with 1 published in an SCI/SSCI journal, 3 in core Chinese journals, and 30 in other non-core journals. The most frequently criticized items in these studies were as follows: "Whether outcome measures of interest taken multiple times before the intervention and multiple times after the intervention" (33 studies, 97.1 %), "whether the sample size sufficiently large to provide confidence in the findings" (28 studies, 82.4 %), "whether the participants in the study representative of those who would be eligible for the test/service/intervention in the general or clinical population of interest" (21 studies, 61.8 %). Positive evaluations highlighted strengths in certain areas: "whether the test/service/intervention clearly described and delivered consistently across the study population" (33 studies, 97.1 %), "whether the study question or objective clearly stated" (27 studies, 79.4 %), "whether the statistical methods examine changes in outcome measures from before to after the intervention" (29 studies, 85.3 %), see Fig. 3.

Randomized controlled trials (RCTs)

This study included a total of 122 RCTs, with 2 articles published in SCI/SSCI journals, 9 in Chinese core journals, and 111 in other non-core journals. Items that commonly received negative assessments among these RCTs were: "Blinding for Different Stakeholders" (31–75 trials, 25.4 %–61.5 %), "adequate concealment of allocation" (51 trials, 41.8 %), "other bias risks" (88 trials, 72.1 %). Conversely, items with predominantly positive evaluations included: "fewer missing outcome data" (97 trials, 79.5 %) and "free of selective outcome reporting" (100 trials, 82.0 %). These findings are detailed in Fig. 4.

*Systematic reviews*

The study evaluated a total of 19 systematic reviews, of which 17 were published in Chinese core journals and 2 in other non-core journals. As shown in Fig. 5, the items frequently receiving negative assessments included: "Funding source disclosure for included studies" (19 reviews, 100.0 %), "predefined review protocol" (18 reviews, 94.7 %), "discussion and explanation of heterogeneity" (16 reviews, 84.2 %), "consideration of individual study bias risk" (16 reviews, 84.2 %). Items that received more favorable assessments included: "Explanation of included study types" (15 reviews, 78.9 %), "duplicate data screening" (15 reviews, 78.9 %), "duplicate data extraction" (16 reviews, 84.2 %). These findings are detailed in Fig. 5.

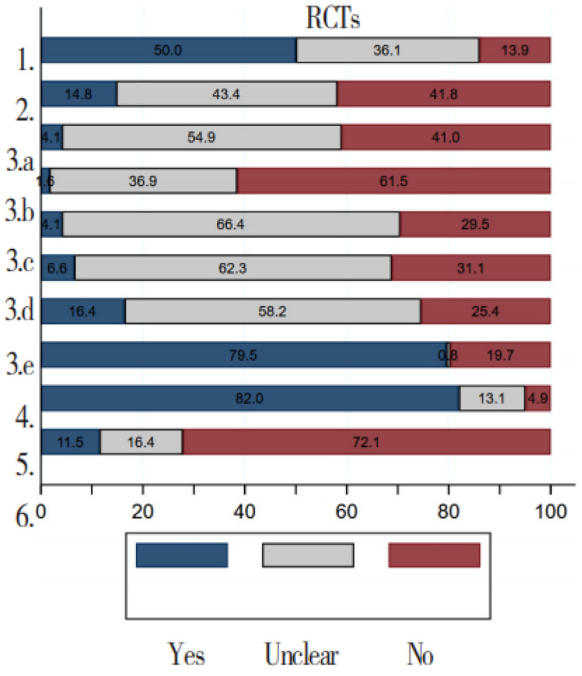


**Fig. 4.** Quality evaluation results of the RCT
Note: 1. Was the allocation sequence adequately generated? 2. Was the allocation adequately concealed? 3a. Was knowledge of the allocated interventions adequately prevented? 3b. Was knowledge of the allocated interventions adequately prevented? 4. Was loss to follow-up (missing outcome data) infrequent? 5. Are reports of the study free of selective outcome reporting? 6. Was the study apparently free of other problems that could put it at a risk of bias?

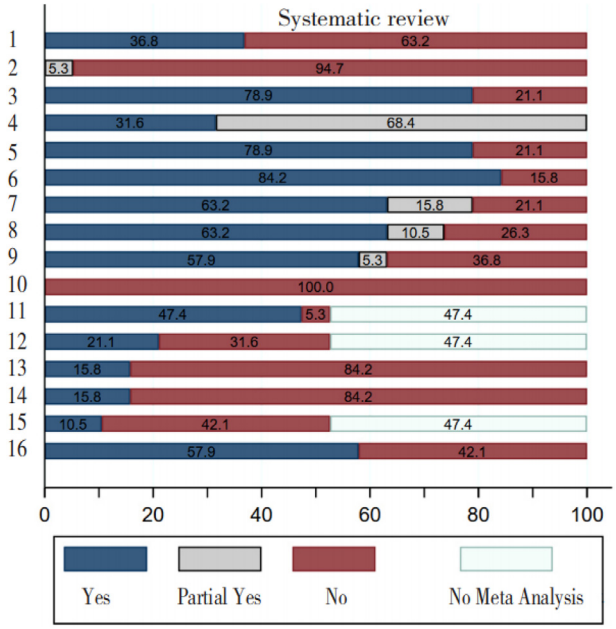


**Fig. 5.** Quality evaluation results of the systematic reviews
Note: 1. Did the research questions and inclusion criteria for the review include the components of PICO? 2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol? 3. Did the review authors explain their selection of the study designs for inclusion in the review? 4. Did the review authors use a comprehensive literature search strategy? 5. Did the review authors perform study selection in duplicate? 6. Did the review authors perform data extraction in duplicate? 7. Did the review authors provide a list of excluded studies and justify the exclusions? 8. Did the review authors describe the included studies in adequate detail? 9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review? 10. Did the review authors report on the sources of funding for the studies included in the review? 11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?
If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis? 13. Did the review authors account for RoB in individual studies when interpreting/discussing the results of the review? 14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review? 15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review? 16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?

*Guidelines/consensus*

As shown in Fig. 6, this study included six consensus/guidelines on primary care/general practice, all published in Chinese core journals. These guidelines demonstrated low compliance rates across several critical criteria. Key areas of non-compliance included failure to “consider patient and public perspectives” (6 studies, 100.0 %), failure to “report search strategies” (5 studies, 83.3 %), failure to “describe study selection processes” (5 studies, 83.3 %), failure to “grade the strength or quality of evidence” (5 studies, 83.3 %), failure to “rate the strength of recommendations” (5 studies, 83.3 %), and failure to “undergo external review” (5 studies, 83.3 %).

## Discussion

This study conducted a representative assessment of the methodological quality of research papers published in 2021 in the fields of primary care and general practice in China. The results indicate widespread, systematic quality issues in this area, with notable deficiencies across research design, implementation, and reporting practices.

*Cross-sectional studies*

From the perspective of methodological quality, the cross-sectional studies included in this assessment reveal two prominent issues. First, in over 40 % of the studies, the sample population did not adequately represent the target population. Additionally, nearly 40 % of the studies used measurement tools or assessment indicators lacking sufficient reliability. Both of these issues may critically limit the scientific validity of the study findings.[14]

From a methodological perspective, it is essential in cross-sectional studies that selected samples sufficiently represent the overall characteristics of the target population. Efforts should be made to ensure alignment between the sample and target populations on sociodemographic factors such as age and gender, as well as other study-relevant characteristics.[15] To achieve this, carefully designed pre-sampling plans and bias-minimizing strategies are necessary. Additionally, researchers need to assess potential differences between respondents and non-respondents to reduce non-response bias, which could otherwise skew findings.[16]

Inadequate sampling methods can lead to findings that do not accurately represent the broader population. For example, a cross-sectional study included in this assessment aimed to examine the association between homocysteine levels and cardiovascular disease risk in older adults in a community in northern Taiwan. The study found a significant association between elevated homocysteine levels and cardiovascular disease risk among residents. However, without a prior sample size calculation, the study surveyed only 396 residents from a single community and lacked detailed descriptions of sampling and recruitment processes. These limitations suggest that the findings may not accurately represent the population of northern Taiwan as a whole. Furthermore,

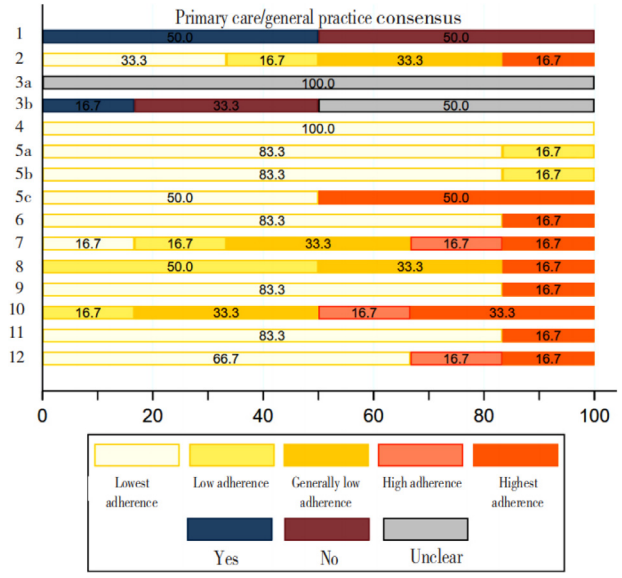


**Fig. 6.** Quality evaluation results of the consensus
Note: 1. Disclosure of Guideline Funding Source; 2. Disclosure and Management of Financial Conflicts of Interests (COIs); 3a. Guideline Development Group (GDG) Composition: Multidisciplinary; 3b. Guideline Development Group (GDG) Composition: Methodologist; 4. Patient and Public Perspectives; 4. Patient and Public Perspectives; 5a. Use of a Systematic Review of Evidence – the Search Strategy; 5a. Use of a Systematic Review of Evidence – the Search Strategy; 5b. Use of a Systematic Review of Evidence – the Study Selection; 5c. Use of a Systematic Review of Evidence – the Synthesis of Evidence; 6. Grading or Rating the Quality or Strength of Evidence; 7. Benefits and Harms of Recommendations; 8. Evidence Summary Supporting Recommendations; 9. Rating the Strength of Recommendations; 10. Specific and Unambiguous Articulation of Recommendations; 11. External Review; 12. Updating.

it remains unclear whether the sampling process effectively minimized bias, randomness, or other confounding factors.[17]

Choosing and applying reliable measurement tools are essential for ensuring the accuracy and credibility of a study's content and results. To address research questions effectively, researchers should ideally select measurement tools that have been validated in similar studies and populations. If suitable tools are unavailable within the local cultural context, using or adapting tools developed in other regions, followed by cross-cultural adjustments and validity checks, is a practical and accepted approach.[18] When existing tools cannot meet the research needs, developing and validating new measurement tools, including conducting reliability tests, should be considered.[19] However, if these steps are neglected, the measurement results may deviate from the actual situation, potentially leading to inaccurate or inconsistent conclusions. In another cross-sectional study that was evaluated, the researchers used an online questionnaire to assess the perceptions of family doctor teams in a certain region regarding the inclusion of pharmacists in providing community pharmacy services. Part of the questionnaire was self-designed, and the final version of the questionnaire was initially tested on 10 family doctor team members and subsequently revised.[20] Given that the researchers' report on the questionnaire validation process was overly brief and the sample size for the pre-survey was too small, the methodology quality assessment group considered this aspect of the study to have significant quality limitations.

*Cohort studies*

In the cohort studies reviewed, key strengths generally included selecting exposed and non-exposed cohorts from the same population, ensuring robust outcome assessments, and demonstrating confidence in exposure assessments. However, a prominent methodological issue was related to the adequacy of follow-up, with approximately 44 % of cohort studies providing insufficient reporting on follow-up information. As one of the essential methods in analytical epidemiological research, the basic principle of cohort studies involves dividing a population into exposed and non-exposed groups based on initial exposure to a specific factor.

The groups are then followed over time to track the incidence of diseases or other outcomes, which enables researchers to calculate and compare the incidence rates between the groups. Through this process, cohort studies assess both the presence and strength of any causal relationship between exposure and outcomes.[21-22] A fundamental requirement for these studies is maintaining comparability between the exposed and non-exposed groups. In many cohort studies, extended follow-up periods increase the risk of participant attrition and can lead to changes in exposure status, which may ultimately affect study outcomes.[23] Therefore, it is crucial for cohort studies to clearly and comprehensively report follow-up information to maintain the integrity and accuracy of the study findings.

In prior literature, commonly omitted follow-up details fall into five primary categories: (1) whether there is any missing follow-up outcome data and the proportion of missing data; (2) the reasons for missing outcome data; (3) whether the missing data is balanced across the exposed and non-exposed groups and whether reasons for missing data are comparable; (4) the extent to which missing data might affect effect estimates; (5) whether appropriate methods were used to handle missing data.

For example, this study reviewed a cohort study examining the relationship between induced abortion history and early-pregnancy stress. While the study outlined its inclusion and exclusion criteria and follow-up duration, it did not provide information on the number of participants initially screened, reasons for baseline exclusions, or details about the number and reasons for loss to follow-up. Another study explored the association between blood pressure variability and type 2 diabetic nephropathy using a cohort design. Although this study reported the number and rate of patients lost to follow-up, it did not specify whether the missing outcome data were balanced across the exposed and non-exposed groups or if suitable methods were used to address missing data in the analysis. These omissions can affect readers' ability to fully assess the findings and evaluate the robustness of the evidence. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement is a widely used checklist aimed at improving the reporting quality of observational research. It includes a version specifically tailored for cohort studies, which serves as a methodological guide for study design, implementation, reporting, and peer review.[24] Given the limited number of cohort studies ultimately included in this review, which did not differentiate between prospective and retrospective cohort types, future research should further analyze and compare the methodological quality across various cohort study designs.

*Pre-Post intervention studies*

The main quality issues in the pre-post intervention studies included in this analysis center around two aspects: the absence of multiple outcome measurements and limitations in sample representativeness and size. The core concept of pre-post intervention studies is to measure an outcome in a group of participants before the intervention, conduct the intervention, and then measure the outcome again afterward, attributing any changes in the outcome to the intervention itself. Compared to RCTs, this approach offers greater flexibility and generally lower costs. However, it also has fundamental limitations, particularly the absence of a control group, which makes it challenging to rule out external factors as potential influences on observed outcomes.[25-26]

To enhance the robustness of pre-post study results, it is common to conduct multiple measurements of outcomes. Repeated measurements allow researchers to better gauge changes in intervention effects across different time points or stages, enhancing the reliability of the find-

ings.[27] Additionally, issues with sample representativeness and sample size in pre-post studies are similar to those found in cross-sectional studies. Limitations in these areas can reduce generalizability and statistical power, confining findings to small-scale populations and undermining confidence in the observed associations.[10]

In addition to the previously discussed aspects, an alternative design approach can further enhance the robustness of pre-post intervention studies: selecting a control group alongside the intervention (exposure) group, with both groups undergoing pre- and post-intervention measurements. By comparing results across four groups—pre- and post-measurements for both the intervention and control groups—this method allows pre-post studies to more closely approximate randomized controlled trials.[28] This method is sometimes also paired with multiple outcome measurements across various time points, further bolstering the reliability of findings. This study design is frequently used in educational intervention research within fields like general practice and primary care. ZOU et al.[29] have provided a comprehensive methodological review on this approach, highlighting its applications and advantages.

*RCTs*

The evaluation of the randomized controlled trials (RCTs) included in this study identified both strengths and areas for improvement. Positively, most studies demonstrated minimal data loss during follow-up and refrained from selective outcome reporting, reflecting strong data integrity and a transparent approach to result reporting. However, significant limitations in design and reporting were identified in 80–90 % of the papers, particularly concerning the implementation of blinding for various stakeholders, the concealment of random allocation, and the handling of other potential sources of bias. These design and reporting gaps present risks that may undermine the validity of the findings, underscoring a need for more rigorous methodological standards in the design and reporting of RCTs.

Inadequate or absent blinding of participants, along with ineffective concealment of random allocation, can significantly affect the behaviors and judgments of different stakeholders involved in a study, potentially leading to biased results and undermining reliability. The Cochrane Handbook notes that lack of blinding in randomized controlled trials (RCTs) can result in inflated estimations of intervention effects, with an average overestimation of 9 %. Similarly, failure to conceal allocation sequences can cause even greater overestimation, with deviations averaging up to 18 %.[30] These findings highlight the importance of rigorous design practices in RCTs. Given that RCTs are among the most precise and methodologically demanding forms of interventional research, ensuring strict adherence to these standards is essential for achieving accurate effect estimates and bolstering the credibility of study outcomes.

The results of this study support prior findings, highlighting a prevalent trend in primary care and general practice research in China: RCTs are authored by only one or two researchers, which may suggest quality concerns within this field.[3] However, this trend also reflects real challenges encountered in conducting these studies. For instance, in acupuncture research, some researchers have experimentally applied blinding techniques for participants, outcome assessors, and statisticians, employing sham acupuncture (i.e., non-meridian point needling) to create a double-blind setup.[31-32] Despite these exploratory efforts, implementing double-blind protocols remains difficult, underscoring the need for further refinement of intervention study standards and protocols.

To enhance the reporting quality of RCTs, researchers are advised to consult relevant guidelines more thoroughly during the design and publication stages. This approach can help mitigate biases that might inflate the perceived effectiveness of interventions or otherwise skew results, contributing to more robust study designs. It is also recommended that journal editors and reviewers evaluate these studies with a focus on whether they provide reliable evidence. The CONSORT 2010 statement offers comprehensive guidance to improve trial reporting, ensuring that readers can thoroughly understand the trial's design, execution, analysis, and interpretation and accurately assess the findings.[33] This guideline, available in multiple languages, including Chinese, provides a methodological framework that supports researchers, reviewers, editors, and readers in assessing the quality and reliability of RCTs effectively.

*Systematic reviews*

The assessment of recent systematic reviews in this field within China indicates some common features. Generally, these reviews demonstrate strong abilities in identifying and selecting relevant data—specifically in literature retrieval and data extraction—while showing notable limitations in synthesizing this data systematically. In particular, the analysis and interpretation of isolated findings that may introduce bias or other risks appear weaker. Given that systematic reviews aim to identify, select, synthesize, and evaluate all available evidence on a research question in a transparent and structured manner,[34] this top heavy tendency could compromise the neutrality and robustness of the reviews' conclusions.

To address this imbalance, researchers could benefit from consulting methodologies established by leading international evidence-based research organizations, such as Cochrane or JBI, when designing systematic reviews. For example, JBI Manual for Evidence Synthesis divides systematic reviews into 12 distinct types, offering case-based guidance on each stage—from introductory concepts and protocol development to data synthesis and analysis.[35] This structured approach is valuable for improving the overall quality of systematic reviews by fostering a comprehensive and unbiased synthesis of evidence.

An additional concern is the frequent omission of funding disclosures in many systematic reviews within this field. While potential conflicts of interest are indeed common, a more pertinent explanation may lie in the fact that research in China has not yet fully aligned with international evidence-based medicine frameworks. Consequently, researchers may prioritize publishing systematic reviews as academic outputs without fully acknowledging their role in generating high-quality evidence that could influence clinical guidelines and health policy.[36] This lack of emphasis on disclosure may also result in a reduced focus on maintaining the neutrality of systematic reviews and managing conflicts of interest effectively.

Furthermore, the findings of this study underscore the importance of collaborative efforts between academic journals and authors to ensure thorough and accurate reporting in systematic reviews. The PRISMA 2020 checklist, widely recognized as a standard for reporting, provides a 27-item guide outlining essential elements such as title, abstract, introduction, methodology, results, discussion, and funding information.[37] By adhering to these guidelines, researchers can improve the quality and transparency of systematic reviews, enhancing readers' ability to understand and critically assess the design and findings of each review.

Deficiencies in systematic review reporting are prevalent not only in primary care and general medicine in China but also across multiple disciplines.[38-39] This challenge is particularly prominent in Chinese-language academic journals.[40] These findings highlight the need for Chinese academic journals to enhance their manuscript review processes, ultimately promoting the publication of high-quality systematic reviews that can serve as valuable evidence in evidence-based medicine.

*Clinical guidelines/consensus statements*

The assessment of clinical guidelines and consensus statements in China indicates that their overall quality in recent years remains relatively low. Common issues include insufficient consideration of patient preferences and values, lack of systematic literature searches, inadequate detailing of inclusion and exclusion criteria, absence of evidence grading and recommendation levels, lack of external peer review, and lack of clear update strategies for guidelines. These gaps underscore

the importance of developing well-structured and contextually relevant guidelines in the field of primary care. Given the limited quantity and generally low quality of research in China, it is essential to establish rigorous, evidence-based guidelines that meet practical needs and to progressively form a consensus within the discipline.

According to evidence-based medicine principles and methodological theory, the Institute of Medicine (IOM) defines clinical practice guidelines as statements developed through a systematic review of evidence and an assessment of the benefits and costs of alternative care options, designed to optimize recommendations for patient care. The IOM emphasizes that high-quality guidelines should incorporate comprehensive evidence, be developed by a multidisciplinary panel (including patient representatives), properly segment patient populations, and thoroughly consider patient values. They should also prioritize clarity, transparency, careful management of bias and conflicts of interest, and clearly specify levels of evidence and recommendation strengths, with provisions for regular updates and revisions.[41]

However, in cases where existing evidence is limited and cannot fully support guideline development, consensus statements by professional associations often serve as a practical alternative. Unlike guidelines, consensus statements primarily reflect the viewpoints of an expert group on a specific topic and often lack transparency in their reporting. This issue is prevalent not only in general practice and primary care research in China but also in international research.[42]

The analysis of the previous five types of studies has confirmed that, at present and likely for some time to come, research in this field within China may struggle to produce sufficient high-quality clinical and health service evidence. This situation could result in an increased volume of lower-quality studies while limiting impactful findings. Given these constraints, a pragmatic approach to advancing evidence-based practice may include identifying areas where high-quality evidence is more readily available and less influenced by population characteristics or environmental factors. Another approach could involve developing primary care guidelines based on international evidence, adopting established international guidelines—such as those created by the U.S. Preventive Services Task Force and the Royal Australian College of General Practitioners for preventive medicine,[43-44] and forming multidisciplinary teams with healthcare providers and patients to create consensus based on national and international data. These strategies would support the progress of evidence-based practice in China's current healthcare environment.

Regardless of the specific approach taken, producing reliable, actionable guidelines and consensus documents that have clinical impact requires adherence to certain evidence-based principles. These include ensuring transparency and openness in the guideline development process, fostering comprehensive involvement from diverse stakeholders (including patients), aligning guidelines with local contexts and national policies, and employing scientifically rigorous consensus methods (e.g., Delphi or nominal group techniques). Additionally, recommendations should be clearly graded and justified, extensive peer review conducted, and professional methodologists engaged throughout the guideline development process.[45-46]

To achieve this goal, it may be advantageous to leverage existing guideline evaluation and development tools, both domestic and international. Widely adopted resources include the GRADE system, which assesses evidence quality and the strength of recommendations[47]; the WHO Handbook for Guideline Development[48]; the McMaster University-developed checklist for guideline creation[49]; the SORT and STEPS tools from the American Academy of Family Physicians, which evaluate clinical and pharmacological evidence, respectively[50-51]; and China's STAR tool for appraising clinical practice guidelines, co-developed by evidence-based medicine experts across several national institutions.[52]

A strong command of these tools would support the rigorous development and evaluation of future evidence-based guidelines and consensus documents in this field. By ensuring these guidelines are credible and trusted by practitioners, patients, and the general public, they can move beyond mere theoretical "impact" or "academic merit." Instead, they could play a substantial role in guiding clinical practice, informing health policies, and contributing positively to the health and well-being of patients and the public alike. In this way, these guidelines and consensus statements could achieve practical relevance, building the trust and influence essential to meaningful progress in healthcare.

*Systemic challenges and recommendations for improvement*

An analysis of studies using the six primary research methodologies—cross-sectional studies, pre-post intervention studies, randomized controlled trials (RCTs), cohort studies, systematic reviews, and clinical guidelines/consensus—indicates that research in primary care and general practice in China faces systemic challenges. Although the quantity of published research in this field places China among global leaders, the overall quality and efficacy of these studies show significant limitations.

Cross-sectional studies, pre-post intervention studies, and RCTs, which dominate in terms of publication frequency, often exhibit notable methodological weaknesses. Cohort studies, although generally reflecting higher methodological quality, are comparatively limited in number. This "upstream" bottleneck in the generation of high-quality evidence potentially diminishes the foundation needed for "midstream" synthesis efforts, such as systematic reviews and clinical guidelines. Further compounding the issue, the "midstream" limitations observed in the rigor of information gathering, analysis, and reporting in systematic reviews and guidelines/consensus documents could weaken the trust that "downstream" healthcare practitioners place in these "scientific findings." This erosion of trust creates a disjunction between evidence generation and practice improvement, as highlighted in evidence-based medicine, where the two processes—"forming evidence" and "enhancing practice"—are meant to function as interconnected cycles that sustain progress. When these cycles are disrupted, the positive feedback loop critical to evidence-based improvement may fail to function effectively.

In practice, this predicament may further lead to a relative absence of scientific evidence tailored to China's specific healthcare system and socioeconomic conditions in the formulation of health policies, service models, and clinical practice guidelines within the fields of primary care and general practice. Consequently, China may have to rely on international evidence developed in countries with healthcare systems, societal conditions, and population characteristics that do not fully align with its own. Alternatively, collective expert opinions, often based on personal experience rather than robust data in China, may fill this gap. This reliance can impact the scientific rigor and practical relevance of China's overall strategies and direction in primary care and general practice.

Addressing this issue will likely require enhancing research capacity within China's primary care and general practice fields. Specifically, improving the ability to design and conduct high-quality research and to develop rigorous systematic reviews, guidelines, and consensus statements will remain crucial for some time. Achieving this goal demands a concerted, top-down approach across the field, engaging professional societies, academic journals, research institutions, and researchers to collectively elevate the quality and applicability of research and evidence.

Based on the research findings and consensus among assessment group members, we propose the following three recommendations for research administrators, researchers, and academic journal editors in the field:

(1) Enhancing research education and training
It is essential to improve the quality of research training within primary care and general practice education, particularly in study design and methodological rigor. This approach aims to create a solid, systematic foundation of theoretical knowledge to strengthen researchers'—especially early-career researchers'—ability to conduct and apply rigorous research even

with limited resources. Such training should continually refine their knowledge and skills to identify, design, and apply scientifically sound methodologies.

(2) Strengthening study design and review standards
Researchers, reviewers, and editors must prioritize adherence to standardized protocols for study design and transparent reporting in original research and systematic reviews. Emphasizing the importance of reporting transparency not only improves understanding and trust in published work but also enhances the studies' capacity to generate reliable evidence. Each research stage, from design through reporting, should focus on creating work that effectively communicates evidence for practical use and application.

(3) Developing guidelines and consensus standards
Establish scientific, feasible methodological standards for developing guidelines relevant to basic healthcare services. These standards should integrate both academic theories while reflecting the current state of the discipline in China. Such an approach would foster the publication of guidelines and consensus statements that engage multiple stakeholders, adhere to transparent and rigorous processes, provide clear and well-reasoned recommendations, and incorporate peer review and plans for regular updates.

An evaluation of the methodological quality of representative 2021 research papers in China's primary care and general practice fields reveals significant limitations in research quality across the board. These limitations are especially pronounced in cross-sectional studies, pre-post intervention studies, and randomized controlled trials. Moreover, the overall quality of the relatively few clinical guidelines and consensus statements remains low. These findings highlight the critical need to improve systematic research training, emphasize adherence to reporting standards, and establish rigorous methodological standards for guidelines and consensus statements. This approach can help produce research outputs that are robust, practical, and relevant, advancing the fields of primary care and general practice in China.

**Review Experts:**

Wannian Liang(Vanke School of Public Health, Tsinghua University, Institute for Healthy China, Tsinghua University)

Hui Yang (Monash University, Australia)

**Members of the Quality Assessment Group for Quantitative Research, Systematic Review and Guidelines/Consensus of Chinese General Practice:**

Lead Writers (Listed in Alphabetical Order by Surname):

Zhikang Ye (Michael G. DeGroote National Pain Centre, McMaster University, Canada)

Yang Wang (Peking University China Center for Health Development Studies)

Kai Lin (Family Medicine Center, the First Affiliated Hospital of Shantou University Medical College)

Yang Zhao (The George Institute for Global Health, University of New South Wales)

Zhijie Xu (Department of General Practice, the Second Affiliated Hospital, Zhejiang University School of Medicine)

Literature Quality Evaluation (Listed in Alphabetical Order by Surname):

Zhiyi Wang(Department of General Medicine, Second Affiliated Hospital, Wenzhou Medical University)

Weijie Gong (Department of Family Medicine, Shenzhen University Health Science Center)

Junli Zhu (School of Public Health, Capital Medical University)

Chenxi Liu(School of Medicine and Health Management, Tongji Medical College, Huazhong University of Science and Technology)

Jinjia Zhang(Department of General Practice, Second Hospital of Hebei Medical University)

Shaofan Chen(School of Medical Administration, Nanjing Medical University)

Chuan Zou (Department of General Practice, Fifth People's Hospital of Chengdu University of TCM)

Kai Lin(Department of General Practice, First Affiliated Hospital, Shantou University Medical College)

Chunmei Lin (Health Development Research Center, National Health Commission)

Hua Jin (Department of General Medicine, Tongji Hospital, Yangpu, Tongji University)

Ran Guo (School of Health Management and Policy, Peking Union Medical College)

Yang Zhao(The George Institute for Global Health, University of New South Wales)

Xin Rao (General practice Center, West China Hospital, Sichuan University)

Mi Yao (Department of General Practice, Peking University First Hospital)

Hongyan Duan(Department of General Practice, Henan Provincial People's Hospital)

Chenwen Zhong(School of Public Health and Primary Care, Faculty of Medicine, The Chinese University of Hong Kong)

Zhongqing Xu(Department of General Practice, Tongren Hospital, Shanghai Jiaotong University School of Medicine)

Xiaotong Gao(First Affiliated Hospital of China Medical University)

Yanli Huang (Management Service Center, Wuhou District Hospital, Chengdu, Sichuan)

Rize Jing (School of Public Administration, Renmin University of China)

Jing Liao(Department of Medical Statistics and Epidemiology, School of Public Health, Sun Yat-sen University)

Zihan Pan(Department of General Practice, Peking University First Hospital)

Literature Screening:

Wang Yang (China Health Development Research Center, Peking University)

Cao Xinyang (Publishing House Co., Ltd. of Journal of Chinese General Practice)

Statistical Analysis:

Wang Yang (Peking University China Center for Health Development Studies)

Xu Zhen (Department of Biology, Hebei Engineering University School of Medicine / Publishing House Co., Ltd. of Journal of Chinese General Practice)

## Declarations

Not applicable.

## Ethics approval and consent to participate

Not applicable.

## Consent for publication

Not applicable.

## Availability of data and materials

Not applicable.

## Funding

Not applicable.

## Authors' other information

Not applicable.

## Declaration of competing interests

L.W., Y.H., W.Y., Z.Y. and L.C. are editorial members of Chinese General Practice Journal, C.X. is editor of Chinese General Practice Journal, they are not envolved in the editorial review or the decision to publish this article. All authors declare that there are no competing interests.

## Acknowledgements

Not applicable.

## References

1. World Health Organization. Operational framework for primary health care: transforming vision into action. 2023 [cited 2023 Apr 10]. Available from: https://www.who.int/publications/i/item/9789240017832.
2. State Administration for Market Regulation of the People's Republic of China. Basic Healthcare and Health Promotion Law of the People's Republic of China. (Chinese). 2019 Dec 28 [cited 2023 Apr 10]. Available from: https://gkml.samr.gov.cn/nsjg/bgt/202106/t20210610_330512.html.
3. Cao X, Wang Y, Xu Z, et al. Research productivity on primary care and general practice in China in 2021. (Chinese). *Chinese General Practice*. 2022;25(34):4232–4240 4258. doi:10.12114/j.issn.1007-9572.2022.0701.
4. Yu D. Research trends in general practice and community health: the future is here. (Chinese). *Chinese General Practice*. 2022;25(34):4227–4231. doi:10.12114/j.issn.1007-9572.2022.0702.
5. Fu Q, Jin H, Yu D. Development and strategy of research capacity in general practice and community health in China from 2001 to 2020. (Chinese). *Chinese General Practice*. 2022;25(34):4252–4258. doi:10.12114/j.issn.1007-9572.2022.0705.
6. Whiting P, Wolff R, Mallett S, et al. A proposed framework for developing quality assessment tools. (Chinese). *Syst Rev*. 2017;6(1):204. doi:10.1186/s13643-017-0604-6.
7. Hummers-Pradier E, Beyer M, Chevallier P, et al. Scope, research needs, and applicable methods in general practice research: a translated summary of the "European General Practice /Family Medicine and Primary Healthcare Research Agenda". *Chinese General Practice*. 2022;25(9):1027–1039. doi:10.12114/j.issn.1007-9572.2022.02.006.
8. Risk of Bias Instrument for Cross-Sectional Surveys of Attitudes and Practices. 2023 Oct 1 [cited 2023 Oct 1]. Available from: https://www.distillersr.com/resources/methodological-resources/risk-of-bias-instrument-for-cross-sectional-surveys-of-attitudes-and-practices-distillersr.
9. Tool to assess risk of bias in cohort studies. 2023 Oct 1 [cited 2023 Oct 1]. Available from:https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-cohort-studies-distillersr.
10. National Heart, Lung, and Blood Institute. Study quality assessment tools. 2020 [cited 2023 Oct 1]. Available from: https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools.
11. Study quality assessment tools - Quality assessment of controlled intervention studies. 2023 Oct 1 [cited 2023 Oct 1]. Available from: https://www.distillersr.com/resources/methodological-resources/tool-to-assess-risk-of-bias-in-randomized-controlled-trials-distillersr.
12. Jue JJ, Cunningham S, Lohr K, et al. Developing and testing the agency for healthcare research and quality's national guideline clearinghouse extent of adherence to trustworthy standards (NEATS) instrument. *Ann Intern Med*. 2019;170(7):480–487. doi:10.7326/M18-2950.
13. Shea BJ, Reeves BC, Wells G, et al. AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*. 2017;358. doi:10.1136/bmj.j4008.
14. Wang XF, Cheng ZS. Cross-sectional studies: strengths, weaknesses, and recommendations. *Chest*. 2020;158(1S). doi:10.1016/j.chest.2020.03.012.
15. Martínez-Mesa J, DA González-Chica, Duquia RP, et al. Sampling: how to select participants in my research study? *An Bras Dermatol*. 2016;91(3):326–330. doi:10.1590/abd1806-4841.20165254.
16. Criqui MH. Potential errors by non-response bias. *Am J Public Health*. 1980;70(12):1301–1302. doi:10.2105/ajph.70.12.1301-c.
17. Shih CC, Shih YL, Chen JY. The association between homocysteine levels and cardiovascular disease risk among middle-aged and elderly adults in Taiwan. *BMC Cardiovasc Disord*. 2021;21(1):191. doi:10.1186/s12872-021-02000-x.
18. Beaton DE, Bombardier C, Guillemin F, et al. Guidelines for the process of cross-cultural adaptation of self-report measures. *Spine*. 2000;25(24):3186–3191. doi:10.1097/00007632-200012150-00014.
19. Boateng GO, Neilands TB, Frongillo EA, et al. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health*. 2018;6:149. doi:10.3389/fpubh.2018.00149.
20. Shi L, Zhu T, Shen M, et al. Analysis of cognition and barriers of community pharmacy care by family doctor teams in Chongming District, Shanghai. (Chinese). *J Pract Pharm*. 2021;39(6):577–580. doi:10.12206/j.issn.1006-0111.202104086.
21. Zhan SY. *Clinical Epidemiology. (Chinese)*. 2nd ed. Beijing: People's Medical Publishing House; 2015.
22. Shen HB, Qi XY. *Epidemiology. (Chinese)*. 8th ed. Beijing: People's Medical Publishing House; 2013.
23. Yan H, Xu YY. *Medical Statistics. (Chinese)*. 3rd ed. Beijing: People's Medical Publishing House; 2015.
24. von Elm E, Altman DG, Egger M, et al. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *PLoS Med*. 2007;4(10). doi:10.1371/journal.pmed.0040296.
25. UK Government. Before and after study: comparative studies. 2023 Feb 15 [cited 2023 Feb 15]. Available from: https://www.gov.uk/guidance/before-and-after-study-comparative-studies.
26. Aggarwal R, Ranganathan P. Study designs: part 4 - interventional studies. *Perspect Clin Res*. 2019;10(3):137–139. doi:10.4103/picr.PICR_91_19.
27. Hill AB. The environment and disease: association or causation? 1965. *J R Soc Med*. 2015;108(1):32–37. doi:10.1177/0141076814562718.
28. Stratton SJ. Quasi-experimental design (pre-test and post-test studies) in prehospital and disaster research. *Prehosp Disaster Med*. 2019;34(6):573–574. doi:10.1017/S1049023X19005053.
29. Zou C, Ou JM, Zeng X, et al. Conducting general practice education intervention research: from design to publication. (Chinese). Chinese General Practice. doi:10.12114/j.issn.1007-9572.2022.0597.
30. Higgins JP, Green S. *Cochrane Handbook for Systematic Reviews of Interventions*. Cochrane Book Series; 2008 [cited 2023]. Available from: *https://training.cochrane.org/handbook* .
31. Wu XK, Stener-Victorin E, Kuang HY, et al. Effect of acupuncture and clomiphene in Chinese women with polycystic ovary syndrome: a randomized clinical trial. *JAMA*. 2017;317(24):2502–2514. doi:10.1001/jama.2017.7217.
32. Zhishun L, Yan L, Huanfang X, et al. Effect of electroacupuncture on urinary leakage among women with stress urinary incontinence - a randomized clinical trial. *Dtsch Z Für Akupunktur*. 2017;60(4):33–34. doi:10.1016/S0415-6412(17)30124-8.
33. Moher D, Hopewell S, Schulz KF, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *Int J Surg*. 2012;10(1):28–55. doi:10.1016/j.ijsu.2011.10.001.
34. Queen's University Library. *Systematic Reviews & Other Syntheses*; 2023 Oct 1 [cited 2023 Oct 1]. Available from: *https://guides.library.queensu.ca/knowledge-syntheses* .
35. JBI Manual for Evidence Synthesis. 2023 Oct 1 [cited 2023 Oct 1]. Available from: https://jbi-global-wiki.refined.site/space/MANUAL.
36. Centre for Reviews and Dissemination. *Health & Environmental Research Online*. York: Centre for Reviews and Dissemination, University of York; 2013.
37. Shamseer L, Moher D, Clarke M, et al. Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015: elaboration and explanation. *BMJ*. 2015;350. doi:10.1136/bmj.g7647.
38. Zhang JX, Wang JC, Han L, et al. Epidemiology, quality, and reporting characteristics of systematic reviews and meta-analyses of nursing interventions published in Chinese journals. *Nurs Outlook*. 2015;63(4):446–455 e4. doi:10.1016/j.outlook.2014.11.020.
39. Zhang Y, Yu DD, Cui DH, et al. Report quality evaluation of systematic review or Meta-analysis published in China Journal of Chinese Materia Medica. *China Journal of Chinese Materia Medica*. 2018;43(6):1254–1260. doi:10.19540/j.cnki.cjcmm.20171220.001.
40. Cao LJ, Yao L, Hui X, et al. Clinical Epidemiology in China series. Paper 3: The methodological and reporting quality of systematic reviews and meta-analyses published by China researchers in English-language is higher than those published in Chinese-language. *J Clin Epidemiol*. 2021;140:178–188. doi:10.1016/j.jclinepi.2021.08.014.
41. Institute of Medicine Committee on Standards for Developing Trustworthy Clinical Practice Guidelines. *Clinical Practice Guidelines We Can Trust*. WashingtonDC: National Academies Press; 2011.
42. Kwong J, Chen H, Sun X. Development of evidence-based recommendations: implications for preparing expert consensus statements. *Chin Med J*. 2016;129:2998–3000. doi:10.4103/0366-6999.195475.
43. U.S. Preventive Services Task Force. USPSTF Home. [cited 2023 Oct 1]. Available from: https://www.uspreventiveservicestaskforce.org/uspstf/.
44. The Royal Australian College of General Practitioners. Guidelines For Preventive Activities in General Practice. East Melbourne: Vic; 2018.
45. Andrews JC, Schünemann HJ, Oxman AD, et al. GRADE guidelines: 15. Going from evidence to recommendation—determinants of a recommendation's direction and strength. *J Clin Epidemiol*. 2013;66(7):726–735. doi:10.1016/j.jclinepi.2013.02.003.
46. JAMA Evidence. Decision making and the patient. [cited 2023 Oct 1]. Available from: https://jamaevidence.mhmedical.com/content.aspx?bookid=847&sectionId=69031507.
47. Guyatt GH, Oxman AD, Vist GE, et al. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008;336(7650):924–926. doi:10.1136/bmj.39489.470347.AD.
48. Guidelines Review Committee. *Quality Assurance of Norms and Standards*. 2nd Edition. WHO handbook for guideline development; 2014. Available from: https://www.who.int/publications-detail-redirect/9789241548960
49. GIN-McMaster Guideline Development Checklist. [cited 2023 Oct 1]. Available from: https://macgrade.mcmaster.ca/resources/ginmcmaster-guideline-development-checklist/.
50. Ebell MH, Siwek J, Weiss BD, et al. Strength of recommendation taxonomy (SORT): a patient-centered approach to grading evidence in the medical literature. *J Am Board Fam Pract*. 2004;17(1):59–67. doi:10.3122/jabfm.17.1.59.
51. Wright J. Introducing STEPS. *Am Fam Physician*. 2003;68(8):1467.
52. Yang N, Zhao W, Pan Y, et al. Development of a rating tool for the scientificity, transparency, and applicability of clinical practice guidelines. (Chinese). *Chin Med J*. 2022;102(30):2329–2337. doi:10.3760/cma.j.cn112137-20220219-00340.