The Scientific Evaluation Standard of Medical Articles for Public: a Delphi Study

Xin Xu

Peking University

Qiyu li

Jinzhou Medical University

Fei Qi

Center for Disease Control and Prevention of Qingdao

Wen-li Yu

Weifang University of Science and Technology

Ying Zhang

Chinese Center for Health Education

Yang Li

University of Shanghai for Science and Technology

Xin-ying Sun

Peking University

Yi-bo Wu (✉ bjmuwuyibo@outlook.com)

Peking University

Research Article

Keywords: evidence-based, Delphi, medical articles for public

Posted Date: October 27th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-954690/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background At present, the quality of articles in the field of medical popularization in China cannot be guaranteed, so it is necessary for experts engaged in scientific popularization to establish a consensus on scientific evaluation criteria to improve it.

Objective This article aims to improve this by reaching a consensus about the scientificity of medical articles for public through Delphi method among experts.

Method The article type is research article. On the basis of AGREE II method, a set of criteria for evaluating the scientificity of popular science articles was established, and then three rounds of Delphi expert consultation were conducted in order to achieve the final consensus.

Results A total of 31 people took part in all three rounds of research, while 4 people lost their visits. Most metric reaches consensus (agreement > 80%) in the final round while there remains argument concerning the "Total number of evidence selected", "Whether a case is regarded as general situation", "If new research evidence exists", "The qualification of the platform for the publication of popular science articles" and "Conflict of interest".

Conclusion This study established a unified scientific evaluation standard for medical popular science articles. The study is of great significance to the establishment of operable scientific evaluation standard of popular science articles. Future research should focus on the diversity in specific aspect.

Introduction

How to properly disseminate health knowledge to the public is an important issue to be discussed globally. In the context of the internet revolution, the public has become accustomed to searching for health information through the internet (1, 2). However, owing to the anonymity and popularity of the Internet, a great deal of health rumor enters the public vision today (3). Compared with them, most of the authoritative advice provided by professionals are lack of competitiveness in readability and communication (4). This situation affects the public's acceptance of the correct popular science information. The evaluation of the writing ability of academic articles has been highly valued (5), and the scientific nature of science textbooks has also been studied a lot (6). However, there exist several differences between popular science writing and academic articles writing (7). In most parts of the world including China, there is no scientific evaluation standard of popular science articles applied in practice. Therefore, as a matter of fact, the scientific evaluation criteria of popular science articles are very vague. Among all the popular science articles, the medicine and health ones are the most concerned by the public, and the most prone to the lack of scientific articles, thus requires the scientific standards the most. Researches concerning the significant linguistic differences telling the truth and lying have been done (8, 9), but it is difficult to evaluate the scientific nature of all medical popular science articles by distinguishing rumors from facts and analyzing the characteristics, because many articles are between true and false.

The practice of evidence-based medicine is the integration of individual clinical expertise with the best available external clinical evidence from systematic research (10, 11). The high-quality evidence is the clinical study data centered on the patient as much as possible to ensure the authenticity of the results. The evidence and its quality are the key of evidence-based medicine. Evidence-based medical practice includes five steps: to raise questions, to search for evidence, to evaluate the evidence, to combine clinical experience with the best evidence to evaluate the

treatment and effect of the patient. The AGREE II is a tool developed to access the methodological quality of practice guidelines (12).

This study aims to develop a consensus on the evaluation criterion of scientificity of medical science popularization works among experts. A Delphi process (13) was developed to involve experts from popular medical writing Committee, CSWA (China Science Writer Association) and ascertain their views on this criterion. The criterion is inspired by the AGREE II standard commonly used in evidence-based medicine mentioned above.

Method

The logic of evidence-based medicine is involved to develop the questionnaire. A step (Figure 1) have been designed and strictly followed through the research.

Delphi panel

All 31 participants were developed from 12 experts by inviting interested colleagues, based on the popular medical writing Committee, CSWA. The average number of years of participating experts engaged in the field of medical popularization of science was more than 9 years, and according to the self-evaluation of participating experts, the average familiarity of experts in all items was more than 0.9 (/ 1.0). The experts involved in Delphi consultation are engaged in the fields of science popularization, such as medicine, pharmacy, nutrition and so on; the popular science work covers science popularization creation, science popularization operation, article evaluation and so on.

Distributing the Delphi survey

The questionnaire was conducted in three rounds, which were sent to the team members in Word via WeChat in 2018.12.16, 2019.02.14 and 2019.05.08. The questionnaire included the introduction of the Delphi method and the explanation of each item, and stressed the importance of completing all three rounds of questionnaires. Each round of questionnaires is scheduled to complete the recovery within three weeks, and most of the questionnaires are collected at the end of the second week. Personalized reminders are sent to unanswered experts, and if the participants determine the completion, the completion time is appropriately extended. All questionnaires were collected within 4 weeks.

Conducting the Delphi survey

In each round of questionnaires, members were asked to fill in the importance score of all six items, and the importance was evaluated by the 5-point Richter scale. 1 - 5 points indicated that the super disagree (SD), disagree (D), Neither agree nor disagree (NAD), agree (A), super agree (SI), respectively. The "recommendation and Evaluation reasons" column is set under each level metric, where the experts are asked to fill in the recommended changes and make the reasons for the corresponding rating. In the second and third rounds of questionnaires, the team members received questionnaires that included the results of the analysis (the weighted average, median, upper and lower quartile and the position of the score among all the experts) and highlighted the need to fill in the comment to dissenting views and the reason for changing the score. The returned questionnaire was analyzed and three items of possible ambiguity were mainly discussed (F. Conflict of interest, whether B.1 regards the controversial content as the academic consensus, B.2 whether the individual case is regarded as the general situation). The members of the contact group clarify and revise the questionnaire according to the feedback, change the ambiguous items, add the items that are not included and carefully delete the unrelated items. In the second and third round of questionnaires, the questionnaire received by the group members included the scores in

the previous round and the analysis results of the previous round of questionnaires, including the weighted average, extreme deviation and standard deviation of group members' scores, The items indicating the disunity of expert opinions (any score < 50%) and the items whose opinions were outside the upper and lower quartiles in the last round is highlighted in different colors. The weighted average is weighted by the degree of expert authority(Cr), which is determined by the familiarity (Cs) (0-1)(the expert self-evaluation of each item) and the basis for judging the metric according to (Ca), $Cr = (Ca+Cs) / 2$. The corresponding relationship between the basis for judging the metric and the Ca value is shown in Table 1.

Table 1
the correspondence between
the basis of judging and the
Ca value

Basis of judging	Ca
Experience	1
Theoretical analysis	0.8
Know from others	0.6
Intuit	0.4

Since only two experts in the first round of questionnaires filled in the column "suggestions and reasons for Evaluation", and the comments were subjective, the reasons collected from the first and second rounds of questionnaires were combined and sent out in the third round, as assessed by the Panel of experts.

Results

Panel participation

All 31 members completed the first round of the survey, while one member lost his visit in the second round and three lost in the third round. The background information of the experts is shown in Table 2.

Table 2
Participant response rate, by panelist characteristic

		round 1		round 2		round 3	
Characteristic	Cr average	N=31	Rate	N=30	Rate	N=27	Rate
Field of expertise							
Nutrition	0.958	5	16.1%	5	16.7%	5	18.5%
Medicine	0.865	17	54.8%	16	53.3%	14	51.9%
Pharmacy	0.846	9	29.0%	9	30.0%	8	29.6%
Experience							
<3 years	0.817	4	12.9%	3	10.0%	3	11.1%
3-8 years	0.866	13	41.9%	13	43.3%	11	40.7%
>9 years	0.868	14	45.2%	14	46.7%	13	48.1%

Delphi Survey

In the first round, 4/6 (66.7%) primary metric and 11/18 (61.1%) secondary metric reached consensus, in which the primary metric F (conflict of interest) and the three secondary metric below it was significantly lower than that of other metric. In the second round, 5/6 (83.3%) primary metric and 11/18 (61.1%) secondary metric reached consensus, which did not change much from the first round of opinion. In the final round, the primary metric F (conflict of interest) remains no consensus, while consensus rate of the secondary metric reaches 13/18(72.2%). The remained equivocal metrics are the primary metric "Conflict of interest" and secondary metric "Total number of evidences selected", "Whether a case is regarded as general situation", "If new research evidence exists" and "The qualification of the platform for the publication of popular Science articles". See Table 3 for details.

Table 3
Summary of item scores by survey round

		round 1			round 2		round 3	
		final consensus	Panel median (25th, 75th percentile)	agreed (%)	Panel median (30th and 70th percentile)	agreed (%)	Panel median (30th and 70th percentile)	agreed (%)
A	Evidence selection	Appropriate	5(5, 5)	0.968	5(5, 5)	1.000	5(5, 5)	1.000
A.1	Total number of evidences selected	Equivocal	4(3.5, 4.5)	0.742	4(3, 4)	0.667	4(4, 5)	0.778
A.2	If choice bias exists	Appropriate	4(4, 5)	0.871	4(4, 5)	0.867	4(4, 5)	0.889
A.3	If advanced evidence exists to support the core point of view	Appropriate	5(4, 5)	0.935	5(4.25, 5)	1.000	5(4, 5)	1.000
A.4	If clear source of information exists	Appropriate	5(4, 5)	0.935	5(4, 5)	0.933	5(4, 5)	0.963
B	Evidence evaluation	Appropriate	5(4, 5)	0.968	5(4, 5)	0.967	5(4.5, 5)	0.926
B.1	Whether the controversial content is regarded as academic consensus	Appropriate	4(4, 5)	0.774	4.5(4, 5)	0.833	4(4, 5)	0.815
B.2	Whether a case is regarded as general situation	Equivocal	4(3, 5)	0.645	4(3, 5)	0.667	4(3, 5)	0.556
B.3	If advanced evidence exists to support expert evidence	Appropriate	5(4, 5)	0.903	4(4, 5)	0.900	4(4, 5)	0.926
B.4	Whether the recommendation is a comprehensive weighing of the pros and cons	Appropriate	5(4, 5)	0.871	4.5(4, 5)	0.900	4(4, 5)	0.926
B.5	If new research evidence exists	Equivocal	4(3.5, 5)	0.742	4(3, 5)	0.633	4(4, 5)	0.778
C	Evidence application	Appropriate	5(5, 5)	0.968	5(5, 5)	1.000	5(5, 5)	0.963

			round 1		round 2		round 3	
C.1	Whether the problem defined is accurate and well related to the evidence	Appropriate	5(4, 5)	0.968	5(4, 5)	0.967	5(4, 5)	0.963
C.2	Whether the conclusion is supported by related evidence	Appropriate	5(4, 5)	0.968	5(4, 5)	0.967	5(4, 5)	0.926
D	Participants	Appropriate	4(3.5,5)	0.742	4(4, 5)	0.867	4(4, 5)	0.889
D.1	Whether the author have relevant professional background	Appropriate	5(4, 5)	0.839	5(4, 5)	0.867	5(4, 5)	0.926
D.2	Whether the author have experience in popular science writing	Appropriate	4(4, 5)	0.806	4(4, 4.75)	0.767	4(4, 5)	0.815
E	Peer evaluation	Appropriate	4(4, 5)	0.806	4(4, 4)	0.800	4(4, 5)	0.852
E.1	Peer review by experts in this field	Appropriate	4(4, 5)	0.935	4(4, 5)	0.900	4(4, 5)	0.963
E.2	The qualification of the platform for the publication of popular Science articles	Equivocal	4(3, 5)	0.710	4(3, 5)	0.600	4(3.5, 5)	0.741
F	Conflict of interest	Equivocal	4(3, 5)	0.581	3.5(3, 5)	0.500	4(3, 5)	0.556
F.1	Whether the publication of the article affects the economic interests of the author	Equivocal	4(3, 5)	0.677	4(3, 5)	0.667	4(3.25, 5)	0.741
F.2	Whether the publication of the article affects the personal reputation of the author	Appropriate	4(4, 5)	0.806	4(4, 5)	0.933	4(4, 5)	0.852

			round 1		round 2		round 3	
F.3	Whether the publication of the article affects the professional interests of the author	Equivocal	4(4, 5)	0.774	4(4, 5)	0.767	4(4, 5)	0.778

Metric adjustment and addition

In accordance with established principles, when the metric satisfies one of the following, it is considered for modification or deletion after consultation with the expert group:

1. There are clear and valuable objections and reasons from the expert opinion;
2. $(SD+D) / n > 80\%$ of the expert opinion.

In the second round of data, Metric c.1 was modified on the basis of feedback from experts. The "accuracy" in the original description "whether the question defined is accurate and evidence-related" has been modified to "clear" in order to better meet the original intention of the item design and the actual situation of popular science work. In addition, in the first and second rounds, it was found that there were relatively many SD reviews of "individual cases" and "conflicts of interest" items (which did not meet the deletion criteria). In the results of the second round of questionnaires, it was found that some experts were obviously inconsistent with the subjective evaluation, and there were possible differences in the description of the above items. That is, the score to be filled in is "the importance of the item as a scientific criterion for the article" rather than the "desirable degree of the item itself". At the end of the second round of questionnaires, the relevant data were modified by contacting the relevant team members, and the relevant contents of the third round of questionnaires were revised and the necessary reminders and explanations were included.

Discussion

The unguaranteed quality of scientific articles in the field of medical popularization of science is an important problem facing the popularization of science in China at present. Determining the scientific evaluation criteria of popular science articles is an important prerequisite for seeking change. This study confirms the consensus of experts on the scientific evaluation criteria of medical popular science article. This consensus shows that it is feasible to establish a unified scientific evaluation standard for medical popular science articles in the medical popular science industry, which covers many specialties. The establishment of this evaluation standard plays a positive role in standardizing the medical popular science industry.

From the consensus results, the expert opinion is concentrated in three areas directly related to the selected evidence, namely, A. Choice of evidence, B. Evidence evaluation, C. The application of evidence, among which the metric A and B present a 100% agreed result. It is shown that, similar to the evaluation standard of scientific papers, the credibility of evidence is the most important thing in evaluating the scientific nature of popular science articles.

However, there are still some disputes on the specific evaluation criteria of evidence. In particular, experts argue on whether newer research evidence should be selected for popular science (77.8%). The opposition view is that the new evidence needs to be repeatedly verified, which is not enough to spread as definite knowledge to the public with weak judgment ability, and there are still many mature knowledge in the field of medical popular science, which is still not understood by the public and needs to be popularized, so it is not cost-effective to carry out such popularization in the situation complicated enough. In addition, the degree of recognition of "Total number of evidences selected" metric is not high (77.8%). The choice of evidence in the evidence rating standard of evidence-based medicine cannot be used as the evaluation standard, but the reason for adding this item to the questionnaire is due to the fact that a large number of pseudo science popularization articles contain only "how to do" but not "why". The disagreement on this metric can be expected before the research.

Of the other three metrics, the most significant differences in expert opinions are concentrated in F. "conflict of interest". The uniformity of expert opinion for the primary metric is only 55.6% ((SA+A) / n*100%). The opinion uniformity of the three secondary metrics is 74.1%, 85.2% and 77.8%, respectively. Only from the scientific point of view, the introduction of any kind of conflict of interest may affect the objectivity of the author, resulting in the bias of author's choice of evidence because of his own point of view, which is not conducive to the scientific nature of the article. However, different from the evidence rating standard of scientific papers, popular science articles are facing two important problems: first of all, different from general scholars, the authors of popular science articles usually need to rely on popular science creation as the main source of income, and it may be too idealistic to require that they do not involve interests at all; in addition, as a means of media education, the popularization of science must include the author's personal position, that is, what the author thinks is correct and should be understood by the public, so it may not be necessary to be completely neutral and objective.

The experts invited in this research have high authority and comprehensiveness. The experts invited for research were all from popular medical writing Committee, CSWA (China Science Writer Association) and have rich experience in the field of medical science popularization in China. The average number of years of experts engaged in science popularization related work is 9 years, and the experts involved in the investigation cover a number of fields involved in the general science popularization in nutrition, medicine and pharmacy.

Strengths and limitations of this study

The Delphi method is used in the study, and its advantages mainly lie in anonymity and iteration. Anonymity is achieved by contacting participants alone, avoiding the influence of expert prestige and authority or opinion leaders in a small range of discussions. It also avoids the problem of group integration and preserves the difference of expert views. In addition, different from the single round of questionnaire survey, because of the three consecutive rounds of Delphi consultation, the members of the group can adjust their views according to the feedback of others over time to promote the achievement of consensus.

However, the potential limitation of the study is the cultural specificity. The design of the questionnaire concerns mostly Chinese society, and the experts who fill out the questionnaire are all from China. In view of the typical problems existing in the popularization of medical science in China, the scores of experts in specific metrics may be higher than the objective standard.

Conclusion

This scientific evaluation criteria of the medical popular science article are in line with the standard requirements in the aspects of expert selection, application of Delphi method, establishment of hierarchical index and expert consultation process, and the index is scientific and credible. This study established a unified scientific evaluation standard for medical popular science articles. That is of great significance to the establishment of operable scientific evaluation standard of popular science articles. In the future, there are some controversial contents mentioned in the discussion that need to be further discussed. The results of the study can provide an important reference for the establishment of unified evaluation criteria for medical popular science articles.

Declarations

Ethics approval and consent to participate:

All methods are implemented in accordance with relevant guidelines and regulations. All experimental protocols were approved by the designated licensing Committee(Scientific Research Office of Shaanxi International Business College).Informed consent was obtained from all subjects.

Consent for publication:

Not applicable.

Data sharing:

Data will be available from the corresponding author on request.

Competing interests

No potential conflict of interest was reported by the authors

Funding

This work was supported by Shanghai Philosophy and Social Science Planning Youth Project in 2020 (Project No.2020EXW005) "Research on the Construction of Scientific Discourse and Public Opinion Guidance Mechanism of Emergent Public Events"

Authors' contributions:

XX is the first author. WY, QF, and YW initiated the concepts of the study. QF, YW, and ZY collected the data. WY, QF, ZY and LY analyzed the data. XX and LQ drafted the manuscript. SX and WY contributed to the interpretation of the results and critical revision of the manuscript for important intellectual content and approved the final version of the manuscript. LQ, YW, ZY, and LY reviews and suggests the manuscript. All authors have read and approved the final manuscript. SX, WY, and XX are the study guarantors.

Acknowledgements:

We would like to thank all the provincial CDC of China for their support and cooperation during this study

References

1. Rideout V. Generation Rx. com: How young people use the Internet for health information. Menlo Park, CA: Kaiser Family Foundation. 2002;22(1):26.
2. Besley JC. The State of Public Opinion Research on Attitudes and Understanding of Science and Technology. *Bulletin of Science, Technology & Society*. 2013;33:12-20.
3. Byford J. *Conspiracy theories: A critical introduction*: Springer; 2011.
4. Oravec JA. On-line medical information and service delivery: implications for health education. *Journal of Health Education*. 2000;31(2):105-10.
5. Rakedzon T, Baram-Tsabari A. To make a long story short: A rubric for assessing graduate students' academic and popular science writing skills. *Assessing Writing*. 2017;32:28-42.
6. Jagger SL, Yore LD. Mind the gap: Looking for evidence-based practice of science literacy for all in science teaching journals. *Journal of Science Teacher Education*. 2012;23(6):559-77.
7. Parkinson J, Adendorff R. The use of popular science articles in teaching scientific literacy. *English for Specific Purposes*. 2004;23(4):379-96.
8. Chua AYK, Banerjee S. *Linguistic predictors of rumor veracity on the Internet*. 2016.
9. Zhang Z, Zhang Z, Li H. Predictors of the authenticity of Internet health rumours. *Health Information & Libraries Journal*. 2015;32(3):195-205.
10. Sackett DL, Rosenberg WM, Gray JM, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. *British Medical Journal Publishing Group*; 2007;455:3.
11. Sackett D. Evidence-based medicine: Seminars in perinatology. Elsevier; 1997;21(1):3-5.
12. Brouwers MC, Kho ME, Browman GP, Burgers JS, Cluzeau F, Feder G, et al. AGREE II: advancing guideline development, reporting and evaluation in health care. *Cmaj*. 2010;182(18):E839-E42.
13. Linstone HA, Turoff M. *The delphi method*: Addison-Wesley Reading, MA; 1975.
14. J JE, Linstone HA, Turoff M. *The Delphi Method: Techniques and Applications*. Technometrics. 1976;18(3):363-4.

Figures

Figure 1

Flowchart of the decision rules adopted for approval and rejection. (SD: Strongly disagree; D: Disagree; NAD: Neither agree nor disagree; A: Agree; SA: Strongly agree)