

Quality Analysis of Chinese Popular Scientific Articles Pertaining to Novel Coronavirus Pneumonia

Shichong Liao (✉ liaoscwhu@163.com)

Wuhan University Renmin Hospital

Liao Shichong

Wuhan University Renmin Hospital

Chen Xinyue

Wuhan University Renmin Hospital

Tu Yi

Wuhan University Renmin Hospital

Sun Shengrong

Wuhan University Renmin Hospital

Zhu Shan

Wuhan University Renmin Hospital

Li Jinxin

Wuhan University

Research article

Keywords: novel coronavirus pneumonia, information quality, scientific articles, Internet

Posted Date: July 21st, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-41998/v1>

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

Abstract

Background: To evaluate the quality, reliability, comprehensiveness and accuracy of the popular science information related to the Corona Virus Disease 2019 (COVID-19) on the Chinese website.

Methods: Searching for scientific articles with the keywords of "novel coronavirus pneumonia" in Baidu, Sogou and 360 search engine at 14:00 on May 22, 2020, and collecting information such as article source, length, upload time, etc. Using JAMA score and 5-point DISCERN scale, two specialists independently analyzed the reliability and quality of scientific articles. Combined with the "Novel Coronavirus Pneumonia Diagnosis and Treatment Program (Trial Version 6)", we made corresponding evaluation on the comprehensiveness and accuracy of articles.

Results: A total of 45 scientific articles were included in the study. The average number of words is 2692.27 ± 1267.61 , and the average upload time is 11.8 ± 10.80 days. Those popular scientific articles are mainly uploaded by government agencies and individuals. However, the JAMA score, DISCERN score, content comprehensiveness and accuracy of the articles were generally low. Among the 45 popular scientific articles, 27 articles were classified as accurate-content group and 18 articles were classified as misleading-content group. Scientific articles of accurate-content groups mainly came from government agencies, universities and hospitals. The publication date of articles in the accurate-content group is closer, the scores of JAMA, DISCERN and content comprehensiveness were higher ($P < 0.05$). The results also showed that JAMA score, DISCERN score, content comprehensiveness and accuracy of popular science articles uploaded by government agencies, universities and hospitals were higher ($P < 0.05$). The results indicated that there was a significant positive correlation between the scores of JAMA, DISCERN, comprehensiveness and accuracy ($P < 0.05$).

Conclusion: Although there are many scientific articles related to the novel coronavirus pneumonia on the Chinese website, the overall detection rate of high-quality scientific articles were low. We are Looking forward to more high-quality medical scientific articles published by government agencies, universities and hospitals. It is suggested that search engines adopt a more optimized sorting method to help more readers get high-quality information.

Background

From the unexplained pneumonia was discovered in Wuhan at the end of December 2019, to the pathogen of pneumonia was identifies as Novel Coronavirus on January 7, 2020, the identification of genomics has been finished, and the virus was officially named 2019-nCoV [1,2]. This virus is different from MERS and SARS, but closely related to them [3]. From the beginning of the disease to June 20, 2020, nearly 215 countries and regions in the world have reported a total of 8657209 cases of diagnosed infections, of which 460440 cases have died[4]. Although the overall fatality rate isn't high, Corona Virus Disease 2019 (COVID-19) is more infectious than MERS and SARS because of its hidden transmission route. As early as February, the number of deaths from this disease has exceeded the total number of deaths from MERS and SARS, and WHO announced the global pandemic of new coronavirus on March 11. In these days, Chinese government has implemented strong intervention measures, including the isolation among cities, grid management of people, establishing hospitals for COVID-19 and publicizing key anti-epidemic information (e.g. wearing masks, washing hands frequently and gathering less). Now the epidemic situation has been basically curbed in China. However, due to uncertain factors

such as hidden transmission of the disease and strong infectivity, the harm caused by it cannot be underestimated. Such as the recent outbreak of new coronavirus pneumonia in Xinfadi wholesale market in Beijing. Therefore, the publicity and education on diagnosis, treatment and prevention of COVID-19 is crucial.

With the prevention measures getting strict, the public transport was stopped and communities were closed in Hubei province, China. People mainly got information related to COVID-19 from the internet. According to the survey, nearly 152 million residents usually obtain popular medical knowledge from the Internet in China [5], and this data may be further increased during the special epidemic period. The three major search engines in China, baidu, sogou and 360, which hold market share over 90%, also became the most important platform for people to acquire COVID-19 related knowledge. Although the internet is very convenient and has broad coverage of knowledge, it is difficult to supervise the quality of information and there is a lack of professional supervisor, which increases the inaccurate information and subjective bias on the internet. Besides, ordinary people lack effective discrimination ability and are easily influenced by false or even harmful information. For example, in the early days of the epidemic, the fear of unknown virus combined with misleading information on the Internet led to a large number of patients with fever came to hospitals, which caused larger widespread infection. This not only made epidemic prevention more difficult, but also made people's health in danger, which raised concerns about the reliability, reality and accuracy of information about COVID-19 online.

However, there has not been relevant research about the quality analysis of information related to COVID-19 on the internet. It is particularly important to carry out quality assessment of COVID-19 related network information as early as possible. This study is the first to assess the reliability, quality, comprehensiveness and accuracy of COVID-19 related popular science articles in the Chinese network, which provides reference for future relevant information, benefiting more people.

Methods

Articles source and scoring criteria

Performing a search using the key word “novel coronavirus pneumonia” with three search engine (baidu, sogou and 360) that holding the highest market share in China (76.42%, 11.35% and 4.71%). The search was proceeded at 14:00 of May 22, 2020. The inclusion criteria are articles that can be read normally and free. Exclusion criteria are: 1. Article is irrelevant to the COVID-19 diagnose and treatment. 2. Article is an advertisement. 3. Article is ratherly too short(< 500 words) 4. Article is a duplicate. Studies have shown that 90% users only browse the first 3 pages of searching results [6]. Therefore, we picked the initial 5 pages from the results given by those 3 search engines(n = 50) as the sample library. The sorting of the picked articles was based on the default program of 3 search engines. The sample library is handed over to specialists for independent reading and recording the basic information of the articles, e.g. publisher, words and publish time. Then we scored articles according to relevant standards.

Articles sorting and scoring

Based on the publishing organization of the article, we divided those popular science articles into 4 groups: by government agencies, by hospital or research institution, by commercial organization or by individual. The reliability of articles was assessed with the American Medical Journal Association(JAMA) score [7] (Table 1). The quality of articles was evaluated according to the modified 5-point DISCERN score, which can be used to

assess the quality of articles non-specifically [8] (Table 2). In order to better evaluate the comprehensiveness and accuracy of the articles, we referred to the "Novel Coronavirus Pneumonia Diagnosis and Treatment Program (Trial Version 6)" and scored the article in terms of 10 aspects, including epidemiological characteristics, causes, clinical manifestations, clinical classification of diseases, auxiliary examinations, diagnosis and identification, treatment, prognosis and quality of life, discharge standards, disease prevention and control. 1 point for each aspect, and higher the score, the more comprehensive the article. In addition, we scored 5 aspects of the information according to the guideline, such as disease disease prevention and control, clinical manifestations, auxiliary examinations, treatment, prognosis and quality of life. Misleading information was recorded as "-1", irrelevant information was "0", partially correct information was "+1" and correct information was "+2". Articles with score > 5 points were included in accurate-content group and articles with no more than 5 points were included in misleading-content group.

Table 1
JAMA score

Standard	Description (1 point each)
Authorship	Authors and contributors, their affiliations, and relevant credentials should be provided.
Attribution	References and sources for all content should be listed clearly, and all relevant copyright information noted.
Currency	Dates that content was posted and updated should be indicated.
Disclosure	Web site "ownership" should be prominently and fully disclosed, as should any sponsorship, advertising, underwriting, commercial funding arrangements or support, or potential conflicts of interest. This includes arrangements in which links to other sites are posted as a result of financial considerations. Similar standards should hold in discussion forums.

Table 2
Modified 5-point DISCERN score

Quality Standard (1 point each)	
1	Are the explanations given in the article clear and understandable?
2	Are useful reference sources given? (publication cited, from valid studies)
3	Is the information in the video balanced and neutral?
4	Are additional sources of information given from which the viewer can benefit?
5	Does the article evaluate areas that are controversial or uncertain?

Statistical analysis

SPSS 22.0 software was used for statistical analysis. Chi-square test was used for difference between groups, and Spearman rank correlation analysis was used for consistency test. The qualitative data was expressed as mean ± standard deviation, median (minimum - maximum), and percentage. The quantitative data was tested by Mann Whitney U and Kruskal-Wallis, and $P < 0.05$ was considered statistically significant.

Results

Screening results of popular science articles

Searching for scientific articles with the keywords of "novel coronavirus pneumonia" in Baidu, Sogou and 360 search engine. 6600000, 14541 and 673000 results were retrieved respectively in them. There were 10 search results per page. We chose the top 5 pages, and a total of 50 articles in each engine was selected as the preliminary sample base. Finally, 45 popular science articles were enrolled to our research (Fig. 1). The excluded results included 55 news, 6 advertisements and 8 articles with too short length. Of the 50 repeated articles, 14 were extracted, while the remaining 36 were excluded. Among the 45 results included in this study, 14 were from government agencies, 10 from universities and hospitals, 10 from commercial websites and 14 from individuals.

General features of the search results

The average word of 45 popular science articles was 2692.27 ± 1267.61 , and the median number of words was 2578 (1067–6432). The average upload time was 11.80 ± 10.80 days, and median time was 7 (1–43) days. The average JAMA score of the articles was 2.00 ± 1.02 points by two specialists, and only one article from the National Health Commission was given 4 points by both specialists. The average DISCERN score is 2.87 ± 1.12 points, and the same one from the National Health Commission was given 5 points by both reviewers. Only 26.67% of the articles showed that the JAMA score was not less than 3 points, and 24.44% of the articles showed that the DISCERN score was not less than 4 points, indicating that the general popular science articles were not reliable enough. In addition, the scores of comprehensibility and accuracy were 5.76 ± 1.80 and 5.16 ± 2.44 . There were 15.56% of the articles were assessed over 8 points for the comprehensiveness, and 11.11% of the articles were assessed over 8 points for the accuracy (Table 3).

Table 3
The general characteristics of 45 popular science articles

	mean	Median (min-max)
Number of words	2692.27 ± 1267.61	2578 (1067–6432)
Upload time (days)	11.80 ± 10.80	7 (1–43)
The JAMA score	2.00 ± 1.02	2.5 (0–4)
The DISCERN score	2.87 ± 1.12	3 (1–5)
The comprehensiveness score	5.76 ± 1.80	5.5 (2–10)
The accuracy score	5.16 ± 2.44	6 (0.5–10)
The results with high score	n (%)	
The JAMA score ≥ 3	12 (26.67%)	
The DISCERN score ≥ 4	11 (24.44%)	
Comprehensiveness score ≥ 8	7 (15.56%)	
Accuracy score ≥ 8	5 (11.11%)	

Stratified analysis based on the accuracy of content

The retrieved results were included in stratified analysis according to the accuracy of content. 27 popular science articles (60%) were divided into accurate-content group and 18 articles (40%) belonged to misleading-content group. The specific characteristics of the two groups were shown in Table 4. There was no significant difference in the number of words between two groups. Articles in the accurate-content group were mainly uploaded by government agencies, universities and hospitals, while in the misleading-content group, the articles were mainly published by commercial websites and individuals, and the differences were statistically significant. Besides, the upload time of articles in the accurate-content group was closer, and the reliability, quality and comprehensiveness scores of articles were significantly higher than those in the misleading-content group ($P < 0.05$).

Table 4
The stratified analysis of two groups according to the accuracy

	accurate-content group (n = 27)	misleading-content group (n = 18)	P	Z/x²
Number of words	3074.25 ± 1208.63	2794.28 ± 1369.28	0.319	-0.996
Upload organization			0.005	12.73
Government agencies	12 (44.44%)	2 (11.11%)		
Universities/Hospitals	6 (22.22%)	1 (5.56%)		
Commercial websites	5 (18.52%)	5 (27.78%)		
Individuals	4 (14.82%)	10 (55.55%)		
Upload time (days)	8.11 ± 6.82	17.33 ± 13.28	0.009	1.923
The JAMA score	2.50 ± 0.84	1.40 ± 0.94	0.001	-3.407
The DISCERN score	3.31 ± 1.02	2.19 ± 0.93	0.001	-3.289
The comprehensiveness score	6.57 ± 1.55	4.56 ± 1.44	< 0.001	-3.758

Table 5
The stratified analysis based on the upload agencies

	Government agencies (n = 14)	Universities/Hospitals (n = 7)	Commercial websites (n = 10)	Individuals (n = 14)	χ^2	p
Number of words	2923.21 ± 1114.96	3713.14 ± 1785.99	2895.80 ± 1063.76	2673.35 ± 1241.48	2.007	0.571
Upload time (days)	10.00 ± 11.31	10.57 ± 11.16	10.40 ± 9.01	15.21 ± 11.60	4.09	0.252
The JAMA score	2.93 ± 0.58	2.35 ± 0.24	1.85 ± 0.82	1.25 ± 1.03	21.2	< 0.001
The DISCERN score	3.64 ± 0.69	4.07 ± 0.61	2.40 ± 0.66	1.82 ± 0.72	28.97	< 0.001
The comprehensiveness score	6.82 ± 1.63	7.29 ± 0.76	5.35 ± 1.29	4.25 ± 1.34	21.01	< 0.001
The accuracy score	6.93 ± 1.69	6.57 ± 1.30	4.35 ± 2.00	3.28 ± 2.26	19.89	< 0.001

Table 6
The correlation between described parameters

		The JAMA score	The DISCERN score	The comprehensiveness score	The accuracy score
The JAMA score	r	1	0.607	0.621	0.615
	p	0	< 0.001	< 0.001	< 0.001
The DISCERN score	r	0.607	1	0.725	0.658
	p	< 0.001	0	< 0.001	< 0.001
The comprehensiveness score	r	0.621	0.725	1	0.664
	p	< 0.001	< 0.001	0	< 0.001

Stratified analysis based on the upload agency

According to the analysis of the general characteristics of popular science articles by different upload agencies, the government agencies, universities/hospitals, commercial websites and individuals have uploaded 14(31.11%), 7(15.56%), 10(22.22%) and 14(31.11%) articles. There was no obvious difference in the number of words and the upload time, but there were statistically significant differences in JAMA score, DISCERN score, comprehensiveness and accuracy. The scores of popular science articles uploaded by government agencies and universities/hospitals are significantly higher than those uploaded by commercial websites and individuals ($P < 0.05$).

The correlation between described parameters

The results shown that the JAMA scores of all popular science articles were positively correlated with the DISCERN scores, comprehensiveness scores and accuracy scores ($P < 0.05$), and the R values are 0.607, 0.621 and 0.615. The DISCERN scores were also positively correlated with comprehensiveness scores and accuracy scores ($P < 0.05$), and the R values were 0.725 and 0.658. The comprehensiveness score was positively correlated with the accuracy score ($P < 0.001$, $r = 0.664$).

Discussion

The COVID-19 is a category B infectious disease that can be spread through droplets, contact, and aerosols. Although the overall mortality rate is not high, but more than 8 million people in 215 countries have been infected[4]. Previous studies have shown that the median number of days from the appearance of the first symptom to death is 14 days. Furthermore, with the increase of age and the impact of basic diseases, the time from onset to death may be shorter [9]. At present, the way that people obtain the information of COVID-19 is mainly browsing the Internet due to the upgrade of containment and prevention measures. Large amount of articles combined with COVID-19 can be retrieved in the three major search engines, but this study only involved the top 5 pages of results of each search engine. The results shown that only 45 articles (30%) met the inclusion criteria, of which only 27 results (18%) were accurate, indicating that the effective detection rate of related popular science articles about COVID-19 on search engines was not high enough. The popular science articles that met the inclusion criteria were mainly published by government agencies and individuals, each accounting for 31.11%. The average word of articles included in the study was 2692.27 ± 1267.61 , and the average upload time was 11.8 ± 10.80 days.

In addition, the study found that the overall reliability, quality, comprehensiveness and accuracy scores of related popular science articles were not high, with average values of 2.0 ± 1.02 , 2.87 ± 1.12 , 5.76 ± 1.80 , 5.16 ± 2.44 . The reason for the low reliability is that the sources, attributions, conflicts of interest, sponsors and citation details of most popular science articles were not mentioned in the content. The reason for the low quality, incompleteness and inaccuracy of the content is that the diagnosis and treatment content were not complete yet, and most of the articles were not updated with the deepening of the awareness of COVID-19. Some articles were published without any conclusive evidence, which tended to lead readers to misjudge the relevant knowledge of the disease. This study is consistent with some researches on the quality of medical-related information abroad: the reliability, quality, and comprehensiveness and accuracy of most medical-related information published online are not entirely satisfactory [10–14].

According to the stratified analysis of content accuracy, it is obvious that popular science articles with accurate content were mainly published by the government agencies, universities and hospitals, while the misleading or wrong articles were mainly published by commercial websites and individuals, and the difference is statistically significant. The accuracy, quality and comprehensiveness scores of those articles with accurate content are also higher than those of the misleading-content group ($P < 0.05$). At the same time, it can be found that the popular science articles with accurate content are often published later, and the content will be updated with the deepening understanding of the disease. For example, in the article published earlier, the prevention methods of COVID-19 only mentioned wearing masks, and measures such as reducing going out, washing hands frequently, and indoor ventilation were not recommended basically. By conducting a stratified analysis of upload agencies, it is shown that the scores (JAMA score, DISCERN score, comprehensiveness score and accuracy score) of articles uploaded by government agencies, universities and hospitals is significantly higher

than the scores of those uploaded by commercial agencies and individuals, which indicates that popular science articles uploaded by government agencies, universities and hospitals have high reliability, better quality, and more comprehensive and accurate content. The reason for this situation is likely to be that the COVID-19 is a new disease, and the first-hand information is mainly kept in the government agencies such as the National Health Commission, the Centers for Disease Control and the hospitals responsible for the treatment of COVID-19. However, due to the lack of reliable resources and professional medical knowledge background, some popular science articles published by commercial websites and individuals have wrong content. This part of the content should be corrected in time to avoid misleading readers.

The medical-related information has strong professionalism, and the methods of assessment for different diseases are often distinctive. Therefore, it is unrealistic for search engines to sort results by disease professionalism and reader's benefit, and finding a simple and universal sorting method is very necessary. This study found there is a positive correlation between any two of the JAMA score, the DISCERN score, the comprehensiveness score and the accuracy score. Thus, it is recommended that search engines can rank search results with using the JAMA score and DISCERN score. This will enable more readers to retrieve high-quality science articles, which can help more readers obtain exact disease-related information.

As the content of popular science articles on the prevention and treatment of COVID-19 is continuously updated, people have gradually broaden their knowledge on the prevention and treatment of COVID-19. This made them treat COVID-19 rationally, helped them understand and support government agencies' prevention measures, and made them pay more attention to self-protection. In terms of treatment, with the renewal of plasma therapy (plasma from rehabilitated patient), cured patients are more proactive in donating plasma. On the other hand, the more professional the articles are, the more difficult they are to read, which makes it hard for non-professionals to understand. For example, some popular science articles mentioned that N95 medical protective masks are better for virus protection, which caused ordinary people to hoard N95 medical protective masks irrationally, resulted in some hospitals having shortages of medical protective supplies. There were also some popular science articles declared that the intermediate host was unclear, and it might be animals such as cats or dogs. As a result, the phenomenon of abandoning pet cats and dogs has occurred in some cities. Therefore, when publishing popular science articles, we should pay attention to avoiding professional terminology to reduce the difficulty of reading, and publish as little unproven content as possible to decrease misunderstanding of the content.

This research also has some limitations. First of all, this study is a cross-sectional study, which can only reflect the situation of the day when the information is collected. Secondly, we only collected the top 5 pages of results from each search engine, and some high-quality articles may be missed. Thirdly, patients may retrieve relevant content on multiple platforms, including some social platforms and video sites, so it is necessary to include them into the future analysis.

In summary, the overall quality of popular science articles on the Chinese website related to the diagnosis and treatment of COVID-19 is generally poor, and patients have limited access to high-quality science information from the Internet. High-quality popular science articles are mainly published by government agencies, universities and hospitals. In the future, professionals from government agencies, universities, colleges and hospitals should be appealed to upload more high-quality medical science articles on the internet, and web

search engines should use more optimized ranking methods to help people obtain accurate and useful popular science information.

Abbreviations

COVID-19:Corona Virus Disease 2019

Declarations

Availability of data and materials

Not applicable.

Ethics approval and consent to participate

This study was approved by the Renmin Hospital of Wuhan University ethical committee.

Consent for Publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study is funded by the Wuhan university and China scholarship council. These entities do not have any influence in the design of the study, analysis and interpretation of the data, or in writing the manuscript.

Authors' contributions

LSC ,CXY and LJY performed the literature review, designed the study, analyzed the results and wrote the manuscripts. TY,SSR and ZS contributed to study design, interpretation of the results, and writing and review of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

- 1.XINHUANET. New-type coronavirus causes pneumonia in Wuhan: expert[EB/OL]. Released on 9 Jan 2020.Available at http://www.xinhuanet.com/english /2020-01/09/c_138690570.htm
- 2.WHO. Coronavirus[EB/OL]. Accessed on 20 Jan 2020.Available at <https://www.who.int /healthtopics /coronavirus>

- 3.Chen Y, Liu Q, Guo D. Coronaviruses: genome structure, replication, and pathogenesis. J Med Virol. 2020. [Epub ahead of print, 22 Jan 2020]. DOI:10.1002/jmv.25681
- 4.Johns Hopkins University and Medicine COVID-19 dashboard by the Center for Systems Science and Engineering at Johns Hopkins University[EB/OL]. Accessed on 20 June ,2020.Available at <https://coronavirus.jhu.edu/map.html>
- 5.China internet network information center[EB/OL].accessed on 5 March 2018.Available at <http://www.cnnic.net.cn/hlwfzyj/hlwzbg/hlwtjbg/>
- 6.iProspect Search Engine User Behaviour Study[EB/OL].Accessed 26 Feb 2018.Available at [http://district4.extension.ifas.fl.edu/Tech/TechPubs/WhitePaper_2006_Search Engine User Behavior.pdf](http://district4.extension.ifas.fl.edu/Tech/TechPubs/WhitePaper_2006_Search%20Engine%20User%20Behavior.pdf).
- 7.Silberg, M. W. Assessing, Controlling, and Assuring the Quality of Medical Information on the Internet[J]. Jama,277(15):1244.
- 8.Charnock D, Needham G, Shepperd S. DISCERN: an instrument for judging the quality of written consumer health information on treatment choices[J]. Journal of epidemiology and community health, 1999,53(2):105-111.
- 9.Wang W, Tang J, Wei F. Updated understanding of the outbreak of 2019 novel coronavirus (2019-nCoV) in Wuhan, China.[Z]. 2020.92,441-447.
- 10.Singh P P, Singh S, Singh A G. YouTube for information on rheumatoid arthritis-A wake up call?[J]. The Journal of rheumatology, 2012,39(5):899-903.
- 11Basch CH, Hillyer GC, MacDonald ZL, Reeves R, Basch CE.Characteristics of YouTubeTM videos related to mammography[J]. J Cancer Educ ,2015,30(4):699–703
- 12.Agarwal N, Hansberry D R, Singh P L, et al. Quality Assessment of Spinal Cord Injury Patient Education Resources[J]. Spine,39(11):E701-E704.
- 13.Aydin M A, Akyol H. Quality of Information Available on YouTube Videos Pertaining to Thyroid Cancer[J]. Journal of Cancer Education.
- 14.Esen E, Aslan M, Sonbahar B Ç, et al. YouTube English videos as a source of information on breast self-examination[J]. Breast Cancer Research and Treatment,2019.

Figures

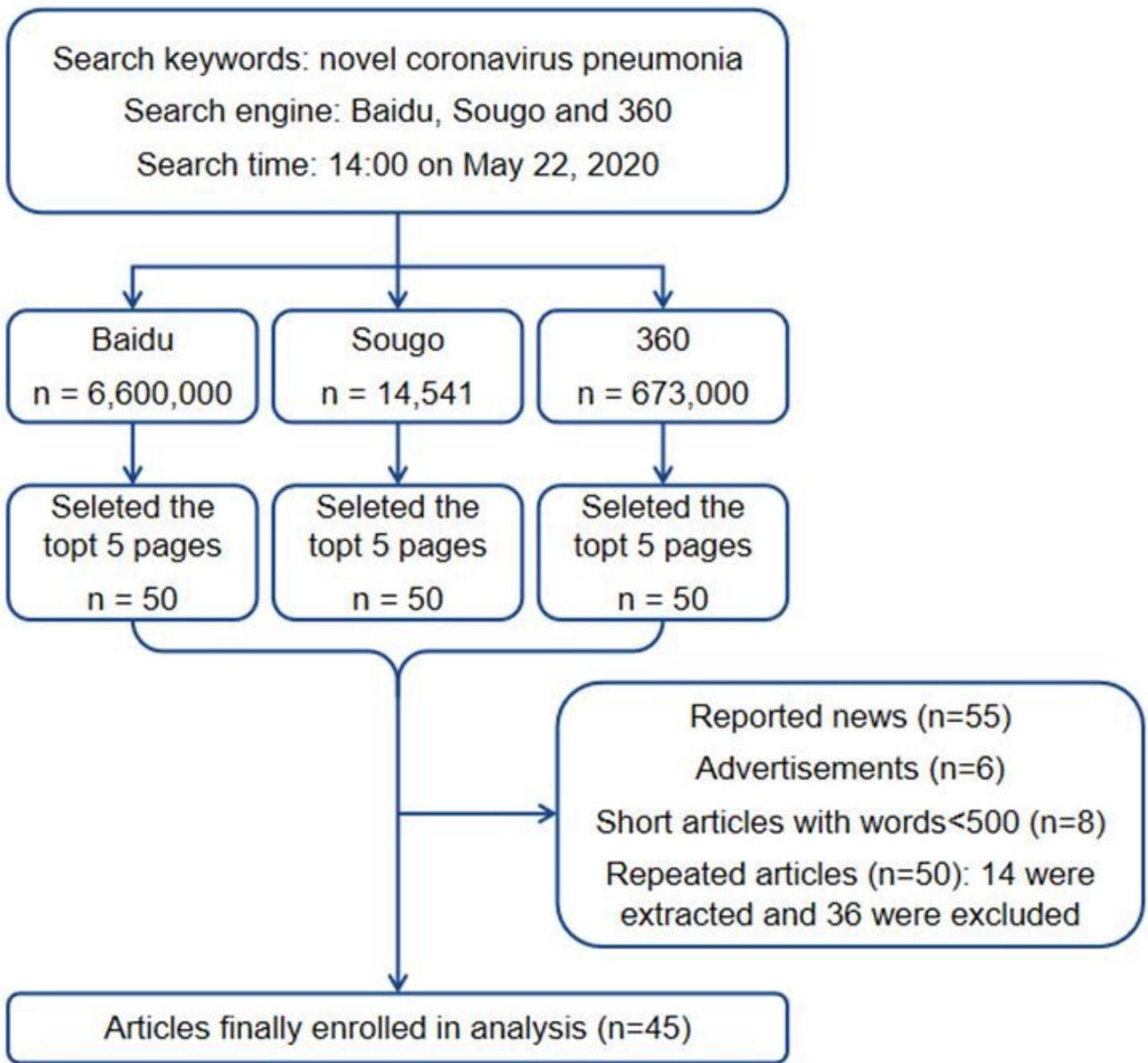


Figure 1

The screening process of popular science articles online.