

# Longitudinal analysis of the quantity and classification of clinical research literature in the PubMed database between 1991 and 2020

**Xiyi Zhao**

Shanghai Jiao Tong University Affiliated Sixth People's Hospital

**Hao Jiang**

Shanghai Jiao Tong University Affiliated Sixth People's Hospital

**Jianyun Yin**

Kunshan Hospital, Nanjing University of Traditional Chinese Medicine

**Hongchao Liu**

Shanghai Jiao Tong University Affiliated Sixth People's Hospital

**Ruifang Zhu**

First Hospital of Shanxi Medical University

**Shencong Mei**

Shanghai Jiao Tong University Affiliated Sixth People's Hospital

**Changtai Zhu** (✉ [zct101@163.com](mailto:zct101@163.com))

Shanghai Jiao Tong University Affiliated Sixth People's Hospital

---

## Research Article

### Keywords:

**Posted Date:** March 8th, 2022

**DOI:** <https://doi.org/10.21203/rs.3.rs-1417817/v1>

**License:**   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

---

# Abstract

**Background:** Clinical research publications have become the dominant source and basis of clinical evidence-based decision-making, and the PubMed database has the most extensive application foundation in the world. Exploring the type and quantity of clinical research publications in the PubMed database is useful for clarifying the changing trends of clinical research development in recent years. Therefore, a longitudinal analysis of the type and quantity of clinical research publications in the PubMed database over three decades was conducted and combined with evidence-based medicine to provide reference for the rational allocation of clinical research resources in future.

**Methods:** The PubMed database was searched to retrieve clinical research according to the type and year of publication from January 1, 1991 to December 31, 2020. The research types were classified as primary and secondary literature using randomized controlled trial (RCT), non-randomized controlled trial (non-RCT), and pragmatic clinical trial (PCT); cohort, case control, and cross-sectional study; case report/series; guideline; expert consensus; narrative, systematic, umbrella, and scoping review; traditional meta-analysis (Tra-meta), systematic review and meta-analysis (Sys-meta), and network meta-analysis (Net-meta). The statistical analyses focused on identifying the changes in the quantity, annual growth rate, and 30-year average growth rate of each specific study type in primary and secondary literature between 1991 and 2020; changes in the constituent proportion of each specific study type in primary and secondary literature; and changes in the ratio of the quantity of guidelines and meta-analysis to primary literature.

**Results:** A total of 1,078,404 primary literatures were retrieved and the constituent proportions were ranked from high to low as case report/series (27.54%), RCT (23.62%), cohort study (21.05%), cross-sectional study (17.49%), case control study (9.15%), non-RCT (1.01%), and PCT (0.15%). Correspondingly, 1,302,173 secondary literatures were retrieved and ranked as narrative review (70.88%), systematic review (15.02%), Sys-meta (13.89%), Tra-meta (4.48%), expert consensus (2.31%), guidelines (1.49%), scoping review (0.68%), Net-meta (0.40%), and umbrella review (0.04%). The average annual growth rate for the primary literature was 10.28%, and ranked from high to low as PCT (83.68%), cohort study (17.74%), cross-sectional study (17.61%), non-RCT (12.11%), case control study (8.86%), RCT (7.68%), case report/series (7.51%); while that for the secondary literature was 10.57%, and ranked from high to low as Net-meta (48.97%), umbrella review (47.09%), scoping review (41.92%), Sys-meta (33.44%), systematic review (33.05%), Tra-meta (12.49%), expert consensus (9.22%), narrative review (8.72%), and guidelines (2.82%). The analysis based on the ratio of the number of secondary literature to the primary literature, the guidelines to primary literature ratio decreased year by year, from 12.08% in 1991 to 1.39% in 2020, for a total ratio of 3.28%, while the meta-analysis to primary literature ratio increased year by year, from 4.59% in 1991 to 27.76% in 2020, for a total ratio of 21.04%.

**Conclusion:** Both the structure and number of clinical studies changed significantly from 1991 to 2020. Based on the trend, the case report/series, case control study, and narrative review are on the decline, while cohort study, cross-sectional study, systematic review, and Sys-meta have increased. Combined

with the grade standard of evidence-based medicine, we recommend enhancing the focus on RCT and cohort study when allocating research resources, as well as improving the quality rather than quantity of Sys-meta and systematic reviews in future.

## Introduction

The PubMed database is a free biomedical database developed by the National Center for Biotechnology Information (NCBI), a division of the U.S. National Library of Medicine (NLM) and has become a major source of literature for biomedical researchers worldwide due to its convenience, accessibility, and extensiveness [1–2]. The PubMed database makes it convenient and efficient for researchers and clinicians to study current guidelines, learn about frontier advancements, and identify future research directions [3].

Sacket et al [4] defined evidence-based medicine (EBM) as “the conscientious, explicit and judicious use of current best evidence in making decisions about the care of individual patients,” and it has become widely accepted due to its strong support of clinical practice guidelines [5–6]. EBM combines the type of study design with the level of evidence to evaluate the reference value of literatures for clinical practice and scientific research [7–8]. According to the evidence pyramid proposed by the Medical Center of the State University of New York (USA) [9], the Grading of Recommendations, Assessment, Development and Evaluations (GRADE) proposed by WHO [10–11], and the JBI levels of Evidence and Grades proposed by Joanna Briggs Institute (Australia) [12], clinical research can be divided into primary and secondary literature. Among the former, RCT has the highest evidence level, case report/series have the highest risk of bias, and other methods such as cohort or case control studies have intermediate evidence level. Among secondary literature, meta-analyses and systematic reviews that included RCT have the highest evidence level, expert consensus has the highest risk of bias, and other types of reviews or meta-analyses that included observational study have an intermediate evidence level [13–14].

A rational allocation of research resources should be conducted to produce more efficient and high-quality results for the advancement of human health; however, the vast amount of research literatures and the multiplicity of study designs pose a great challenge. A longitudinal analysis of the type and quantity of clinical publications can quantitatively measure the constituent proportion and changing trends of various types of studies, and provide reference for the rational allocation of clinical research resources. Although there have been some studies on the changing trends in the literature for specific diseases [15–17], a characterization of the clinical research literature in the field of overall medical health is still lacking. Therefore, our research quantitatively measures the constituent proportion and changing trends of clinical research literatures over three decades.

## Materials And Methods

### *Literature search and data extraction*

First, the research type was clarified as primary and secondary literature. Primary literature comprises observational and interventional studies, with the former consisting of four parts: cross-sectional study, cohort study, case report/series, and case control study, while interventional study consists of RCT, non-RCT, and PCT. Secondary literature is defined as a type of study that relies on primary literature for further analysis, including guidelines, expert consensus, reviews (narrative, systematic, umbrella, and scoping), and meta-analyses (Tra-meta, Sys-meta, and Net-meta). The study type classification is presented in Fig. 1. Second, a literature search was conducted using keywords of study type and the filter tool provided by PubMed, which classifies literature according to type, with year of publication from January 1, 1991 to December 31, 2020. By typing “clinical research [all],” we searched all PubMed literature on clinical research for the last three decades, and by selecting the categories of RCT, PCT, and meta-analysis in the filter tool, combined with the keywords of study type such as Net-meta and umbrella review, a specific literature number of various study types during the period 1991–2020 was obtained.

## Statistical analyses

Statistical analyses were performed using GraphPad Prism 6.02 (GraphPad Software, San Diego, CA, USA) and Microsoft Excel 2019. The statistical analyses focused on identifying the changes in quantity, annual growth rate, and 30-year average growth rate of each specific study type in primary and secondary literature between 1991 and 2020; changes in the constituent proportion of each specific study type in primary and secondary literature; and changes in the ratio of the quantity of guidelines and meta-analysis to primary literature. In this constituent proportion change, we further defined meta-analysis to Sys-meta, Tra-meta, and Net-meta, while the primary literature included RCT, PCT, non-RCT, cohort study, and case control study (Fig. 1). The annual growth rate = (Number of current year publications – number of previous year publications)/number of previous year publications\*100%. The average annual growth rate was calculated using the POWER function included in Microsoft Excel 2019; the specific formula is: the average annual growth rate = POWER [number of publications in year a/number of publications in year b, 1/(a-b)]-1. In this formula, a is the year 2020 and b is the year 1991, and if the number of publications was 0 in 1991 for this study type, then b is the year when the study type first appeared.

## Results

A total of 1,078,404 primary literatures were retrieved between 1991 and 2020 and the quantity and constituent proportion were ranked from high to low as 297,045 for case report/series (27.54%), 254,698 for RCT (23.62%), 226,954 for cohort study (21.05%), 188,563 for cross-sectional study (17.49%), 98,653 for case control study (9.15%), 10,871 for non-RCT (1.01%) and 1,620 for PCT (0.15%) (Supplementary Table 1). Correspondingly, 1,302,173 secondary literatures were retrieved and ranked as 922,973 for narrative review (70.88%), 195,541 for systematic review (15.02%), 61,246 for Sys-meta (13.89%), 58,283 for Tra-meta (4.48%), 30,084 for expert consensus (2.31%), 19,459 for guidelines (1.49%), 8,806 for scoping review (0.68%), 5,216 for Net-meta (0.40%), 565 for umbrella review (0.04%) (Supplementary Table 2).

Specifically, as for the quantity of publications of primary study, in the year 1991, the total number of case report/series was the highest at 3,320; RCT study was second at 2,037; the number of cross-sectional study and cohort study was relatively small at no more than 200, and PCT had not yet appeared. PCT study appeared for the first time in 2011, while in the year of 2013, the number of RCT reached 14,661, for the first time surpassing the 14,551 of case report/series, which was the highest number of primary study in that year. By 2016, the cohort study number had reached 18,004, exceeding RCT (17,140) for the first time and consistently occupying the top of the primary study list in the following years. Cross-sectional study, being one of the primary study types with the lowest number of publications in 1991, surpassed case control study for the first time in 2005 and then exceeded RCT in 2018, eventually emerging with case report/series as the most primary study type after cohort study in 2020.

Correspondingly, regarding the quantity of publications of secondary literature, the total number of narrative review reached 5,790 in 1991, followed by guidelines with 366, expert consensus with 260, Tra-meta with only 139, systematic review with only 8, which was still in the initial stage, while Net-meta, Sys-meta, umbrella review, and scoping review were not yet available. The quantity of systematic review reached 1,168 in 2000, surpassing guidelines, expert consensus, and Tra-meta for the first time, and has been the top two study type of secondary literature after narrative review ever since. In the year 2002, the number of Tra-meta surpassed guidelines and expert consensus, and became the third most numerous study type of secondary literature until 2016. Sys-meta first appeared in 1994, surpassed guidelines and expert consensus in 2008 and 2009, and Tra-meta for the first time in 2016, becoming the most numerous type of meta-analysis and the third most numerous type of secondary literature (maintaining this to date). Scoping review, Net-meta, and umbrella review first appeared in 1999, 2002, and 2006, respectively, and the quantity of publications of such new study types has grown rapidly, reaching 3,119, 1,305, and 222 in 2020, respectively (Fig. 2).

Regarding the composition ratio of primary literature, the proportion of case report/series and RCT among primary literature tended to decrease year by year, conversely, the proportion of cross-sectional study and cohort study had a tendency to increase year by year. Among secondary literature, the proportion of guidelines and narrative review tended to decrease year by year, with the proportion of systematic review and Sys-meta tending to increase year by year (Figs. 3–5).

As for the average annual growth rate, the primary literature was 10.28%, ranked from high to low as PCT (83.68%), cohort study (17.74%), cross-sectional study (17.61%), non-RCT (12.11%), case control study (8.86%), RCT (7.68%), and case report/series (7.51%); while the secondary literature was 10.57%, ranked from high to low as Net-meta (48.97%), umbrella review (47.09%), scoping review (41.92%), Sys-meta (33.44%), systematic review (33.05%), Tra-meta (12.49%), expert consensus (9.22%), narrative review (8.72%), and guidelines (2.82%).

Specifically, in the original literature, PCT had the lowest base and the most change in annual growth rate, while that of other original literature types was relatively stable; cross-sectional study and cohort study

maintained a positive growth trend, and the annual growth rate increased steadily in the last three years. The annual growth rate of RCT maintained a positive trend before 2018, but there has been a slight fluctuation in the most recent three years, although the number of publications tends to be stable. The growth rate of non-RCT, case report/series in 2020 was significantly higher than earlier. In the secondary literature, the change rate of the annual growth rate of Net-meta, umbrella review, and scoping review was large because of their low base, while that of annual growth rate of other secondary literature research types was relatively stable. The annual growth rate of guidelines, expert consensus, and narrative review was relatively lower; while the systematic review and Sys-meta (which are the improved methods of narrative review and Tra-meta), basically maintained a positive growth trend, significantly higher than that of narrative review and Tra-meta (Supplementary Tables 3–4, Fig. 6).

Based on the analysis of the ratio of secondary literature to the number of original studies, the proportion of all guidelines and meta-analysis to the original studies that were included from 1991 to 2020 were 3.28% and 21.04%, respectively, and the average annual growth rates of the included original studies, guidelines, and meta-analysis that were included were 10.78%, 2.82%, and 17.87%, respectively. Specifically, the ratio of guidelines to original literature decreased year by year, from 12.08% in 1991 to 1.39% in 2020, with a generally stable decline rate; while the proportion of meta-analysis and included original literature increased year by year, from 4.59% in 1991 to 27.76% in 2020, of which the increase was at its highest from 2008 to 2014, after which the growth rate stabilized (Figs. 7–8).

## Discussion

PubMed is the most widely used free database in the field of medical health. In the last 30 years, the increase of scientific researchers, and the update of scientific research methods, the number of papers and research design has exploded. This study extracted the number of publications of different research types published on PubMed for the last three decades, and defined the change in trend of the number of publications of different research designs during this period by analyzing the change in trend of their number, growth rate, composition ratio, and the proportional relationship between original literature and secondary literature research. Combined with the quality level of EBM corresponding to different research designs, this study provides researchers with an overview of different research design publications for the 30 years under study and provides a change trend of medical research.

In the original studies, case report/series (as the research type with the largest number of early studies), have a low average annual growth rate, and their proportion in the original studies decreased year by year. At present, they are no longer the original literature type with the largest proportion. Although the sum of the proportion of cohort study and cross-sectional study was only about 8% in 1991, their average annual growth rate reached about 18%. Currently, the proportion of these types has surpassed that of RCT and become one of the largest types of original literature.

The number of RCT increased steadily from 1991 to 2017, its proportion in the original studies ranked second from 1991 to 2014. The number of RCT surpassed case report/series for the first time in 2015 but

fell to the fourth place from 2016 to 2020, which may be related to the standardization and high ethical requirements of RCT.

As a control experiment closer to the application in the real world, PCT first appeared in 2011. Although it currently accounts for a low proportion of original literature and there is no consensus on its evidence quality level, considering its strong clinical application and the advantages that the research design basically follows the control experiment, we think PCT may attract more researchers' attention in the future.

In the secondary literature research, the narrative review accounted for 88%, while the guidelines and expert consensus accounted for nearly 10% in 1991. Later, the average annual growth rates of narrative review and guidelines became the lowest in secondary literature research, and their proportion also decreased year by year. Currently, although narrative review is still the secondary literature with the highest proportion, its gap with systematic review and Sys-meta (the second and third systematic review), has been greatly reduced compared with 1991, from more than 85% to less than 20% at present. The proportion of guidelines has even fallen below 1%, and we believe this trend will develop further in the future.

Compared with narrative review and meta-analysis, systematic review and Sys-meta with their more scientific and systematic research methods show a significantly higher growth rate than traditional methods. The number of Sys-meta surpassed Tra-meta in 2016, and the difference between systematic review and narrative review also decreased from 88.1–27.94%. Combined with the fact that systematic review and Sys-meta are higher than narrative review and Tra-meta in the level of evidence quality, we believe that this trend will be further deepened in the future.

Scoping review, which can describe the research results and research scope of a specific research field in more detail by investigating or exploring the research status, degree, and methodology of a research field or topic, appeared for the first time in 1999; the research method of Net-meta, which can analyze the relationship between more than two interventions based on multiple studies by means of indirect comparison or mixed comparison, appeared for the first time in 2002; while in 2006, the umbrella review appeared to summarize broader evidence when the systematic review and meta-analysis of a medical research topic reached a certain number. These new research types increased rapidly, with an average annual growth rate of more than 40%, further enriching the methods of secondary literature research.

Our research shows that the average annual growth rate of meta-analysis (17.87%) from 1991 to 2020 is significantly higher than that of included original literature (10.78%) and guidelines (2.82%). The proportion of guidelines to included original literature decreased year by year, from 12.08% in 1991 to 1.39% in 2020, while the proportion of meta-analysis to original literature increased year by year, from 4.59% in 1991 to 27.76% in 2020. For the ratio between meta-analysis and original literature, each meta-analysis needs to include a certain amount of original literature to ensure the robustness of research results of meta-analysis. However, there were many overlapped and partially overlapped meta-analyses, which means that the meta-analysis with high quality and high evidence level is reduced. Therefore, we

suggest that attention paid to meta-analysis be reduced and original literature strengthened, especially RCT with its high evidence level.

Compared with other studies focusing on the characteristics of literature related to specific diseases, our study is the first to analyze the changes in the number, growth rate, composition ratio, and the proportional relationship between the original literature and the secondary literature of all clinical original studies and secondary literature studies against the background of the explosive growth in the number of various research methods and publications in the last 30 years. This study defines the classification characteristics and change trend of clinical research literature and, combined with the quality level of EBM evidence, provides researchers with an overview of different research design publications in recent 30 years, as well as an EBM reference for the rational allocation of research resources.

Our research also has some limitations. First, we only extracted the number of publications of different research types without evaluating their quality. Therefore, the increase of research types with high evidence quality such as RCT, Sys-meta, and systematic review cannot be entirely equal to the improvement of research evidence quality. Second, we only used the PubMed database for literature retrieval, and the retrieval was dependent on the use of keywords and PubMed filtering tools. There may be a small amount of missed detection and repeated retrieval.

In general, the characteristics of clinical research literature have changed significantly from 1991 to 2020. Case report/series, case control study, and narrative review showed a downward trend, while cohort study, cross-sectional study, systematic review, and Sys-meta developed rapidly. Combined with the evidence quality level of EBM, we suggest that more attention should be paid to RCT and cohort study when allocating research resources to improve the quality rather than quantity of Sys-meta and systematic review.

## **Declarations**

### **Acknowledgements**

Not applicable.

### **Author Contributions Statement**

CTZ, SCM, RFZ and JYY designed the study. XYZ, CTZ, and SCM retrieved literature and extracted the data. CTZ, HJ, XYZ, HCL, SCM, RFZ and JYY performed the analysis. XYZ, CTZ, SCM, HJ, RFZ and JYY wrote the paper. XYZ, JYY and CTZ prepared figures 1-8. All authors contributed to the final revision of the project. All authors reviewed the manuscript.

### **Funding**

None.

## Availability of data and materials

Not applicable.

## *Competing interests*

The authors declare that they have no competing interests.

## *Consent for publication*

Not applicable.

## *Ethics approval and consent to participate*

This article does not contain any studies with human participants or animals performed by any of the authors.

## References

1. Yang H, Lee HJ. Research Trend Visualization by MeSH Terms from PubMed. *Int J Environ Res Public Health*. 2018 May 30;15(6):1113.
2. Falagas ME, Pitsouni EI, Malietzis GA, Pappas G. Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths and weaknesses. *FASEB J*. 2008 Feb;22(2):338–42.
3. Lu Z. PubMed and beyond: a survey of web tools for searching biomedical literature. *Database (Oxford)*. 2011 Jan 18; 2011: baq036.
4. Sackett DL, Rosenberg WM, Gray JA, Haynes RB, Richardson WS (1996) Evidence based medicine: what it is and what it isn't. *BMJ* 312: 71–72.
5. Woolf SH, George JN (2000) Evidence-based medicine. Interpreting studies and setting policy. *Hematol Oncol Clin North Am* 14: 761–784.
6. Beckmann JS, Lew D. Reconciling evidence-based medicine and precision medicine in the era of big data: challenges and opportunities. *Genome Med*. 2016 Dec 19;8(1):134.
7. Atkins D, Eccles M, Flottorp S, Guyatt GH, Henry D, Hill S, et al. (2004) Systems for grading the quality of evidence and the strength of recommendations I: critical appraisal of existing approaches The GRADE Working Group. *BMC Health Serv Res* 4: 38.
8. Sackett DL. Rules of evidence and clinical recommendations on the use of antithrombotic agents. *Chest*. 1989;95:2S–4S.
9. [http://library downstate.edu/ebmdos/2100.htm](http://library.downstate.edu/ebmdos/2100.htm)
10. Guyatt GH, Oxman AD, Vist GE, Kunz R, Falck-Ytter Y, Alonso-Coello P, Schünemann HJ; GRADE Working Group. GRADE: an emerging consensus on rating quality of evidence and strength of recommendations. *BMJ*. 2008 Apr 26;336(7650):924–6.

11. Balshem H, Helfand M, Schünemann HJ, Oxman AD, Kunz R, Brozek J, Vist GE, Falck-Ytter Y, Meerpohl J, Norris S, Guyatt GH. GRADE guidelines: 3. Rating the quality of evidence. *J Clin Epidemiol*. 2011 Apr;64(4):401–6.
12. <https://jbi.global/jbi-approach.html#tabbed-nav=levels-of-Evidence>
13. Burns PB, Rohrich RJ, Chung KC. The levels of evidence and their role in evidence-based medicine. *Plast Reconstr Surg*. 2011 Jul;128(1):305–310.
14. Hung BT, Long NP, Hung le P, Luan NT, Anh NH, Nghi TD, Hieu MV, Trang NT, Rafidinarivo HF, Anh NK, Hawkes D, Huy NT, Hirayama K. Research trends in evidence-based medicine: a joinpoint regression analysis of more than 50 years of publication data. *PLoS One*. 2015 Apr 7;10(4):e0121054.
15. Zou Y, Li Q, Xu W. Scientific research output in orthopaedics from China and other top-ranking countries: a 10-year survey of the literature. *BMJ Open*. 2016 Sep 16;6(9):e011605.
16. Deshazo JP, Lavallie DL, Wolf FM. Publication trends in the medical informatics literature: 20 years of "Medical Informatics" in MeSH. *BMC Med Inform Decis Mak*. 2009 Jan 21;9:7.
17. Lee JE, Kim YY. How Should Biobanks Prioritize and Diversify Biosample Collections? A 40-Year Scientific Publication Trend Analysis by the Type of Biosample. *OMICS*. 2018 Apr;22(4):255–263.

## Figures

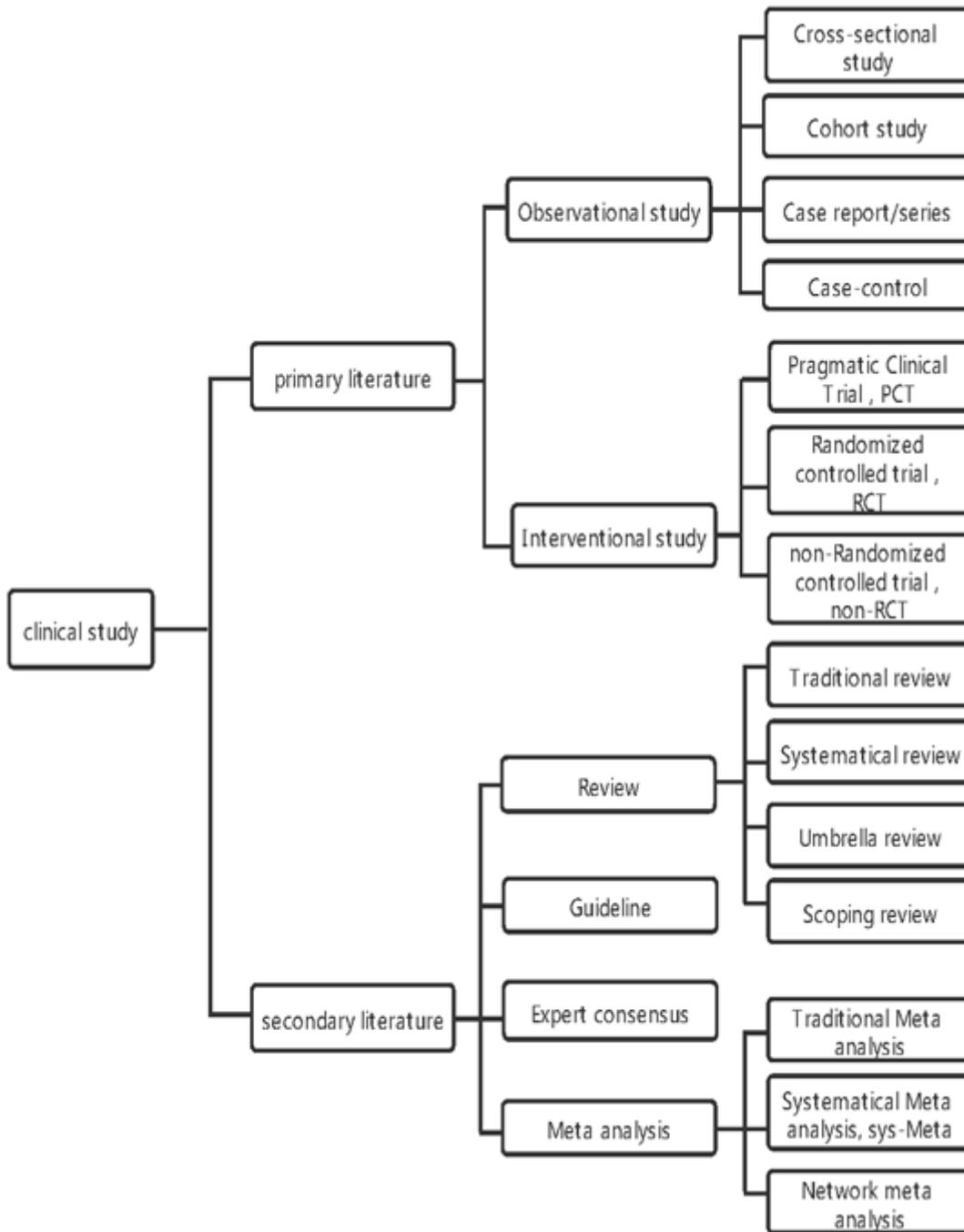


Figure 1

Classification of study types

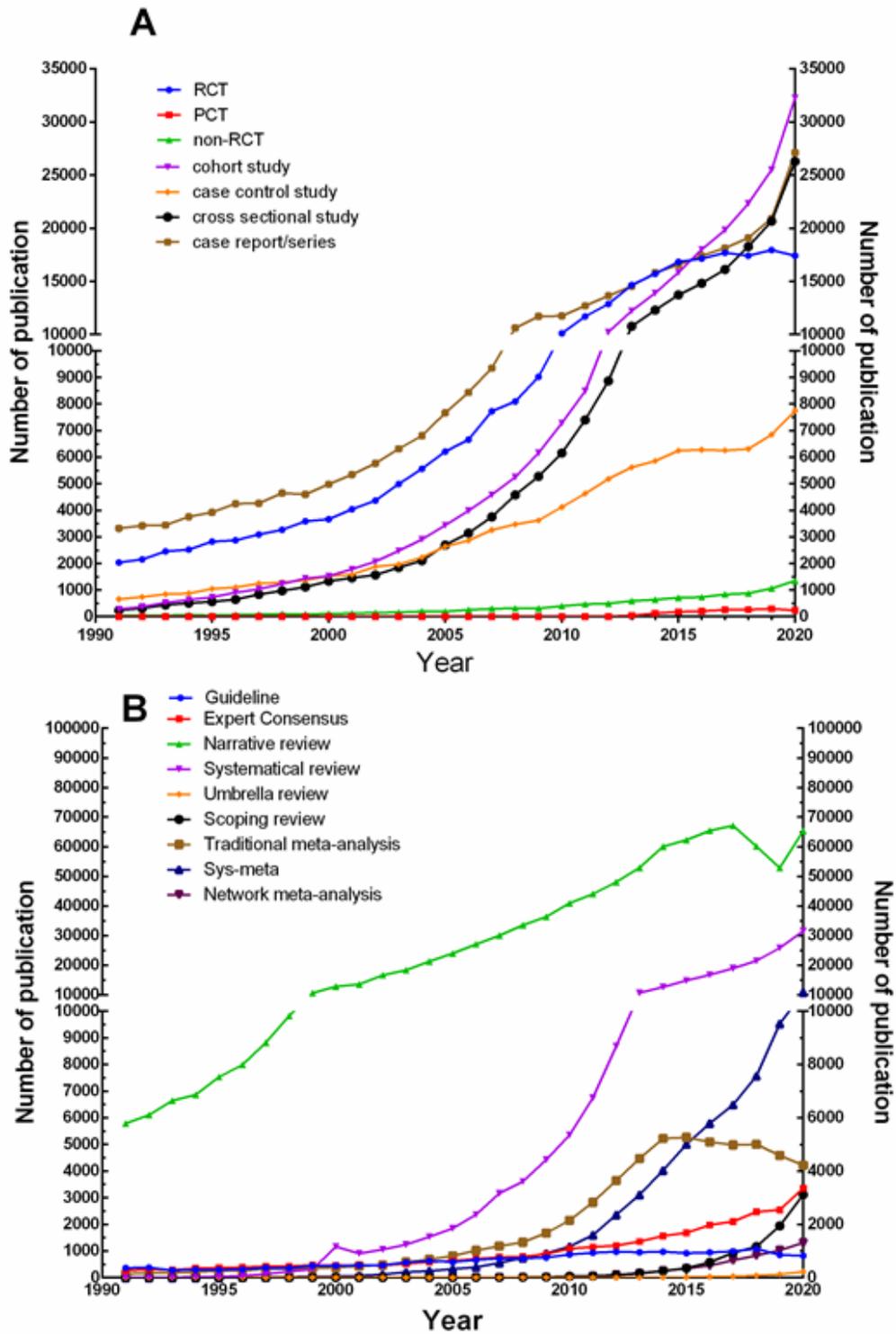


Figure 2

Quantity trend change (truncated distribution)

A: Trends in the number of original literature; B: Trends in the number of secondary literature

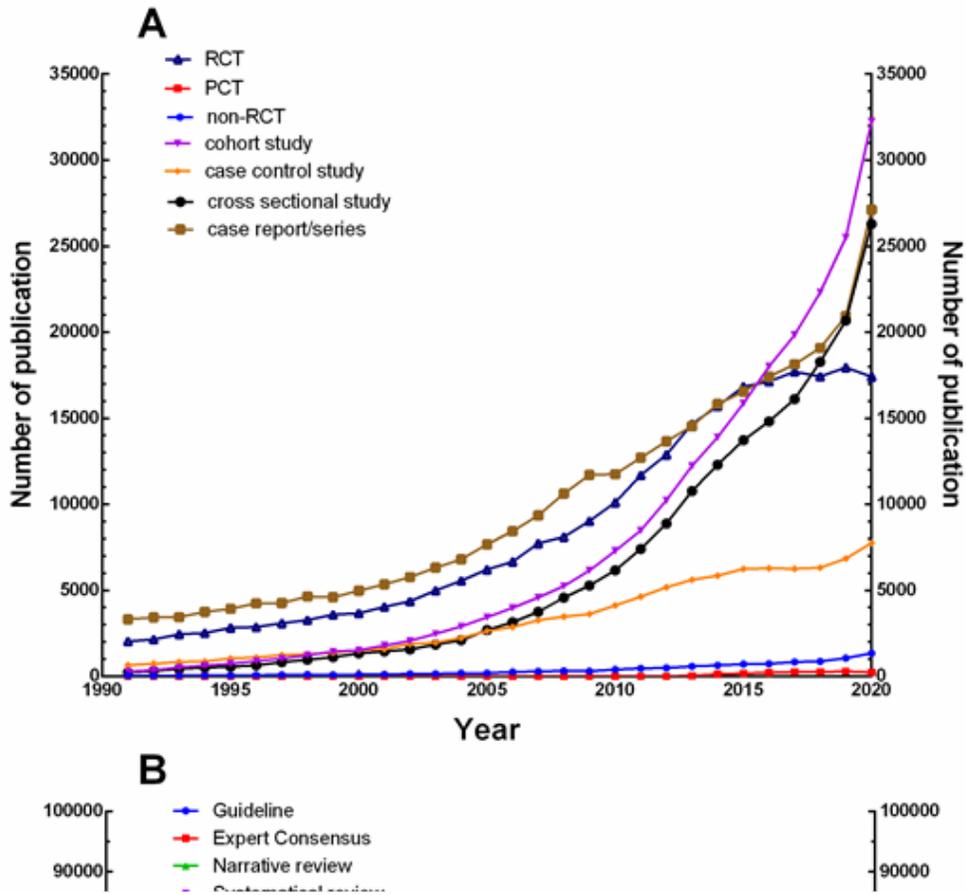


Figure 3

Quantity trend change (average distribution)

A: Trends in the number of original literature; B: Trends in the number of secondary literature

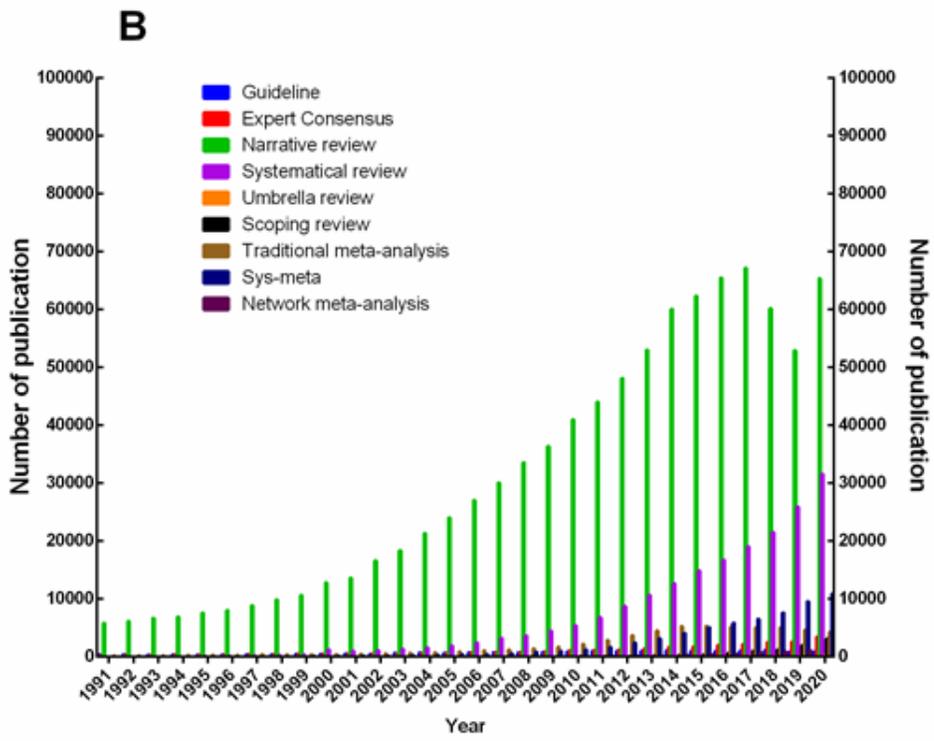
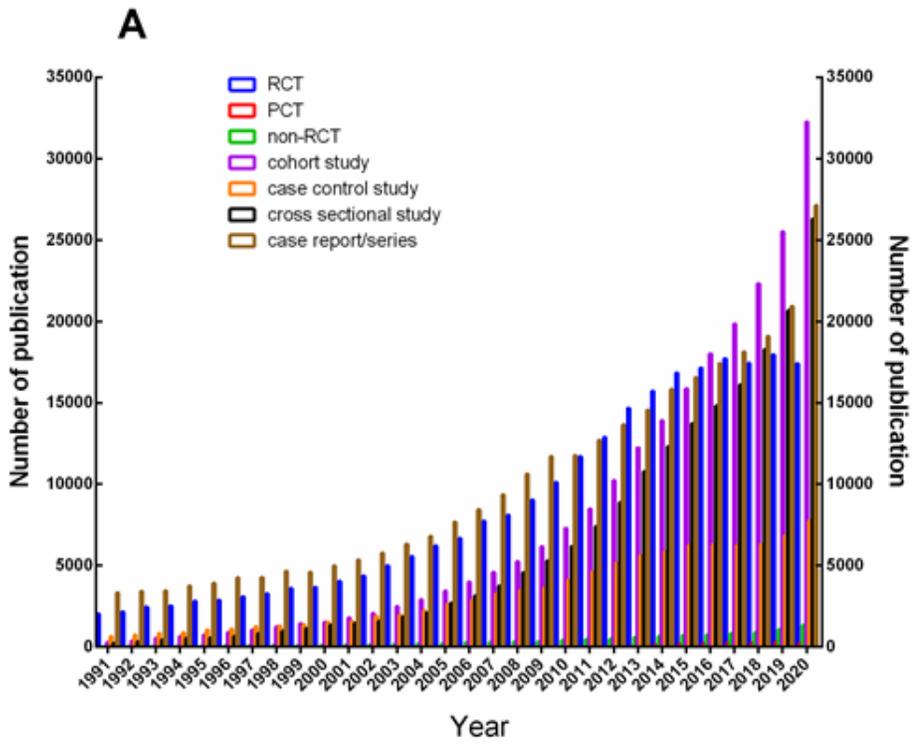


Figure 4

Quantity change trend of each study type

A: Trends in the number of original literature; B: Trends in the number of secondary literature

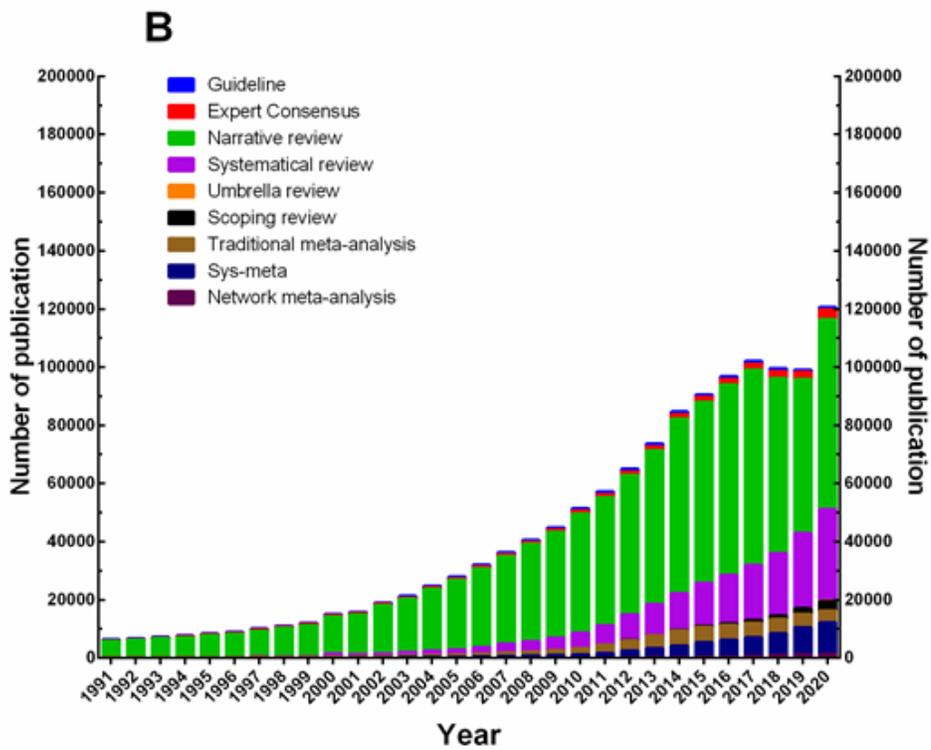
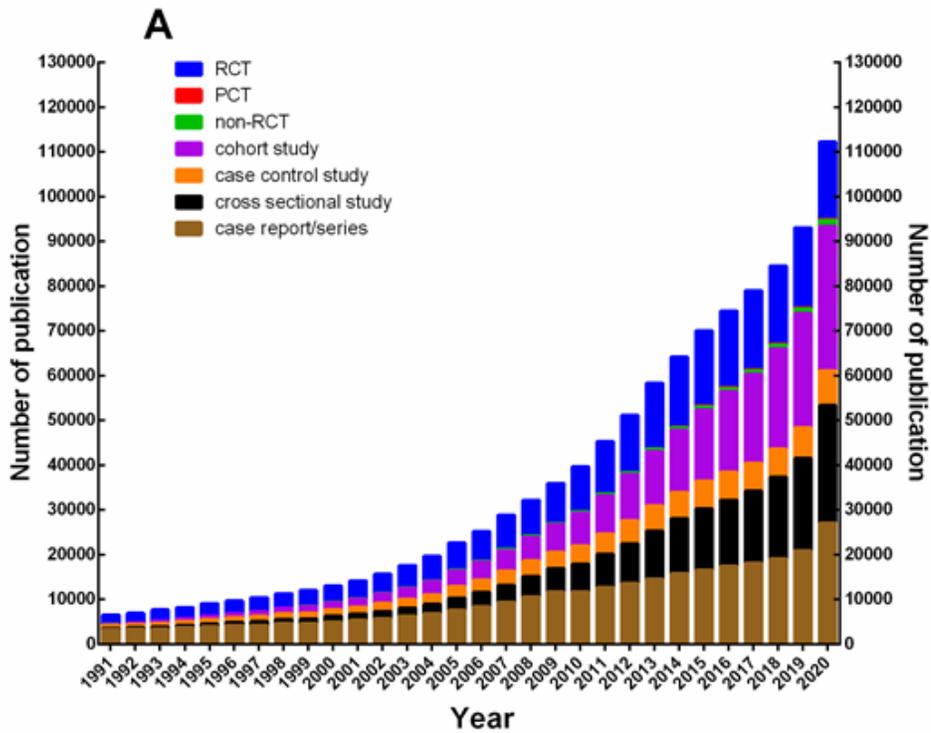


Figure 5

Quantity change trend of each study type

A: Trends in the number of original literature; B: Trends in the number of secondary literature

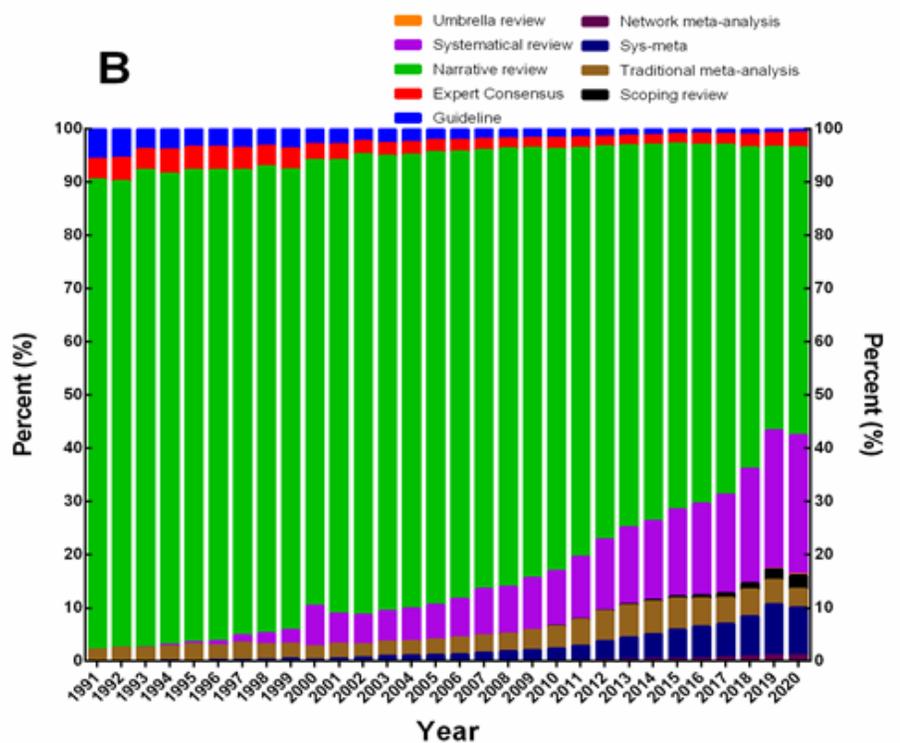
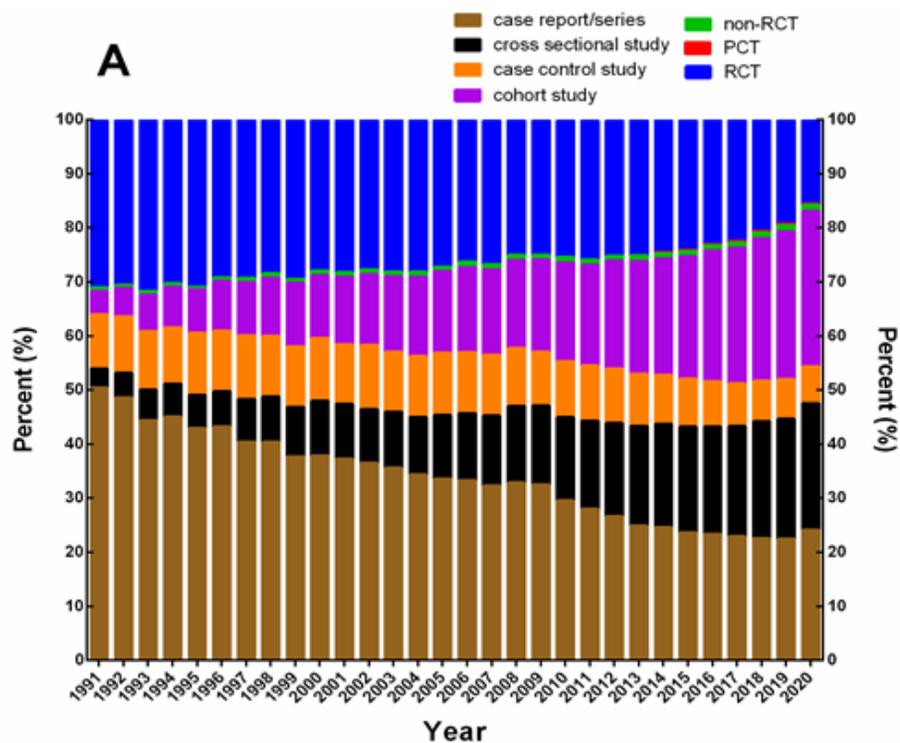


Figure 6

Proportion change of each research type

A: Trend change of original literature proportion; B: Trend change of secondary literature proportion

Figure 7

**Trend chart of growth rate of each study type**

A, B: change trend of annual growth rate of original literature; C, D: change trend of annual growth rate of secondary literature; E: Bar chart of the average annual growth rate of the original literature; F: bar chart of average annual growth rate of secondary literature

\*In this figure, due to the annual growth rate of some new research types in the original and secondary literature changes greatly, for the convenience of readers, we use two broken line graphs with different ordinate values to represent the annual growth rate of the original literature and secondary literature research according to the severity of the change range

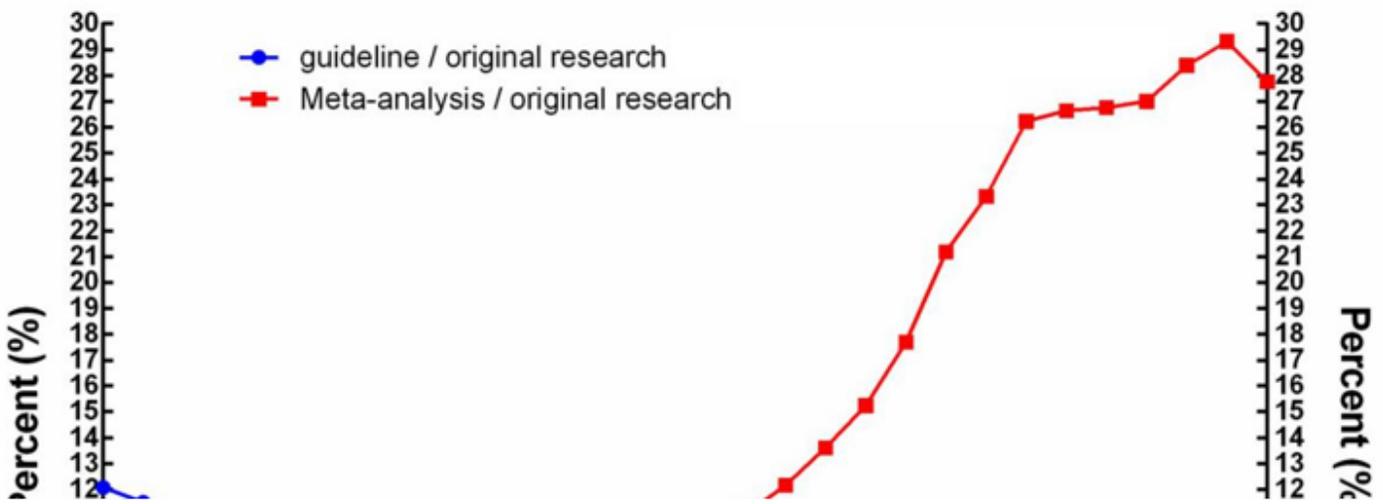


Figure 8

**Trend chart of proportion change between secondary literature and original literature (broken line chart)**

Blue: the ratio of guideline to the original study; Red: the ratio of meta-analysis to the original study

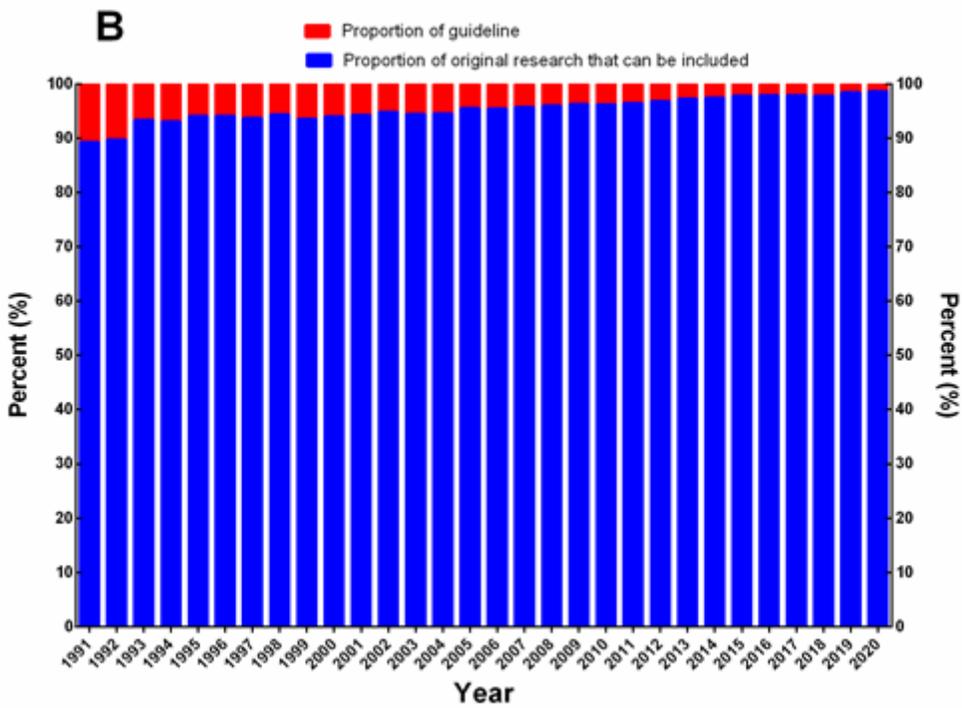
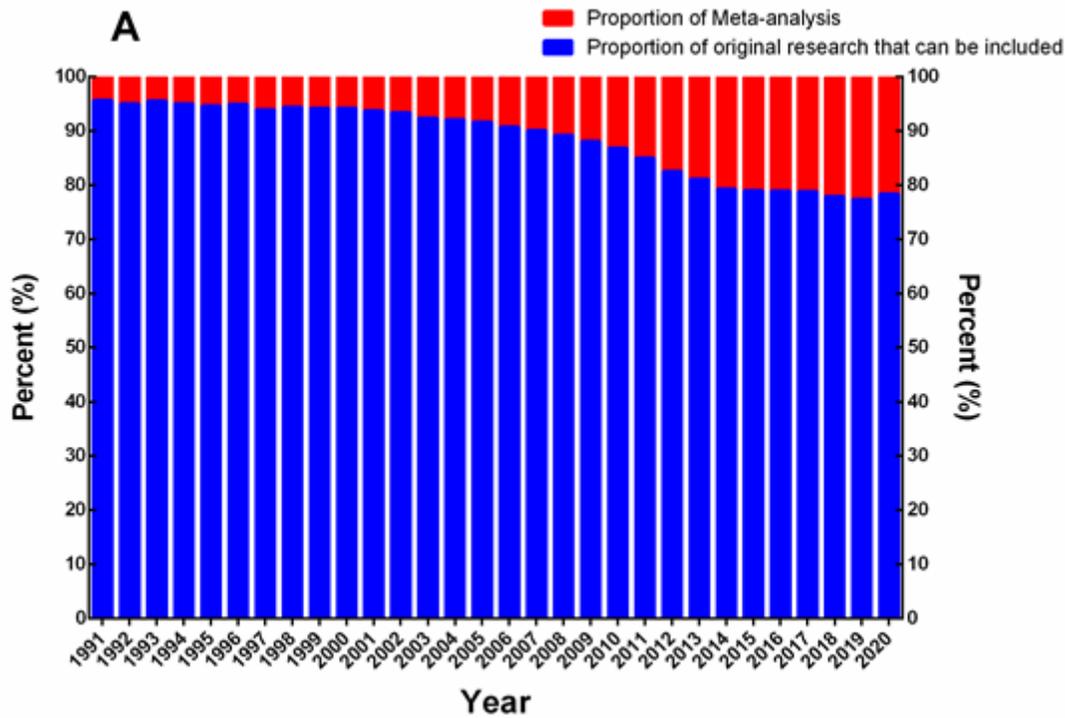


Figure 9

Trend chart of proportion change between secondary literature research and original research (bar chart)

Blue: the ratio of guideline to the included original study; Red: the ratio of meta-analysis to the included original study

## Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementarymaterials.docx](#)