




Analysis and prediction of research hotspots and trends in pediatric medicine from 2,580,642 studies published between 1940 and 2021

Ze-Ye Liu¹ · Yuan Huang² · Jing Xu³ · Li Xiang⁴ · Zhan-Hao Su⁵ · Yi-Wei Liu⁴ · Hao Zhang^{4,6} 

Received: 11 January 2023 / Accepted: 2 May 2023 / Published online: 9 June 2023
© Children's Hospital, Zhejiang University School of Medicine 2023

From 1990 to 2019, there has been a substantial increase in the global population from 5.3 billion to 7.7 billion (about 44.6%) [1, 2]. Simultaneously, the global death population among people under 20 years of age has increased from 6.22 million to 13.93 million, accounting for an increase of 124% [1–3]. This means that the mortality rate of children/adolescents is almost three times higher than the growth rate of the global population. Additionally, the disparities between adolescent mortality in different countries and sexes are widening [3, 4]. Given this alarming trend, scientific research is crucial for addressing the global health challenges of children/adolescents and can aid in decision-making.

The current priorities in funding and policies in the field of child health research are increasingly focused on innovation in an endeavor to improve clinical outcomes. The United States and the United Kingdom have established relevant working groups in this area and issued several guidelines [5, 6]. However, summarizing all existing research based on published literature is challenging, given the broad areas of research in pediatric medicine.

The journal classification systems currently used are neither well-refined to identify and classify all relevant studies nor capable of elaborating on the topics or concepts [7].

Semi-automated methods that identify key topics based on text analysis offer an alternative solution to this issue [8, 9]. The research and development of software such as “CiteSpace” and “VOSviewer” have demonstrated similar concepts [8, 9]. Researchers have previously summarized the hotspots and analyzed the research trend in cardiovascular medicine during 2004–2013 using artificial intelligence (AI)-based natural language processing (NLP) technology, involving the analysis of more than 400,000 literature reports [7]. Machine learning models have also been used for classifying over 200,000 pediatric research literature in the field of pediatrics [10]. However, these conclusions require updating and will benefit from meticulous, large-scale, and long-term analytical studies. Therefore, this study aimed to identify the themes and evolutionary trends in pediatric medical research by assessing studies published during 1940–2021 (particularly in the last 40 years) and evaluating the change in these trends over time.

The dataset used in this study included data on titles, abstracts, and keywords obtained from 2,580,642 pediatric medical publications from 1940 to 2021 via searching 14 pediatric medical search terms based on Medical Subject Headings (MeSH) phrases provided by PubMed, eliminating repetition, and using specific search terms as shown in Supplementary Table 1. All the titles and abstracts from these publications were used, and the time information and noun

✉ Hao Zhang
drzhanghao@126.com

¹ Department of Structural Heart Disease, National Center for Cardiovascular Disease, Fuwai Hospital, Chinese Academy of Medical Sciences and Peking Union Medical College, Beijing 100037, China

² State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Pediatric Cardiac Surgery Center, Fuwai Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College, Beijing, China

³ State Key Laboratory of Cardiovascular Disease, Fuwai Hospital, National Center for Cardiovascular Diseases, Fuwai Hospital, Chinese Academy of Medical Sciences, and Peking Union Medical College, Beijing, China

⁴ Heart Center and Shanghai Institute of Pediatric Congenital Heart Disease, Shanghai Children's Medical Center, National Children's Medical Center, Shanghai Jiaotong University School of Medicine, No. 1678 Dongfang Road, Shanghai 200127, China

⁵ Department of Cardiovascular Surgery, Guangdong Cardiovascular Institute, Guangdong Provincial People's Hospital, Guangdong Academy of Medical Sciences, Guangzhou 510060, China

⁶ Shanghai Clinical Research Center for Rare Pediatric Diseases, Shanghai Children's Medical Center, School of Medicine, Shanghai Jiao Tong University, Shanghai 200127, China

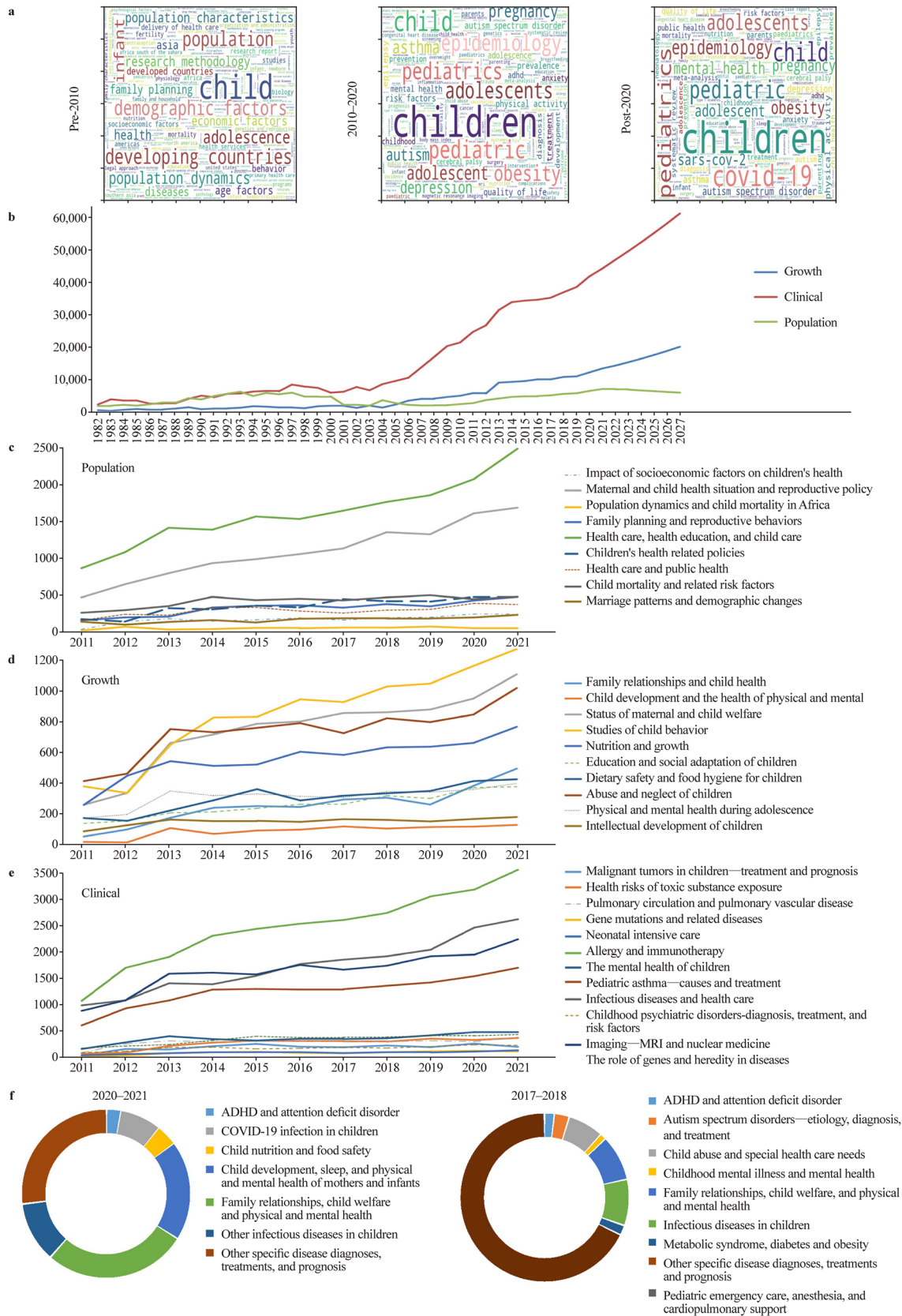


Fig. 1 Hot topics and trends in pediatric medicine research. **a** Frequency of each keyword in the research in the three periods; the higher the frequency, the larger the corresponding word. (The top five keywords with the highest frequency of occurrence in each period were marked in red.) **b** Trends in the number of studies in the three fields of clinical, population, and development during 1982–2021 and projections over the next 5 years. **c–e** Increasing trends in clinical, growth, and population studies in the last 10 years. Here, 9, 10, and 12 of the fastest-growing research hotspots are displayed (**c–e**, respectively). They have increased to at least double of original quantity in the past 10 years. **f** Distribution of the document clusters during 2017–2018 and 2020–2021. The top nine largest document clusters presented in 2017–2018 accounted for 98.1% of the total documents in that period. The top seven largest document clusters presented in 2020–2021 accounted for 97.9% of the total documents in that period. *COVID* coronavirus disease, *ADHD* attention deficit and hyperactivity disorder, *MRI* magnetic resonance imaging

phrases (pieces of text of various lengths) were extracted using the NLP framework developed by Python 3.9 software. The generic terms commonly used in most documents that generate a series of highly specific text fragments and themes to improve the accuracy and highlight class features were filtered out. Five experts from the field of pediatric medicine ranked the topics based on the first 30 text fragments representing the topics; they then cross-reviewed them and eventually merged similar topics. Two complementary approaches, latent dirichlet allocation (LDA) and the *k*-means clustering algorithm (*k*-means) were used to categorize the literature [11, 12]. Supplementary Fig. 1 illustrates the specific data analysis flow.

This study was conducted based on three time periods: pre-2010, 2010–2020, and post-2020. During these periods, word clouds were drawn for keywords, and the top five keywords with the highest frequency of occurrence (in a larger font on the figure) in each period were identified. For the pre-2010 period, the major keywords were “developing countries,” “population,” “demographic factors,” “infant,” and “adolescence”; for the 2010–2020 period, they were “pediatrics,” “epidemiology,” “adolescents,” “obesity” and “adolescent”. The post-2020 period keywords with the highest frequency were “coronavirus disease 2019 (COVID-19),” “pediatrics,” “adolescents,” “epidemiology” and “obesity”; common words such as “child” and “children” were not included (Fig. 1a).

In total, 111 pediatric medical topics were identified and are listed alphabetically in Supplementary Table 2. These topics were further categorized into “clinical,” “growth,” and “population” studies, e.g., the characteristics and treatment of pediatric malaria in clinical studies, cognitive development of children in developmental studies, and the health status of children in developing countries in population studies. AutoReg was used to predict the growth trend of the three categories of studies in the next 5 years based on the number of publications in the first 40 years (Fig. 1b)

[13]. The trend of clinical studies was observed to be rapidly increasing.

The “hotspots” of the significantly increased publications in clinical, growth, and population studies over the last decade are displayed in Fig. 1c–e, respectively. Allergy and immunotherapy studies were found to be extremely “popular” in clinical studies, which may be due to the high incidence of such diseases [2, 15]. More than 50% of children had at least one allergic disease, significantly reducing their quality of life [16]. Conventional fields of study including infectious diseases, congenital diseases, and pediatric imaging were observed to be expanding. Children’s psychological health, exposure to toxic substances, and the impact of coronavirus emerged as the latest research hotspots. In population studies, studies related to health education, healthcare, and fertility policies were extremely “popular.” This may be due to large-scale, comprehensive public health interventions can effectively improve the health of children [4, 17]. Our results indicated that the number of studies targeting socioeconomic factors and people in developing nations rapidly increased. Research on child behavior, family relationships and child health, child nutrition, child maltreatment, and neglect showed a continuous growth trend. Moreover, research on child development and physical and mental health grew rapidly. Supplementary Table 3 lists the comparison of the rapidly growing themes in 2021 with those in 2012. Here, the number and growth rate of publications is depicted.

A complementary clustering algorithm of LDA and the *k*-means approach was applied in two publication datasets, 2017–2018 and 2020–2021, leading to the generation of nine and seven large document clusters, respectively. Figure 1f shows a comparison of the document clusters between the two epochs. Infectious diseases in children, mental health, family relations, child welfare, specific disease diagnosis, and treatment play crucial roles in both periods. Research on family relations, child welfare, and physical and mental health was found to gain increased attention. COVID-19 infection in children was newly defined during 2020–2021. Moreover, research on child nutrition and food safety was prominent. Autism spectrum disorders, child maltreatment, and special healthcare needs were separately identified areas during 2017–2018; however, they were not recognized during 2020 and 2021.

The results obtained by the two methods were compared to determine the high consistency of the two methods (Supplementary Table 4), and detailed clusters for 16 document clusters were obtained through the LDA method. For example, the cluster “family relationships, child welfare, and physical and mental health” exhibited different areas of focus, i.e., child behavior and parent–child relationships, child-rearing practices and potential impact, education, and

social adaptation of children. Moreover, the “infectious disease” cluster was further divided into malaria, parasites, viruses, and other fields.

In addition to focusing on the primary research areas, understanding the significance of research findings in fields with a lower focus is essential. Inverse document frequency (IDF) calculations were performed for keywords by year of research [14] to observe the research trends in relatively “niche” areas. Keywords with extensive research focus in a specific year compared with that in previous years were identified. We calculated the annual IDF for each keyword. This value indicated the year in which a particular keyword appeared most frequently to locate those parts that greatly vary over time despite their small number, irrespective of whether it may have appeared in previous research. Supplementary Table 5 indicates the keywords that have appeared most frequently in their field each year (2012–2021), including keywords like “ultrasound contrast agents,” “2019 novel coronavirus disease,” and “oral habits” in the last 3 years. A complete list of keywords from the last 76 years is presented in Supplementary Table 6.

The NLP techniques enable the identification of themes and detect clusters across pediatric medical publications, allowing us to view trends in pediatric medical research at the macro level. The research hotspots in the field of pediatric medicine are gradually changing over time, particularly in the last decade. A previous study used a dataset of publications retrieved from PubMed to classify research results in pediatrics, and this work required extensive expert review of a subset of published articles [10]. Despite the use of advanced automated analysis techniques, experts still played a key role in interpreting and linking concepts, as well as validating results. Moreover, experts refer to terms on the list of classical specificity in the field and incorporate their insights with pediatrics classification while naming current topics and clusters. This approach has aided in the identification of many emerging research areas, such as “health inequities and child mortality study” and “abuse and neglect of children” [10, 18].

This study found that allergy and immunotherapy were the most prominent and rapidly expanding areas in clinical research, which could be due to the high prevalence of diseases and their substantial impact on life [2, 15, 16]. Furthermore, it may indicate the increased number of ongoing clinical trials in this field for exploring novel therapeutic tools. Infectious diseases continue to dominate pediatric medicine, indicating that children, as immunocompromised individuals, remain the focus of prevention and control of infectious diseases. In response to the new coronavirus pandemic, coronavirus-related studies in the field of pediatric medicine have emerged as research groups, demonstrating the high sensitivity of the analytical approach in this study.

In recent years, there has been an increasing focus on research related to child behavior, women’s welfare, child

maltreatment, physical and mental health, and social family factors. This reflects a positive trend towards providing comprehensive care and attention to the physical and psychological well-being of children. Additionally, the emphasis on family factors is consistent with the current trend transferring the shift in focus of childcare away from hospitals and directing it toward communities and families [19]. Child nutrition is a hot topic and the number of related studies is growing [20]. This may be due to that some communities and families continue to struggle to provide safe and adequate nutrition to children. The number of researches on population health conditions, health education, and fertility policy, has been rapid growth over the last decade as the related clinical studies have increased. Meanwhile, research topics are gradually shifting towards developing nations such as Africa.

Traditional research on specific childhood diseases remains noteworthy, with an emphasis on the importance of evidence-based treatment guidelines and outcome studies. The rapid growth of topics such as health economics and quality of medical services suggests current concerns about long-term prognosis and disease care. The importance of preventive medicine is being highlighted by increased research on risk factors. Notably, areas such as metabolic syndrome, environmental pollution, and obesity have shown an increasing trend in population, clinical, and growth studies, implying that pediatric medicine research is highly susceptible to cross-domain integration and concatenation.

In conclusion, our findings suggest that pediatric medicine is currently promoting the development of overall physical and mental health in children. While clinical research focuses on infections and immune-related diseases, growth and population studies are increasingly concerned with the development of physical and mental health and welfare in children. Furthermore, there has been an increased focus on the impact of food safety and family factors on the health of children. The requirement for the overall healthy development of children, new technologies, and social support are the potential driving factors for these changes [21]. The results of our study, along with expert insights, funding allocation, and policy formulation, will help pave the way for future research.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s12519-023-00731-9>.

Acknowledgements We would like to thank all the subjects who were involved in this study for their participation and co-operation.

Author contributions ZH: conceptualization, writing–review and editing. LZY, HY, XJ: formal analysis, writing–original draft, visualization. XL, SZH, LYW: writing–review and editing. All the authors approved the final version of the manuscript. LZY, HY and XJ contributed equally to the work.

Funding This study was supported by the National Natural Science Foundation of China (Grant Nos. U20A2018 and 31971048).

Data availability The datasets generated and analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest No financial or non-financial benefits have been received or will be received from any party related directly or indirectly to the subject of this article.

Ethics approval Not applicable.

References

- Institute for Health Metrics and Evaluation. GBD results tool 2019; 2019. <https://ghdx.healthdata.org/gbd-results-tool>. Accessed 26 Dec 2022.
- Diseases GBD, Injuries C. Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2020;396:1204–22.
- Collaborators GBDAM. Global, regional, and national mortality among young people aged 10–24 years, 1950–2019: a systematic analysis for the Global Burden of Disease Study 2019. *Lancet*. 2021;398:1593–618.
- Deshmukh V, Lahariya C, Krishnamurthy S, Das MK, Pandey RM, Arora NK. Taken to health care provider or not, under-five children die of preventable causes: findings from cross-sectional survey and social autopsy in rural India. *Indian J Community Med*. 2016;41:108–19.
- Cathie K, Sutcliffe AG, Bandi S, Coghlan D, Turner SW, Powell C. Priorities for child health research across the UK and Ireland. *Arch Dis Child*. 2022;107:474–8.
- Boat TF, Whitsett JA. How can the pediatric community enhance funding for child health research? *JAMA Pediatr*. 2021;175:1212–4.
- Gal D, Thijs B, Glanzel W, Sipido KR. Hot topics and trends in cardiovascular research. *Eur Heart J*. 2019;40:2363–74.
- Synnestvedt MB, Chen C, Holmes JH. CiteSpace II: visualization and knowledge discovery in bibliographic databases. *AMIA Annu Symp Proc*. 2005;2005:724–8.
- van Eck NJ, Waltman L. Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*. 2010;84:523–38.
- Levy-Mendelovich S, Barbash Y, Budnik I, Levy-Erez D, Somech R, Soffer S, et al. Pediatric literature trends: high-level analysis using text-mining. *Pediatr Res*. 2021;90:212–5.
- Kozłowski D, Semeshenko V, Molinari A. Latent Dirichlet allocation model for world trade analysis. *PLoS ONE*. 2021;16:e0245393.
- Chen ZL. Research and application of clustering algorithm for text big data. *Comput Intell Neurosci*. 2022;2022:7042778.
- Shih H, Rajendran S. Comparison of time series methods and machine learning algorithms for forecasting taiwan blood services foundation's blood supply. *J Healthc Eng*. 2019;2019:6123745.
- Naeem MZ, Rustam F, Mehmood A, Mui Zzud D, Ashraf I, Choi GS. Classification of movie reviews using term frequency-inverse document frequency and optimized machine learning algorithms. *PeerJ Comput Sci*. 2022;8:e914.
- Wu CY, Huang HY, Pan WC, Liao SL, Hua MC, Tsai MH, et al. Allergic diseases attributable to atopy in a population sample of Asian children. *Sci Rep*. 2021;11:16052.
- Haanpaa L, Af Ursin P, Nermes M, Kaljonen A, Isolauri E. Association of allergic diseases with children's life satisfaction: population-based study in Finland. *BMJ Open*. 2018;8:e019281.
- Khanna R, Hota P, Lahariya C. Health research strengthening and operational research needs for improving child survival in India. *Indian J Pediatr*. 2010;77:291–9.
- Dupuis LL, Robinson PD, van de Wetering MD, Tissing W, Seelisch J, Digout C, et al. Lexicon for guidance terminology in pediatric hematology/oncology: a white paper. *Pediatr Blood Cancer*. 2020;67:e28170.
- Lahariya C. Health and wellness centers to strengthen primary health care in India: concept, progress and ways forward. *Indian J Pediatr*. 2020;87:916–29.
- Norris SA, Frongillo EA, Black MM, Dong Y, Fall C, Lampl M, et al. Nutrition in adolescent growth and development. *Lancet*. 2022;399:172–84.
- Lahariya C. 'Ayushman Bharat' program and universal health coverage in India. *Indian Pediatr*. 2018;55:495–506.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.