Original Article

DOI: 10.5582/irdr.2024.01062

A bibliometric study of rare diseases in English and Chinese databases from 1985 to 2024 based on CiteSpace

Mengru Tian[1], Shuyi Wang[1], Zhihui Meng[1], Xin Zhang[2], Jinxiang Han[1,3,*], Yanqin Lu[1,3,*]

[1] Key Laboratory for Biotech-Drugs of National Health Commission, Key Laboratory for Rare & Uncommon Diseases of Shandong Province, Biomedical Sciences College & Shandong Medicinal Biotechnology Centre, Shandong First Medical University & Shandong Academy of Medical Sciences, Ji'nan, China;
[2] Digestive Diseases Hospital of Shandong First Medical University, Ji'ning, China
[3] Department of Endocrinology and Metabology, The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, Ji'nan, China.

SUMMARY: This study utilizes CiteSpace (version 6.2.R3) to visually analyze literature related to rare diseases, summarizing the current research status and hotspots in the field. The goal was to provide broader perspectives and references for researchers in rare diseases. A comprehensive search for relevant literature in the rare diseases domain was conducted through the China Knowledge Network (CNKI) and Web of Science (WOS), spanning the years 1985 to 2024. Then, CiteSpace software was utilized to create a visual map of the annual publication volume, authors, institutions, keywords, and other content. After screening, 2,293 Chinese and 2,262 English articles were included in the study. Over the last several decades, the diagnosis and treatment of rare diseases have been a common research focus in both China and foreign countries, but there is a significant research depth and breadth gap. In China, there is a shortage of core authors and high-quality literature, and the level of collaboration among research teams is significantly lower compared to the robust international cooperation between authors and institutions. High-frequency and central keywords in the field include "orphan drugs", "children", and "genetic mutations", reflecting research hotspots in this domain. Research on rare diseases has been increasing annually, with key directions focusing on orphan drug development, novel therapeutic agents, genetic therapies, and healthcare security. In the research field of rare diseases, emphasis should be placed on early detection, early prevention, and early treatment. The application of genetic diagnostic techniques in clinical practice will have a broader prospect. This will be one of the direction for future research in this area.

Keywords: rare disease, visual analysis, knowledge map, CiteSpace

1. Introduction

Rare diseases are distinguished by their uncommon occurrence or low incidence rates. There is a lack of uniformity globally in the criteria used to define the incidence of rare diseases, with no universally accepted standard currently in place. The World Health Organization (WHO) categorizes rare diseases as those with a prevalence ranging from 0.65 to 1 per 1000 individuals (*1*). In the "*China Rare Disease Definition Research Report 2021*", published in 2021, China has established specific criteria for identifying rare diseases, proposing that a condition be classified as rare if it meets any of the following thresholds: a neonatal incidence rate below 1/10,000, a prevalence rate below 1/10,000, or a total affected population less than 140,000 (*2*). Rare diseases are mostly caused by specific gene mutations, often involving multiple organs and systems of the human body, showing a chronic, progressive, and consumptive development. They are characterized by a low incidence rate for individual diseases and a small proportion of the patient population (*3*). This results in a multitude of interrelated issues in the diagnosis and treatment of rare disease patients, including the intricacy of diagnosing these conditions, a high incidence of misdiagnosis, and a low prevalence of standardized therapeutic approaches (*4*). Additionally, the development of orphan drugs is marked by significant costs, with only a limited number of these medications successfully making it to the market.

Due to the enormous population base in China, the total number of patients with rare diseases is expected to exceed 20 million, creating a situation where "rare diseases are not uncommon" (*5*). These patients urgently

need effective diagnostic and treatment methods. Therefore, Chinese scholars attach great importance to rare disease research, investing more resources and funds in this field to bring about new discoveries and treatment strategies. This not only aids in promoting the government's formulation of public health policies targeting rare diseases, but also ensures the protection of patients' rights and interests. In addition, it can facilitate global cooperation in rare disease research, enhancing China's academic influence in this field. Systematically organizing and summarizing the literature related to rare diseases, and encapsulating research progress, current status, hot topics, and potential trends in the field of rare diseases plays a positive role in guiding researchers in selecting the right research direction and promoting long-term development of the field. The WOS and CNKI are the most comprehensive core journal indexing databases both in China and globally. They can help identify trends in scientific research from different linguistic and national perspectives. Based on this, this study uses CiteSpace software to conduct a bibliometric analysis of research related to rare diseases from different perspectives. The study will analyze the research in the field of rare diseases and draw a knowledge map, providing valuable references for rare disease research in China.

2. Data sources and research methods

2.1. Literature retrieval and data screening

A Chinese literature search was conducted on CNKI using the search query "rare disease" or "infrequent disease", covering the period from January 1985 to November 2024. This search yielded a total of 4,176 original Chinese articles. An English literature search was performed using the WOS database with the search query "rare disease" or "rare diseases", covering the same period from 1985 to 2024, resulting in a total of 4,981 original English articles.

To ensure the quality of the literature retrieval and the scientific validity of subsequent bibliometric analysis, data screening was conducted. All duplicate entries were removed, and editorial materials, book chapters, letters, and other types of literature were excluded. Additionally, any literature with incomplete or missing author information was excluded. After multiple rounds of screening, 2,293 Chinese literature records and 2,362 English literature records were retained (Figure 1).

2.2. Research methods

The search results from CNKI were exported in "Refworks" format, whereas the results from WOS were exported using "Plain Text - Full Record and References" as the data source, information on authors, research institutions, countries, publication years, and keywords

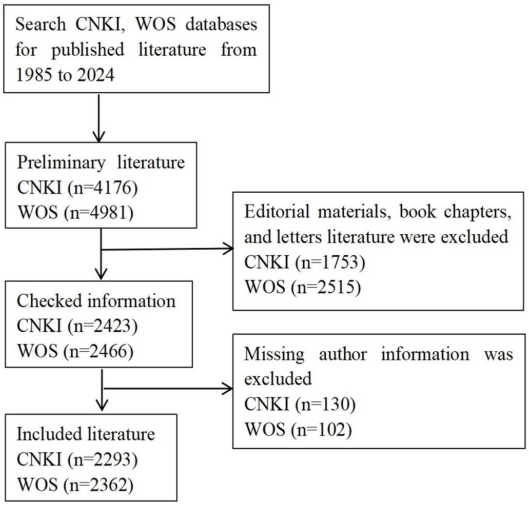

Figure 1. Flowchart of literature search and data incorporation.

were included. CiteSpace software was then used for visual analysis. The parameter settings include a "Time Slicing" span from "1985.1" to "2024.11" and a "Years Per Slice" of 1, with node types selected as author, institution, and keyword for analysis. Cosine similarity was used for connection strength, and the Pathfinder method was employed for pruning.

3. Results

3.1. Statistics of article publication volume

Through the statistical analysis of the annual publication volume, we can clearly understand development status and future trends of a specific field in China and other countries. The overall trend of the annual publication volume in the field of rare diseases shows a wave-like upward trend, since 2012, this trend has become increasingly evident. According to the CNKI database, the number of publications has risen from 47 in 2012 to 240 in 2023. Similarly, the WOS database reveals an increase of 174 articles between 2012 and 2023. Despite minor fluctuations in certain years, the upward trend in research output is apparent. It's important to note that our data for 2024 is incomplete, as we haven't included publications from November 2024 onwards, which might make the 2024 figures seem lower in comparison. Overall, rare diseases have become a significant area of research over the past decade, and it is crucial for modern scholars to continue investing in this field (Figure 2).

3.2. Distribution of countries/regions

We used CiteSpace to produce a country/region co-occurrence map (Figure 3). The map consists of 90 nodes and 902 connected lines with a network density of 0.2252, indicating that 90 countries/regions contribute to the field for rare diseases. In the WOS database, the country

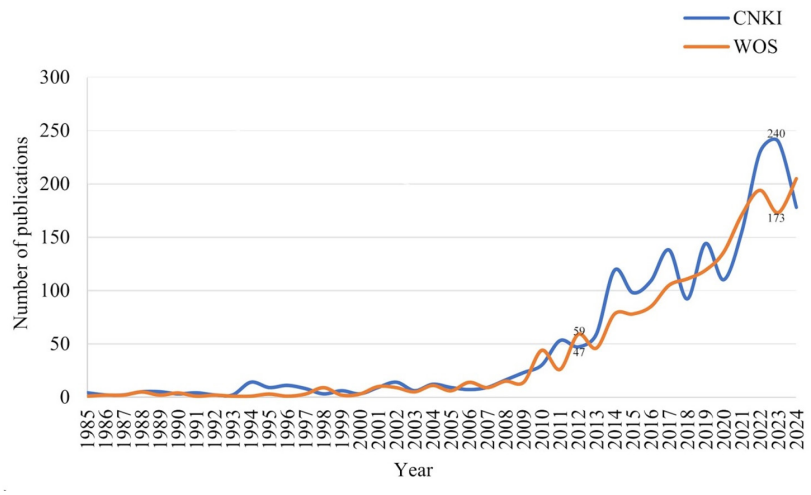

Figure 2. Annual number of publications for rare disease-related studies.

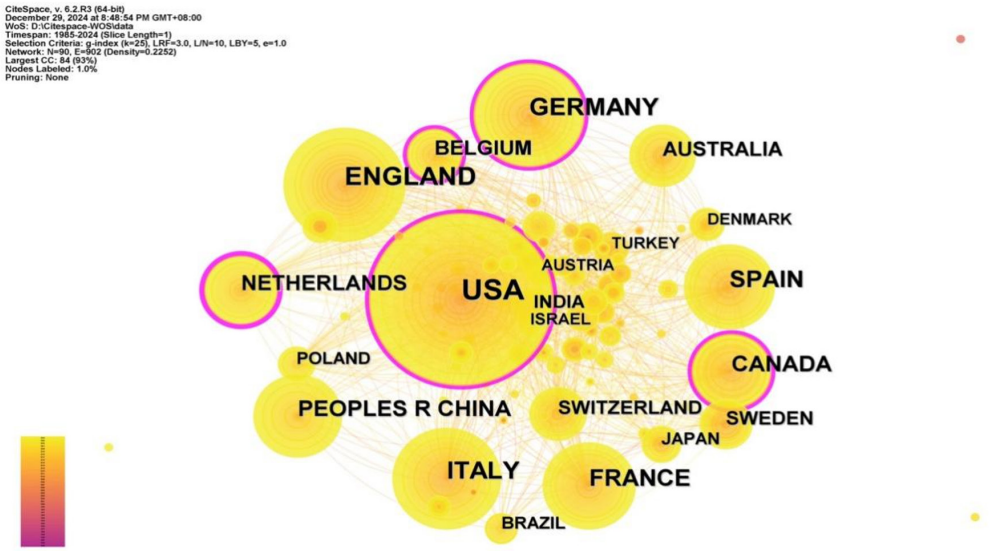

Figure 3. Network of collaborating countries.

with highest number of publications is the United States, with 1,518 related papers. Second is the United Kingdom (719 articles), followed by Italy (592 articles), Germany (492 articles), and China is ranked relatively lower (201 articles). There is a significant gap between the United States, which ranks first, and other countries, which indicates that the United States also occupies a core position in the field of rare disease research (Table 1). The purple nodes in the co-occurrence map indicate intermediary centrality above 0.1, with the Netherlands (0.22), Belgium (0.14), Germany (0.1), Canada (0.1), and the US (0.11) playing a "bridging" role in the field. Considering a combination of the volume of publications and intermediary centrality, the US and the UK have contributed the most to the field of pharmacovigilance for rare diseases. From the perspective of distribution of publishing journals, the United States and other developed European countries occupy a dominant position (Table 2).

3.3. Institution analysis

Using CiteSpace6.2.R3 software for institutional analysis, the distribution of publishing institutions is displayed (Figure 4, A and B). Nodes in the maps represent publishing institutions, with lines between nodes indicating collaborations between different institutions. Thickness of the lines reflects the degree of collaboration. The visual map of publishing institutions in the CNKI database includes a total of 360 nodes and 117 lines, resulting in a network density of 0.0018. The institution with the highest number of publications in China is Peking Union Medical College Hospital, Chinese Academy of Medical Sciences, with a total of 162 articles (Table 3). As a top medical and research center in China, it collaborates with the Institute of Basic

Table 1. The top ten countries/regions with the most publications quantity and betweenness centrality

Rank	Publications	Countries	Centrality	Countries
1	1518	USA	0.22	NETHERLANDS
2	719	ENGLAND	0.14	BELGIUM
3	592	GERMANY	0.11	USA
4	492	ITALY	0.1	GERMANY
5	491	FRANCE	0.1	CANADA
6	377	SPAIN	0.09	POLAND
7	318	CANADA	0.08	FRANCE
8	230	NETHERLANDS	0.08	AUSTRALIA
9	201	CHINA	0.06	ENGLAND
10	174	AUSTRALIA	0.06	ITALY

Table 2. Top 10 journals by publication volume

CNKI		WOS	
Publication Title	Number of publications	Publication Title	Number of publications
Pharmaceutical Economy Newspaper	82	Value in health	82
Journal of Rare and Uncommon Diseases	78	Orphanet Journal of Rare Diseases	78
China Hospital CEO	67	European Journal of Human Genetics	67
Chinese Journal of New Drugs	64	Molecular Genetics and Metabolism	64
International Journal of Pharmaceutical Research	54	American Journal of Gastroenterology	54
Journal of Clinical Pediatrics	51	American Journal of Respiratory and Critical Care Medicine	51
Chinese health	48	Advances in Experimental Medicine and Biology	48
Health For Everyone	45	Expert Opinion on Orphan Drugs	45
Health News	40	Chest	40
Progress in Pharmaceutical Sciences	40	International Journal of Environmental Research and Public Health	40

Medical Sciences of the Chinese Academy of Medical Sciences, Tsinghua University, and other prestigious institutions to jointly establish the National Key Laboratory for Difficult and Severe Diseases and Rare Diseases, as well as the Lymphangioleiomyomatosis/ Tuberous Sclerosis Complex Rare Disease Special Fund, collectively promoting research on rare diseases in China. In contrast, the visual map of publishing institutions in the WOS database comprises 280 nodes and 478 lines, with a network density of 0.0122. The institution with the highest number of publications overseas is the National Institute of Health and Medical Research. The institution was established in 1964 and boasts abundant research resources and a top-notch research team. These maps reveal a stark contrast in the collaborative landscape of rare disease research: institutions from other countries exhibit a high level of connectivity, whereas Chinese research institutions appear more isolated with limited collaborative efforts. This disparity underscores the imperative for fostering stronger international cooperation in this field.

3.4. Author analysis

The author's co-occurrence map visually depicts the core researchers in the field and the intensity of their collaborations. Each node represents an author, while the links between authors indicate collaboration. The color palette of nodes and links visually depicts starting year of collaborations, with maroon marking the beginning in 1985 and progressing through various hues to yellow for the year 2024. The size of each node is directly proportional to number of co-authors associated with each author, while the size of the node's circle reflects total number of publications by that author. The lines connecting nodes illustrate degree of collaboration between authors.

In the map, collaborative groups in the field can be identified. In the Chinese literature, the primary collaborative group is represented by Shuyang Zhang and Bo Zhang, with a total of 436 nodes, 308 connections, and a network density of 0.0032 (Figure 5A). They are all focused on the field of rare diseases in China, conducting research on pathogenesis of rare diseases and development of new drugs. In the English literature, Taruscio D, Graessner H and Boycott KM were the top three authors in terms of publication volume. The three authors constituted the center of the collaboration network, with a node count of 413, a connection count of 394, and a network density of 0.0046 (Figure 5B). These three experts have all made significant contributions to the field of rare diseases. The top ten authors are ranked in order of frequency of appearance in the literature (Table 4).

3.5. Keyword analysis

Figure 4. A network knowledge map of institutions in the field of rare diseases obtained using CiteSpace6.2.R3 software based on the CNKI database (A) and WOS database (B), respectively.

3.5.1. Keyword co-occurrence analysis

Pathfinder algorithm along with pruning of merged and sliced networks were selected in the software to obtain the co-occurrence graph of rare disease keywords by analyzing keywords from the literature in the WOS and CNKI databases separately (Figure 6, A and B). Keywords serve as a high-level summary of core topics and content of the literature. This graph reflects the frequency of occurrence and research hotspots of keywords in the field (*6*). The size of the squares in the graph indicates the frequency of keyword occurrence, with larger squares representing a higher frequency of central keywords. The CNKI database shows that the top five high-frequency keywords in China are: rare diseases (frequency = 675), orphan drugs (frequency = 164), rare drugs (frequency = 61), diagnosis (frequency = 45) and medical insurance (frequency = 40). The top five foreign high-frequency keywords are: rare diseases (frequency = 449), rare disease (frequency = 282), diagnosis (frequency = 110), orphan drugs (frequency = 92) and children (frequency = 87). It is clear that orphan medications, diagnostic methodologies, and therapeutic approaches are shared priorities in the realm of rare disease research. Centrality is a key indicator for analyzing the importance of keywords. Nodes with a centrality higher than 0.1 are regarded as key nodes. Their centrality is organized in descending order, with comprehensive details provided in Table 5. It can be learned that the centrality ranking first in the CNKI database is "rare medications", followed

Table 3. Top 10 major institutions for rare disease Chinese and English Databases

CNKI		WOS	
Publication institution	Number of publications	Publication institution	Number of publications
Peking Union Medical College Hospital, Chinese Academy of Medical Sciences	162	Institut National De La Sante Et De La Recherche Medicale	220
China Pharmaceutical University	122	University Of London	169
Beijing University	56	Assistance Publique Hopitaux Paris APHP	166
Shenyang Pharmaceutical University	50	National Institutes Of Health NIH USA	152
Shandong University	41	Universite Paris Cite	133
Fudan University	37	Harvard University	125
National Medical Products Administration	36	University College London	177
Huazhong University Of Science And Technology	32	Ciber Centro De Investigacion Biomedica En Red	108
Peking Union Medical College Hospital	28	Ciberer	95
Tsinghua University	25	University Of Toronto	89

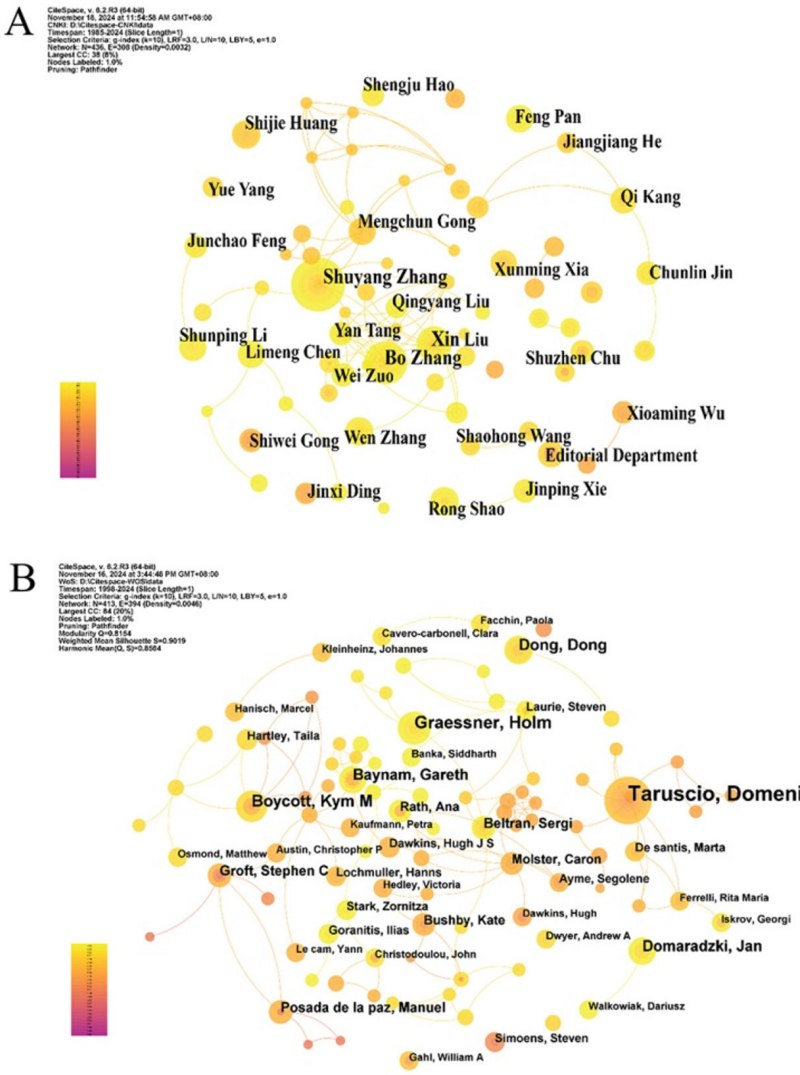

Figure 5. A network knowledge map of authors in the field of rare diseases obtained using CiteSpace6.2.R3 software based on the CNKI database (A) and WOS database (B), respectively.

by "etiologies", with research focusing on development of orphan drugs. In the WOS database, the centrality ranking first is "epidemiology", followed by "mutations", with research emphasis on gene therapy for rare diseases. This demonstrates the differences in research directions between China and the international community.

3.5.2. Keyword clustering analysis

Table 4. Statistics of the high-frequency authors in the field of rare diseases

CNKI				WOS			
Count	Centrality	Year	Author	Count	Centrality	Year	Author
43	0.01	2017	Shuyang Zhang	25	0.02	2012	Taruscio, Domenica
26	0	2019	Bo Zhang	13	0	2022	Graessner, Holm
18	0	2019	Xin Liu	11	0.01	2014	Boycott, Kym M
13	0	2016	Shijie Huang	10	0	2020	Dong, Dong
11	0	2022	Shunping Li	9	0	2021	Domaradzki, Jan
11	0	2019	Feng Pan	9	0.04	2017	Baynam, Gareth
11	0.01	2017	Mengchun Gong	7	0.03	2020	Beltran, Sergi
11	0	2021	Wen Zhang	7	0	2010	Posada de la paz, Manuel
11	0	2020	Rong Shao	7	0.01	2010	Groft, Stephen C
10	0	2022	Liming Chen	6	0.04	2014	Bushby, Kate
9	0	2017	Xunming Xia	6	0	2017	Rath, Ana
9	0	2018	Qi Kang	5	0	2010	Simoens, Steven
9	0	2010	Editorial Department	5	0	2023	Goranitis, Ilias
8	0	2022	Jinping Xie	5	0	2008	Ayme, Segolene
6	0	2023	Yan Tang	5	0.01	2023	Stark, Zornitza

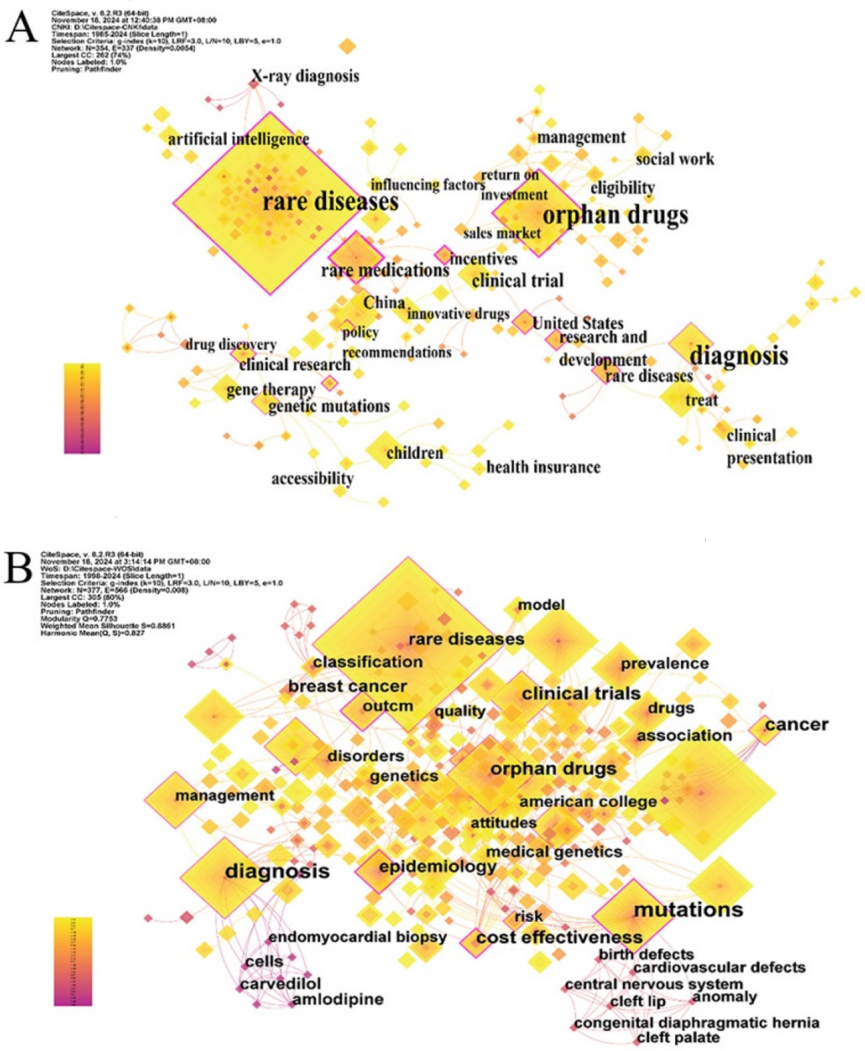

Figure 6. A network knowledge map of keywords in the field of rare diseases obtained using CiteSpace6.2.R3 software based on the CNKI database (A) and WOS database (B), respectively.

Table 5. Highly centralized keywords in the field of rare diseases

CNKI				WOS			
Count	Centrality	Year	Keywords	Count	Centrality	Year	Keywords
61	0.78	2002	rare medications	25	0.32	2008	epidemiology
6	0.6	2007	incentives	85	0.28	2005	mutations
675	0.53	1990	rare diseases	11	0.28	2010	Cost-effectiveness
164	0.42	2010	orphan drugs	9	0.19	2009	design
13	0.3	2004	United States	110	0.18	2000	diagnosis
9	0.27	2011	Research and development	54	0.16	2007	clinical trials
4	0.26	2022	medication management	37	0.16	2010	gene
15	0.24	2002	rare diseases	92	0.13	2008	orphan drugs
5	0.24	2008	new drug	54	0.12	2012	quality of life
14	0.23	2005	clinical research	52	0.12	2012	management
45	0.19	2001	diagnosis	14	0.12	1999	cancer
15	0.15	2011	genetic mutations	449	0.1	2008	rare diseases

Keyword clustering analysis was conducted using the natural logarithm method (LLR) within the spectral clustering algorithm. The modularity value (modularity Q, Q value) serves as an indicator of stability of the clustering network, with a Q value exceeding 0.3 generally considered to indicate a significant clustering effect. The average silhouette value (weighted mean silhouette, S value) represents the similarity within clustering nodes, with a typical range of 0.5 to 1.0 and a critical value of 0.5. A higher S value suggests greater similarity within clustering, making clustering more reasonable (*7*). In this study, the CNKI clustering module's Q value is 0.8561, with an average silhouette value S of 0.984, indicating clear keyword clustering and well-defined topics in the Chinese literature (Figure 7A). The WOS clustering module's Q value is 0.8154, and the average silhouette value S is 0.9019 (Figure 7B), which demonstrates a significant clustering structure and accurate, reasonable clustering results. This highlights that the clinical treatment of rare diseases and development of orphan drugs are key areas of research both in China and globally.

3.5.3. Keyword emergence analysis

Emergent terms denote keywords whose frequency increases over a specific timeframe, reflecting the research hotspot and development trends during that period. CiteSpace6.2.R3 software was employed to plot emergent terms in the study of rare diseases globally, to delve into the research trends within this field (Figure 8, A and B). The emergent strength quantifies the importance of keywords, indicating extensive citation of the term in a short time (*8*). In the visualizations, light blue bands indicate keywords that have not emerged, dark blue bands represent keywords that have appeared, and red bands signify keywords that are frequently cited during the depicted period.

In the CNKI database, the emerging keywords in rare disease research predominantly feature rare drugs, rare diseases, and orphan drugs. It is noteworthy that since 2020, there has been a growing interest among researchers in the healthcare system for rare diseases, indicating an increased focus on medical insurance aspects of treatments for these conditions. In the WOS database, focus of rare disease research is largely on clinical trials, qualitative research, patient-reported outcomes, European reference networks, and orphan drugs. Two databases both indicate that development of orphan drugs and genetic diagnosis are globally shared key research areas in the field of rare diseases. In addition, there is a focus of international attention on molecular mechanisms of rare diseases and health management of the rare disease population.

3.6. References analysis

3.6.1. Co-cited references

Highly cited references can indicate hot topics in a research field. By analyzing these co-cited references, the dynamic changes in research topics within a specific time range can be identified. Citation frequency analysis identified the top 10 most-cited English articles (Table 6). High citation frequency indicates foundational or widely referenced works in this field, providing insights into key developments and influential studies. The standards and guidelines for sequence variants jointly formulated by the American College of Medical Genetics and Genomics and the Association for Molecular Pathology were interpreted, and the effectiveness of protein-coding genes in disease screening was discussed. The article "Rare-disease genetics in the era of next-generation sequencing: discovery to translation" published by Boycott KM, *et al.* not only has a high citation frequency but also a high centrality (Table 7). It can be inferred that genetic technology, genetic medicine, and the human phenotype ontology model have significant influence and importance in the field of rare disease research.

3.6.2. Reference bursts

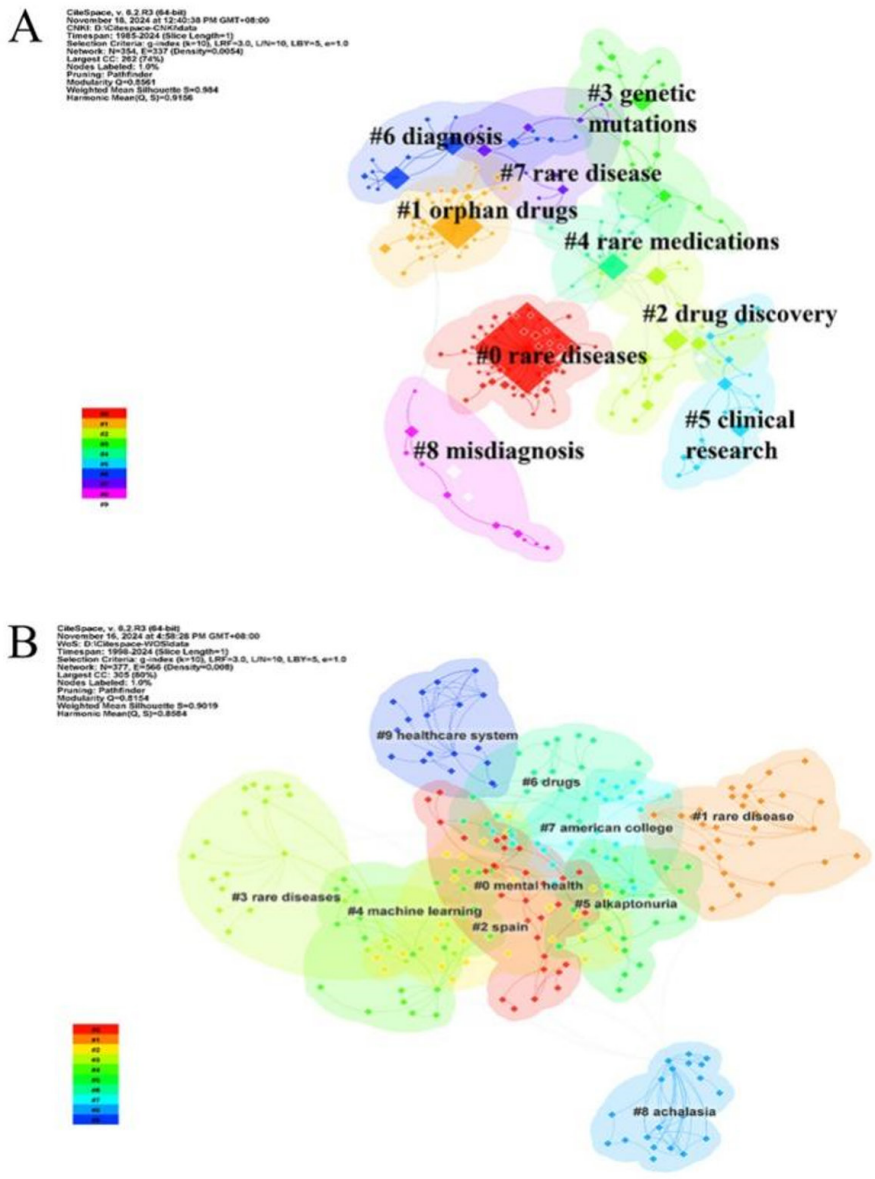

Figure 7. A clustering map of rare disease keywords obtained using CiteSpace6.2.R3 software based on the CNKI database (A) and WOS database (B), respectively.

A reference has a high burst value when it is abruptly cited heavily during a specific period, which often denotes that the research results of this reference represent innovative discoveries or new frontiers in this field (*9*). Figure 9 shows the top 25 references with the strongest burst values and their relevant information. There are three burst values exceeding 10, with citations from Wakap SN, *et al*., having a burst strength of 26.46, Haendel M, *et al.* with a burst strength of 12.29, and Betts *et al.* with a burst strength of 10.66.

4. Discussion

The rare disease research field has attracted widespread attention from scholars and achieved significant accomplishments. Here, we identified the top-producing countries, institutions, and authors. They have established representative research teams with tight internal cooperation, providing solid academic support for subsequent studies. We also determined the ten most cited influential authors and the most co-cited journals, whose publications have been widely recognized by the academic community.

4.1. Comparison of CNKI and WOS databases

In terms of the number of published papers, the United States is far ahead, with its publication volume being approximately 7.5 times that of China. Although China's paper output is still less than that of some European countries, since 2012, the number of research papers on rare diseases in the CNKI database has been continuously increasing. This trend reflects the rising attention to rare disease research in China, and there is potential for

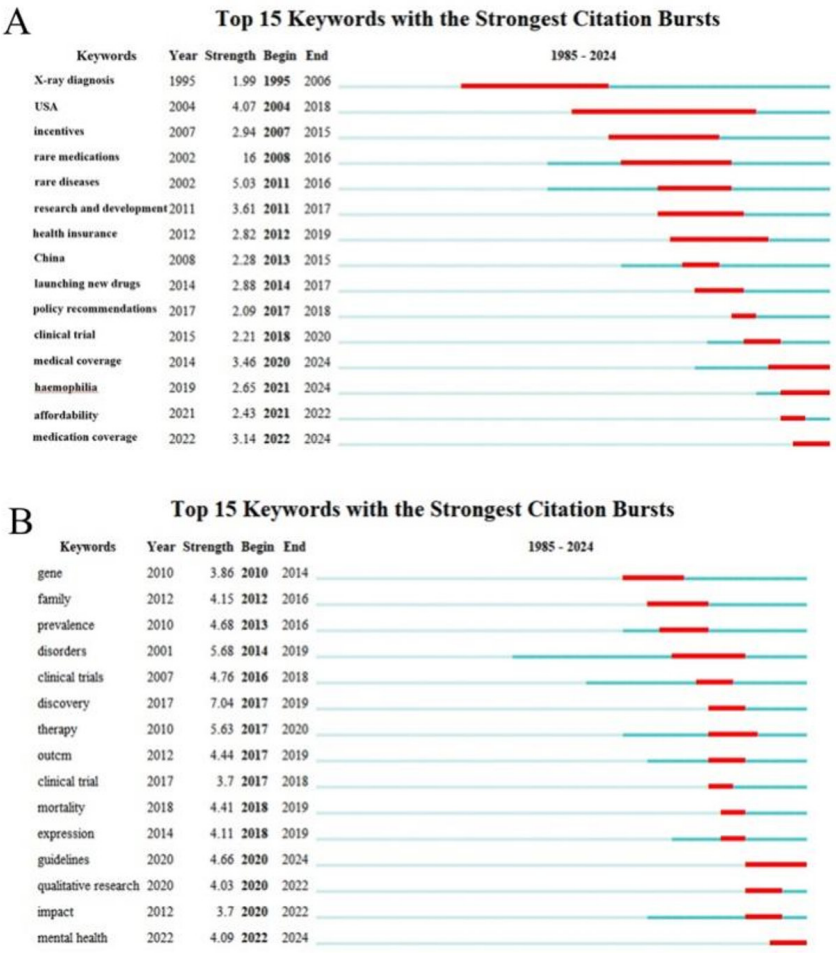

Figure 8. Top 15 keywords with the strongest bursts obtained using CiteSpace6.2.R3 software based on the CNKI database (A) and WOS database (B), respectively.

Table 6. Cited high-frequency words in the WOS database

The name of the document	First author	Year of publication	Citation frequency	DOI
Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database	Wakap, SN	2020	746	10.1038/s41431-019-0508-0
Peripartum cardiomyopathy - National Heart, Lung, and Blood Institute and Office of Rare Diseases (National Institutes of Health) workshop recommendations and review	Pearson, GD	2000	570	10.1001/jama.283.9.1183
Office of rare diseases neuropathologic criteria for corticobasal degeneration	Dickson, DW	2002	547	10.1093/jnen/61.11.935
Rare-disease genetics in the era of next-generation sequencing: discovery to translation	Boycott, KM	2013	521	10.1038/nrg3555
Why rare diseases are an important medical and social issue	Schieppati, A	2008	407	10.1016/S0140-6736(08)60872-7
Common variants at SCN5A-SCN10A and HEY2 are associated with Brugada syndrome, a rare disease with a high risk of sudden cardiac death	Bezzina, CR	1991	394	10.2307/2982708
The detection of clusters in Rare Diseases	BESAG, J	1991	380	10.2307/2982708
100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report	Peter M Visscher	2021	357	10.1056/NEJMoa2035790
On the need for the Rare Disease assumption in Case-control studies	GREENLAND, S	1991	350	10.1093/oxford journals. age.a113439
The Matchmaker Exchange: A Platform for Rare Disease Gene Discovery	Philippakis, AA	2015	346	10.1002/humu.22858

Table 7. The top 5 cited references with the highest centrality

Count	Centrality	Year	References
24	0.16	2017	Boycott KM, 2017, AM J HUM GENET, V100, P695, DOI 10.1016/j.ajhg.2017.04.003
19	0.15	2017	Köhler S, 2017, NUCLEIC ACIDS RES, V45, PD865, DOI 10.1093/nar/gkw1039
61	0.1	2019	Ferreira CR, 2019, AM J MED GENET A, V179, P885, DOI 10.1002/ajmg.a.61124
19	0.1	2013	Boycott KM, 2013, NAT REV GENET, V14, P681, DOI 10.1038/nrg3555
10	0.1	2017	Annemans L, 2017, ORPHANET J RARE DIS, V12, P0, DOI 10.1186/s13023-017-0601-9

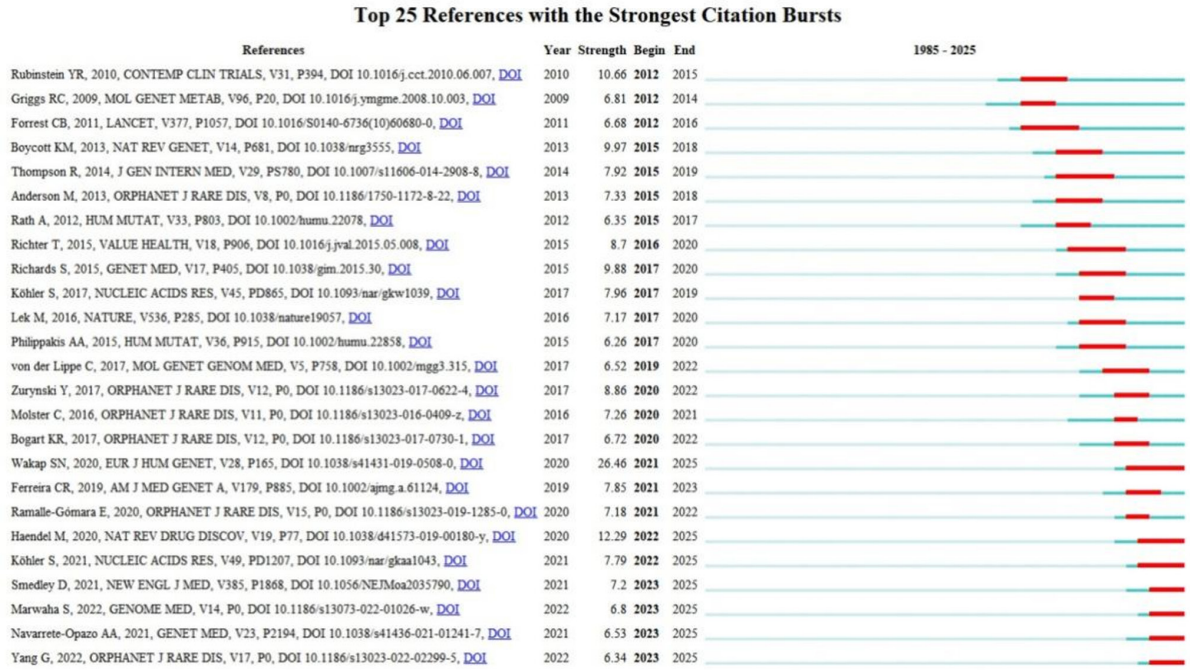

Figure 9. Twenty-five references with the strongest citation bursts.

improvement in the quality of research outcomes. The strong research capacity in economically developed countries like the United States and the European Union is due to various factors. For example, the United States is the first country in the world to formulate specific laws related to rare diseases and has established a comprehensive system for diagnosis, treatment, research, and development of orphan drugs, as well as innovative incentive mechanisms. In addition, significant financial support has been provided in these areas.

From the perspective of focus, based on the keyword frequency ranking and clustering results in CNKI, Chinese journals pay more attention to rare disease drugs, while foreign journals focus more on gene mutations and treatment. The reason behind this phenomenon may be rooted in policy differences. The United States was the first to enact specific legislation for rare diseases and has extensive experience in building a regulatory framework for these conditions. Following the Orphan Drug Act in 1983, the U.S. has continuously introduced policies like the Implementation Measures for the Orphan Drug Act and the Rare Disease Act, gradually integrating incentives for rare disease drug research and development across the entire lifecycle of pharmaceuticals. Similarly, the European Commission established the Orphan Drug Regulation in 1999, providing a legal basis for EU countries' orphan drug policies. In 2000, the European Medicines Agency formed the Committee for Orphan Medical Products to evaluate applications for rare disease drug status. These countries are pivotal in the research and manufacture of rare disease treatments. They conduct extensive genetic research, employ gene therapies, uncover disease mechanisms, and lay groundwork for new drugs and treatments. However, in China, despite existing policies that expedite approval processes, reduce taxes, and include rare disease drugs in medical insurance, most such drugs are still imported, indicating a gap compared to the U.S. and other Western countries. Hence, more time is required to enhance policies and systems concerning rare disease research, pushing China's efforts in rare disease prevention and control to a new level.

4.2. Research hotspots in the field of rare diseases

4.2.1. Orphan drug research and development

Rare diseases, often referred to as orphan diseases,

have led to the term orphan drugs for those used in their prevention, diagnosis, and treatment. The complexity and diversity of these diseases, coupled with the limited understanding and small patient populations for each condition, make the development of orphan drugs particularly challenging, lengthy, and risky. This scarcity has consistently been a major obstacle in the treatment of rare diseases. To date, over 7,000 rare diseases have been identified globally, yet only about 400 have available treatments or interventions, underscoring the pressing need for orphan drug development (*10*). In the top 25 references with the strongest citation bursts, one-third of the literature includes a focus on orphan drug research and development. The second-ranked article, "Clinical research for rare disease: opportunities, challenges, and solutions" by Grigs RC, *et al.*, mentions providing incentives for sponsors to develop promising drugs for the treatment, prevention, or diagnosis of rare diseases (*11*). The United States was an early pioneer in this field, with the Orphan Drug Act providing not only a clear definition of rare diseases, but also offering targeted research grants, tax incentives, and a 7-year market exclusivity period for orphan drugs post-launch (*12,13*). These incentives, both push and pull, have established risk-sharing mechanisms and expedited review and approval processes, substantially reducing the risks associated with orphan drug development. The European Union's Orphan Drug Regulation goes further, offering a 10-year market exclusivity period and research support, along with incentives like tax breaks and funding for pharmaceutical companies (*14*). Within the CNKI database, phrases such as "rare medications", "orphan drugs", and "new drugs" emerge with remarkable frequency, appearing hundreds of times and thus qualifying as high-frequency keywords. China, too, has introduced various policies to stimulate the development of treatments for rare diseases and is actively working to regulate and promote the growth of orphan drugs.

Different from drugs for common diseases, the development, launch, and patent approval of orphan drugs are supported by multiple international organizations and policies of many countries, leading to a much faster market penetration than non-orphan drugs (*15*). In recent years, drugs for rare diseases have accounted for over 35% of the total number of drugs approved by the FDA (*16*). Techniques such as gene therapy, antibody therapy, enzyme replacement therapy (ERT), and drug repurposing have increasingly become the focus of orphan drug research and development. The development of orphan drugs is a shared challenge and research hotspot in the global pharmaceutical industry. The emergence of new research technologies has transformed the landscape of disease treatment and offered a glimmer of hope for patients with rare diseases.

4.2.2. Genomics and genetics

Genome research can identify gene variants associated with rare diseases through whole genome sequencing or exome sequencing. Utilizing cellular and molecular genetic techniques to screen and diagnose diseases, applying genetic technology for precision treatment, exploring pathogenesis, and guiding drug development. The high-frequency keywords in the CNKI and WOS databases both show "gene" and "mutation". Gene therapy has achieved significant results in treatment of various rare diseases, such as Wiskott-Aldrich syndrome, and Fanconi anemia (*17*). In the analysis of co-cited literature, Boycott KM has two highly cited and highly prominent articles: "International Cooperation to Enable the Diagnosis of All Rare Genetic Disease" and "Rare-disease genetics in the era of next-generation sequencing: discovery to translation" (*18,19*). The articles respectively elaborate on the impact of the genome on the mechanisms of rare diseases and the use of genetic diagnosis and treatment to improve the identification rate of genes causing rare genetic diseases. These findings indicate that genetic diagnosis and treatment are a hot topic in the field of rare diseases.

In 2017, the world's first gene therapy drugs were approved for market, officially opening the era of personalized treatment and gene therapy. To date, more than 40 gene therapy drugs have been approved for clinical application worldwide, with indications including hereditary autoimmune diseases, blood disorders, neurological diseases, and solid and non-solid tumors, with treatment costs being generally high. In contrast, China's gene therapy industry is still in its infancy. So far, more than 20 products in China have entered the clinical trial phase.

In 2024, researchers from institutions such as Boston Children's Hospital, Harvard Medical School, and Dana-Farber Cancer Institute in the United States conducted sequencing and analysis of the exomes and genomes of more than 8,000 family members suspected of having rare single-gene diseases, assessing the diagnostic yield of genomic sequencing within this cohort. The study results confirmed that genomic sequencing can provide diagnostic evidence for some families affected by rare single-gene diseases, increasing the diagnostic rate for rare diseases by 8% (*20*). In recent years, China has also attached great importance to the development and application of genetic testing technology. In September 2022, the Central Special Lottery Public Welfare Fund initiated the UPWARDS project (Upgrade of Precision in Diagnostics and Treatment of Rare Diseases), led by Peking Union Medical College Hospital, with the participation of hospitals in the National Rare Disease Diagnosis and Treatment Collaboration Network. To date, 514 hospitals have joined the project, and a total of 100,500 people have been tested, playing a significant role in improving China's capabilities and level of diagnosis and treatment for rare diseases.

4.2.3. Medical insurance coverage

The challenge of rare diseases in public health has garnered significant global attention, becoming a major hurdle for public health systems in EU countries. The EU has integrated efforts to enhance the quality of life for those afflicted with rare diseases into its regulatory framework. The focus on improving patients' quality of life and expanding medical insurance coverage for rare diseases stands at the cutting edge of research in this field. The diversity of rare diseases, coupled with the small number of patients, limited market demand, and high costs of drug development, has led to a lack of interest from pharmaceutical companies in researching treatments for these conditions (*21*). Against this backdrop, numerous countries around the world have enacted orphan drug legislation over the past three decades to safeguard the rights of rare disease patients. The U.S. was the first to enact legislation related to rare diseases, establishing comprehensive systems and innovative incentives for the diagnosis, treatment, and development of orphan drugs, supported by substantial funding. The third most highly cited article, titled "The burden of rare diseases" by Ferreira CR, *et al*., mentions the health care services for patients with rare diseases (*22*). These patients face significant living expenses, making it particularly important to pay attention to their physical and mental health. Hence, establishing and enhancing rare disease drug security policies to ensure fair access to treatment for rare disease patients has become a global research priority.

In the United States, the review process for orphan drugs is, on average, at least 6 months shorter than that for conventional medications. Moreover, the US FDA typically waives or reduces the review fees for orphan drugs, and new orphan drugs can sometimes be exempt from application fees entirely (*23*). Following the EU Medicines Agency's adoption of the Orphan Medicinal Products Regulation in 2000, continuous revisions have been made, laying the groundwork for a comprehensive rare disease management system (*24*). Due to the substantial healthcare and social care system resources required for rare diseases, the fifth most highly cited article published in the Orphanet Journal of Rare Diseases, titled "Recommendations from the European working group for value assessment and funding processes in rare diseases (ORPH-VAL)" primarily discusses several decisions regarding the research and development, pricing, and reimbursement of orphan drugs, with the aim of ensuring equitable access to effective treatments for the most vulnerable populations (*25*).

Although China's articles on medical insurance for rare diseases rank lower internationally, in the CNKI database, the keyword emergence map shows that "incentive" ranks second, with "policy recommendations" also among the top 10, indicating China's significant focus on rare disease policies. In recent years, China has actively sought its own approach to rare disease research and care. For instance, the government has established a rare disease directory and integrated it into the medical insurance system, set up a national collaborative network for rare disease diagnosis and treatment to provide support, intensified drug negotiation efforts to significantly reduce medication prices, and improved a multi-tiered medical insurance system for rare diseases to alleviate the financial burden on patients (*26*). Governments, society, the market, and patients themselves are all actively exploring solutions to ensure the care and treatment of rare diseases. In May 2018, the National Health Commission and other departments jointly released the *First Batch of Rare Disease Directory*, encompassing 121 rare diseases, with 141 drugs approved for marketing in China capable of treating 53 of these conditions, thus addressing the void in China's rare disease treatment landscape (*27*). In September 2023, the National Health Commission and other departments issued the *Second Batch of Rare Disease Directory*, adding 86 new rare diseases, with 57 approved drugs available to treat 39 of these rare diseases (*28*).

4.2.4. Artificial intelligence and big data

In the most frequently cited literature, the article "Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database" published by Nguengang Wakap S, *et al.*, precisely utilizes the Orphanet database to analyze rare diseases (*29*). Orphanet (*https://www.orpha.net/*) is currently recognized as an authoritative and respected public database for rare diseases at the international level. It contains high-quality information and research results about rare diseases. Orphanet's classification integrates various types of information, categorizing rare diseases into 32 groups and initially coding them by adopting the ICD coding method. It has established a unique naming method and ORPHA coding for rare disease. Additionally, as mentioned in the 2017 article "The human phenotype ontology" by Köhler S, *et al.*, HPO is increasingly being used by different groups such as international rare disease organizations, registries, clinical laboratories, biomedical resources, and clinical software tools as a standard for phenotypic abnormalities (*30*). This will contribute to the initial efforts of global data exchange to determine disease etiology. This illustrates the importance of data collection and registration for the diagnosis and treatment of rare diseases.

With the development of technology, harnessing the powerful computing capabilities of artificial intelligence to determine the relationship between genes and disease onset, and exploring potential drug targets has become a new trend. For example, Exome Disease Variant Analysis (EVA) is a variant pathogenicity prediction and

annotation tool published in 2019 and has been applied to various rare diseases (*31*). This tool identifies causal mutations by using whole exome sequencing of trios (family members or father-mother-child), and researchers have applied it to several cases of familial diseases to demonstrate its clinical applicability. The ability of artificial intelligence to integrate and analyze data from different sources holds great potential in overcoming challenges related to rare diseases, such as low diagnostic rates, small patient numbers, and geographical dispersion.

4.2.5. Limitations of the study

Considering the purpose of this article, no visualization analysis was conducted on the research hotspots or directions for a specific rare disease. In the future, further research and analysis can be conducted in a specific direction or from a particular perspective, thereby clarifying specific research hotspots and progress in that field. Since the CiteSpace software can only support the visualization analysis of papers from a single database, although both CNKI and Web of Science, two well-known citation databases, were analyzed, there is still the possibility of biased or omitted literature analysis. Additionally, due to the limitations of the CiteSpace software and the CNKI database, it is not possible to directly export citation data, resulting in a lack of co-citation analysis for Chinese literature in this study.

5. Conclusion

This study employs CiteSpace software to conduct a systematic review of literature on rare diseases from the CNKI and WOS databases. By analyzing research institutions, authors, keywords, and highly cited documents, a visual knowledge map is created. The goal is to uncover the current status, hotspots, and trends in this field, offering valuable insights for global rare disease research. As data science and technology advance, the visualization research in the rare disease field will be further enhanced. This not only fosters international collaboration and communication, but also speeds up transition of research findings into clinical practice, ensuring more effective support and care for patients with rare diseases. Meanwhile, artificial intelligence and big data in rare diseases will empower prevention, diagnosis and treatment of rare diseases, as well as disclosing the relationship between complex diseases and rare diseases in the near future.

Funding: This work was supported by a grant from the Natural Science Foundation of Shandong Province (General program ZR2023MH276) and Academic Promotion Program of Shandong First Medical University (LJ001).

Conflict of Interest: The authors have no conflicts of interest to disclose.

References

1. Melnikova I. Rare diseases and orphan drugs. Nat Rev Drug Discov. 2012; 11:267-268.
2. Lu Y, Han J. The definition of rare disease in China and its prospects. Intractable Rare Dis. 2022; 11:29-30.
3. Li Y. Policy recommendations regarding rare diseases in China - An analysis based on a survey of living conditions of the people with rare diseases. China Soft Science. 2014; 2:77-89. (in Chinese)
4. Parshin V, Osminin S, Komarov R, Vetshev S, Strakhov Y, Ivashov I. Rare diseases of esophagus: Surgical treatment of cysts in adults. Case report. Int J Surg Case Rep. 2021; 81:105732.
5. Li X, Zhang X, Zhang S, Lu Z, Zhang J, Zhou J, Li B, Ou L. Rare disease awareness and perspectives of physicians in China: a questionnaire-based study. Orphanet J Rare Dis. 2021; 16:171.
6. Synnestvedt MB, Chen C, Holmes JH. CiteSpace II: visualization and knowledge discovery in bibliographic databases. AMIA Annu Symp Proc. 2005; 2005:724-8.
7. Sabe M, Pillinger T, Kaiser S, Chen C, Taipale H, Tanskanen A, Tiihonen J, Leucht S, Correll CU, Solmi M. Half a century of research on antipsychotics and schizophrenia: A scientometric study of hotspots, nodes, bursts, and trends. Neurosci Biobehav Rev. 2022; 136:104608.
8. Zhang L, Zheng H, Jiang ST, Liu YG, Zhang T, Zhang JW, Lu X, Zhao HT, Sang XT, Xu YY. Worldwide research trends on tumor burden and immunotherapy: a bibliometric analysis. Int J Surg. 2024; 110:1699-1710.
9. Liao H, Tang M, Luo L, Li C, Chiclana F, Zeng XJ. A bibliometric analysis and visualization of medical big data research. Sustainability. 2018; 10:166.
10. Zhang S, Chen L, Zhang Z, Zhao Y. Orphan drug development in China: progress and challenges. Lancet. 2019; 394:1127-1128.
11. Griggs RC, Batshaw M, Dunkle M, Gopal-Srivastava R, Kaye E, Krischer J, Nguyen T, Paulus K, Merkel PA. Rare Diseases Clinical Research Network. Clinical research for rare disease: opportunities, challenges, and solutions. Mol Genet Metab. 2009; 96:20-26.
12. Pedrini E, Negro A, Di Brino E, Pecoraro V, Sculco C, Abelli E, Gnoli M, Magrelli A, Sangiorgi L, Cicchetti A. Real-world data and budget impact analysis (BIA): Evaluation of a targeted next-generation sequencing diagnostic approach in two orthopedic rare diseases. Front Pharmacol. 2022; 13:785705.
13. Weerasooriya SU. The impact of orphan drug policies in treating rare diseases. Health Info Libr J. 2019; 36:179-184.
14. Liu BC, He L, He G, He Y. A cross-national comparative study of orphan drug policies in the United States, the European Union, and Japan: towards a made-in-China orphan drug policy. J Public Health Policy. 2010; 31:407-420; discussion 420-421.
15. Berlin AC, Eggan ER. Realizing the potential of patient organizations in translational research-delivering treatments for rare diseases. Acad Med. 2014; 89:1573.
16. Taruscio D, Capozzoli F, Frank C. Rare diseases and orphan drugs. Ann Ist Super Sanita. 2011; 47:83-93.

17. Emmert D, Rasche T, Sellin J, Brunkhorst R, Bender TTA, Weinstock N, Börsch N, Grigull L, Conrad R, Mücke M. Rare diseases in the differential diagnosis of myalgia. Schmerz. 2022; 36:213-224. (in German)
18. Boycott KM, Rath A, Chong JX, *et al.* International Cooperation to Enable the Diagnosis of All Rare Genetic Diseases. Am J Hum Genet. 2017; 100:695-705.
19. Boycott KM, Vanstone MR, Bulman DE, MacKenzie AE. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. Nat Rev Genet. 2013; 14:681-691.
20. Carey JC. The importance of case reports in advancing scientific knowledge of rare diseases. Adv Exp Med Biol. 2010; 686:77-86.
21. Chen S, Dong D. Improving insurance protection for rare diseases: Economic burden and policy effects - simulation of people with Pompe disease in China. Int J Health Policy Manag. 2023; 12:6282.
22. Ferreira CR. The burden of rare diseases. Am J Med Genet A. 2019; 179:885-892.
23. Stafinski T, Glennie J, Young A, Menon D. HTA decision-making for drugs for rare diseases: comparison of processes across countries. Orphanet J Rare Dis. 2022; 17:258.
24. Groft SC, Posada M, Taruscio D. Progress, challenges and global approaches to rare diseases. Acta Paediatr. 2021; 110:2711-2716.
25. Annemans L, Aymé S, Le Cam Y, Facey K, Gunther P, Nicod E, Reni M, Roux JL, Schlander M, Taylor D, Tomino C, Torrent-Farnell J, Upadhyaya S, Hutchings A, Le Dez L. Recommendations from the European Working Group for Value Assessment and Funding Processes in Rare Diseases (ORPH-VAL). Orphanet J Rare Dis. 2017; 12:50.
26. He J, Song P, Kang Q, Zhang X, Hu J, Yang Y, Tang M, Chen D, Hu S, Jin C. Overview on social security system of rare diseases in China. Biosci Trends. 2019; 13:314-323.
27. National Health Commission of the People's Republic of China. The first batch of rare disease catalog in China. *https://www.gov.cn/zhengce/zhengceku/2018-12/31/content_5435167.htm* (accessed December 1, 2024). (in Chinese)
28. National Health Commission of the People's Republic of China. The second batch of rare disease catalog in China. *https://www.gov.cn/zhengce/zhengceku/202309/content_6905273.htm* (accessed December 1, 2024). (in Chinese)
29. Nguengang Wakap S, Lambert DM, Olry A, Rodwell C, Gueydan C, Lanneau V, Murphy D, Le Cam Y, Rath A. Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database. Eur J Hum Genet. 2020; 28:165-173.
30. Köhler S, Vasilevsky NA, Engelstad M, *et al.* The human phenotype ontology in 2017. Nucleic Acids Res. 2017; 45:D865-D876.
31. Wojtara M, Rana E, Rahman T, Khanna P, Singh H. Artificial intelligence in rare disease diagnosis and treatment. Clin Transl Sci. 2023; 16:2106-2111.

Received November 21, 2024; Revised February 11, 2025; Accepted February 16, 2025.

**Address correspondence to:*
Jinxiang Han and Yanqin Lu, The First Affiliated Hospital of Shandong First Medical University & Shandong Provincial Qianfoshan Hospital, No. 16766 Jingshi Road, Ji'nan, Shandong 250013, China.
E-mail: jxhan@sdfmu.edu.cn (JH), yqlu@sdfmu.edu.cn (YL)

Released online in J-STAGE as advance publication February 22, 2025.