Original Article

DOI: 10.5582/irdr.2024.01059

Rare disease publishing trends worldwide and in China: A CiteSpace-based bibliometric study

Qi Kong[1,2,§], Chenxin Fan[3,4,§], Ying Zhang[3], Xinlei Yan[3], Liming Chen[3,4], Qi Kang[5,*], Peihao Yin[1,2,*]

[1] Putuo People's Hospital, School of Medicine, Tongji University, Shanghai, China;
[2] Putuo Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
[3] Scientific Innovation Volunteer Team of Rare Diseases, Shanghai TCM-Integrated Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
[4] Yueyang Hospital of Integrated Traditional Chinese and Western Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China;
[5] Shanghai Health Development Research Center (Shanghai Medical Information Research Center), Shanghai, China.

SUMMARY: This study aimed to understand research trends, determine frontier topics, and explore the developments in and the differences between research conducted in China and the rest of the world. We analyzed the research status of rare diseases in China and globally over the past decade using bibliometric methods. We focused on rare disease literature indexed in the Web of Science (WoS) and China National Knowledge Infrastructure (CNKI) databases from January 2013 to December 2023. We selected studies based on inclusion and exclusion criteria. CiteSpace 6.1.R6 software were used to prepare knowledge graphs and perform comparative analyses of authors, institutions, content, and hot topics between both databases. A total of 10,754 articles from the WoS and 969 from the CNKI met the inclusion criteria. In the past 10 years, the diagnosis and treatment of rare diseases have been a common research focus in both China and the world. China has emphasized more on "orphan drugs". "Genes" and "management" were focused globally. The United States had the greatest number of publications. China ranks high in terms of publication volume and institutional ranking. Research interest in rare diseases has gradually increased worldwide, with European and American countries maintaining a leading position. China has made significant contributions. China's research is lagging compared to global trends, lacking collaboration with other countries. The diagnosis and treatment of rare diseases remain central themes, whereas genetic research, artificial intelligence, and sociological studies on rare disease populations are emerging as hot topics.

Keywords: rare diseases, orphan diseases, visualization, knowledge graph, CiteSpace

1. Introduction

Rare diseases refer to a group of clinically heterogeneous disorders characterized by considerably low prevalence and a small population of affected individuals (*1*). Different countries and regions have established different specific definitions for rare diseases. The U.S. Food and Drug Administration (FDA) defines rare diseases as "any disease or condition that affects < 200,000 persons in the United States (*2*)". The European Medicines Agency (EMA) defines rare diseases as "life-threatening or chronically debilitating conditions that affect no more than 5 in 10,000 people in the EU (*2*) ". According to The Report on the Definition of Rare Diseases in China 2021, China defines rare diseases as "diseases with a neonatal incidence rate of less than 1 in 10,000, a prevalence rate of less than 1 in 10,000, and a total affected population of less than 140,000 (*3*) ". According to estimates, rare diseases affect over 300 million patients worldwide, with over 7,000 rare diseases accounting for 10% of all human diseases. Among these, 80% are of genetic origin (*4*). However, currently, effective treatment methods are unavailable for more than 90% of rare diseases (*5*). The direct medical costs, non-medical costs, and productivity losses associated with rare diseases impose a significant burden on society (*6*). Therefore, the treatment and management of rare diseases have become important global public health issues (*7*).

According to statistics, the number of patients with rare diseases in China is estimated to be approximately 20 million, with an average of 200,000 new cases reported annually (*3*). Rare diseases are receiving increasing attention in China. In 2018, China released The First Batch of Rare Disease Catalog in China, which mentioned 121 rare diseases (*8*). This made China the first country to classify rare diseases using a specific list

(*9*). In 2023, China released The Second Batch of Rare Disease Catalog in China, which added 86 (*10*) new rare diseases to the list. Meanwhile, globally, developed countries such as the United States, France, and Germany have implemented national rare disease programs and orphan drug regulations and provide substantial funding to support research innovation and the establishment of healthcare systems for rare diseases (*11*). Academic research on rare diseases has experienced unprecedented and rapid development. WoS and CNKI databases are the most comprehensive databases for core journal indexing in China and worldwide. They can help identify scientific research trends from publications in different languages and country perspectives (*12*).

This study aims to provide a comprehensive overview of the research trends in rare diseases over the past decade using CiteSpace software. We summarize the development process in the field of rare diseases and provide directions for further advancements in this area.

2. Methods

2.1. Retrieval strategy

The data for this study were obtained from the WoS Core Collection and CNKI databases. For the WoS database, a topic-based search was conducted using the keywords "(rare disease) OR (rare cancer) OR (rare tumor)". The keyword "rare disease" was set as a "must inclusion", and the time range spanned from January 1, 2013, to December 31, 2023. The selected document types were "Review Article" and "Article."

For the CNKI database, the search strategy included the keywords "(rare disease (in Chinese)) OR (rare disorder (in Chinese)) OR (in Chinese))" for the topic search. The time range spanned from January 1, 2013, to December 31, 2023. The selected document type was "Journal Article".

2.2. Inclusion criteria

The inclusion criteria for the academic papers were as follows: the articles must have keywords such as "rare disease (in Chinese)", "rare disorder (in Chinese)", or "rare tumor (in Chinese)", and the research topic must be related to rare diseases or rare tumors. The inclusion was limited to academic journal papers available in the Chinese language for CNKI. No language restrictions were used for papers retrieved from the WoS.

2.3. Exclusion criteria

The exclusion criteria for literature selection were as follows: *i*) Duplicate publications, *ii*) Literature unrelated to the research topic of rare diseases or rare tumors, *iii*) Literature without relevant keywords and/or the complete text, *iv*) Conference papers, patents, newspapers, and project reports.

2.4. Literature screening and data extraction

Two researchers conducted a thorough review of the literature based on the predetermined inclusion and exclusion criteria. They screened the titles and abstracts to exclude irrelevant data. In case of discrepancies, a third person was consulted for consensus. 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.

2.5. Analysis method

CiteSpace 6.1.R6 is a Java-based program that supports visual exploration and knowledge discovery in literature databases. The analysis was conducted using CiteSpace to aid the visualization of the research landscape (*13*). The time slice was set to 1 year, and the top N was set to 50. Pruning was set to Pathfinder and Pruning sliced networks. The analysis included the K-means clustering analysis of keywords and the identification of emerging research frontiers. The results were presented in the form of timeline graphs and keyword burst graphs in the visual interface of CiteSpace.

3. Results

3.1. Literature search results

A total of 11,347 papers were retrieved from WoS. After data inspection, deduplication, and cleaning, 10,754 papers were included in the study. These papers were exported in the form of "Plain Text - Full Record and References". An initial screening of CNKI yielded 2,152 papers. After further evaluation, 969 papers were selected

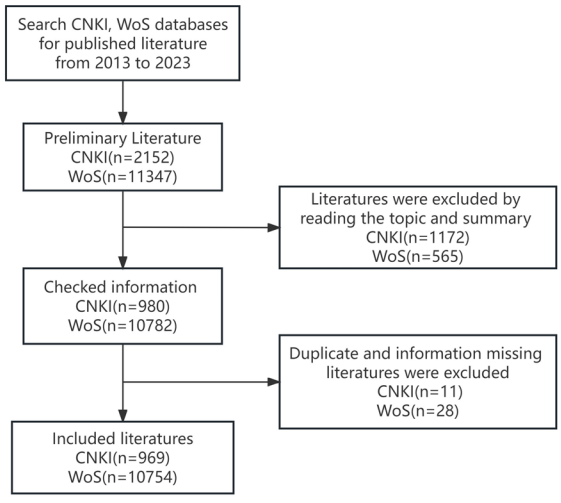

Figure 1. Flowchart of literature search and data incorporation.

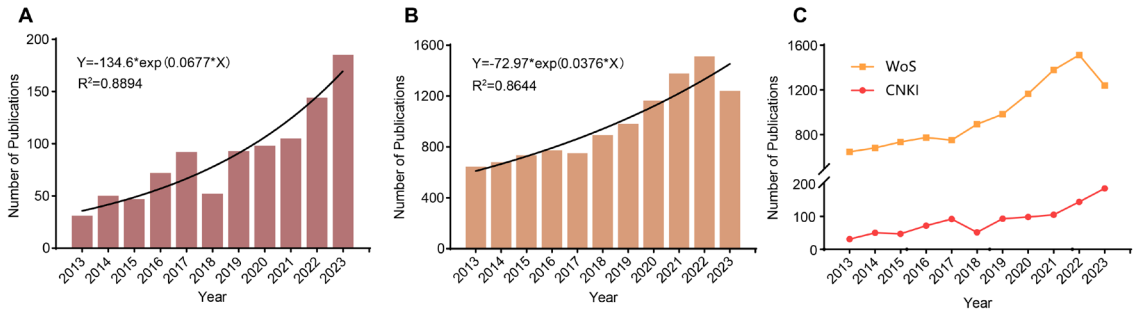

Figure 2. Chart of annual publishing and growth trends. (A) Trends in the number of publications per year in the CNKI database; **(B)** Trends in the number of publications per year in the WoS core database; **(C)** Comparison of publication growth trends between the CNKI and WoS core database.

and included in the study (Figure 1).

In the CNKI database (Figure 2A), the overall number of publications was relatively low and showed slow growth. There was a decline in 2018, followed by a recovery to normal levels. A rapid increase was observed from 2022. In 2023, the number of publications was 5.97 times greater than that in 2013. In the WoS database, the number of publications in the rare disease field showed a steady growth trend (Figure 2B). The growth rate increased significantly from 2018, although there was a marginal decline in 2023.

Based on a comparison of the domestic and international publication counts (Figure 2C), the field of research on rare diseases is continuously expanding both domestically and internationally. Compared to domestic journals, international journals show a higher level of attention and depth of research in this field.

Using a nonlinear index to fit the growth trend, the curve fitting equation for CNKI was set at $Y = -134.6*\exp(0.0567*X)$, $R^2 = 0.8894$; WoS is $Y = -72.97*\exp(0.0376*X)$, $R^2 = 0.8644$.

3.2. Database literature spatial distribution (core countries/institutions)

3.2.1. Countries and Regions

The country with the greatest number of publications in the WoS database was the United States, with 2,423 papers (22.5%). China ranked second with 1969 papers (18.3%), followed by Italy and Germany, both with over 1,000 papers (10.4% and 9.4%). However, there was a significant gap between the top two countries and the rest (Table 1). This indicates the core position of the United States in the field of rare disease research. China also demonstrated strong research capabilities. Among the top 10 countries, six were European countries. From the Network of Collaborating Countries (Figure 3), extensive and frequent collaboration was observed among European and American countries. However, currently, China has limited research collaborations with other countries. In terms of literature centrality,

Table 1. The top 10 productive countries with publications concerning rare diseases

Rank	Article counts	Centrality	Countries
1	2,423	0	USA
2	1,969	0	PEOPLES R CHINA
3	1,123	0	ITALY
4	1,020	0	GERMANY
5	881	0.01	FRANCE
6	821	0.07	ENGLAND
7	649	0	JAPAN
8	598	0.07	SPAIN
9	465	0.05	CANADA
10	379	0.09	NETHERLANDS

the Netherlands had the greatest intermediary centrality (0.09), indicating close collaborative connections with other countries. The Netherlands has not only published a large volume of research output but also achieved high-quality results in the field of rare disease research, thus playing a crucial role.

3.2.2. Research Institutions

Among publishing institutions (Table 2), Chinese research institutions have abundant research output in both domestic and international databases. The Beijing Union Medical College Hospital, Chinese Academy of Medical Sciences, as a top medical and research center in China, is the only institution that ranks among the top 10 institutions in both domestic and international databases. It houses the National Key Laboratory for Rare and Difficult Diseases and focuses on diseases such as Gitelman syndrome, transthyretin amyloidosis cardiomyopathy, and hereditary retinal degeneration. Zhejiang University has the greatest number of publications in the WoS database and primarily focuses on rare diseases such as spinal muscular atrophy and Wilson's disease. Combining the results of the analysis of countries in the previous section, it is evident that European and American countries hold critical positions in the field of rare disease research. Meanwhile, the research capabilities of China are noteworthy.

Figure 3. Network of collaborating countries. The circle indicates the country; the larger the size of the circle, the greater the number of publications from the country. Links between nodes describe cooperation between the countries.

Table 2. The top 10 productive institutions ranked by the numbers of publications

Rank	Article counts	Institutions (CNKI)	Article counts	Institutions (WoS)
1	74	Peking Union Medical College Hospital	151	Zhejiang Univ
2	61	China Medical University	132	Capital Med Univ
3	31	Peking University	127	Harvard Med Sch
4	31	Peking Union Medical College	126	Univ Milan
5	21	Shandong University	123	Mayo Clin
6	20	National Medical Products Administration	118	Univ Toronto
7	18	Shenyang Pharmaceutical University	105	Univ Penn
8	15	Shanghai Health and Health Development Research Center	98	Chinese Acad Med Sci
9	14	West China Hospital, Sichuan University	94	Sichuan Univ
10	12	Shanghai Jiao Tong University	90	INSERM

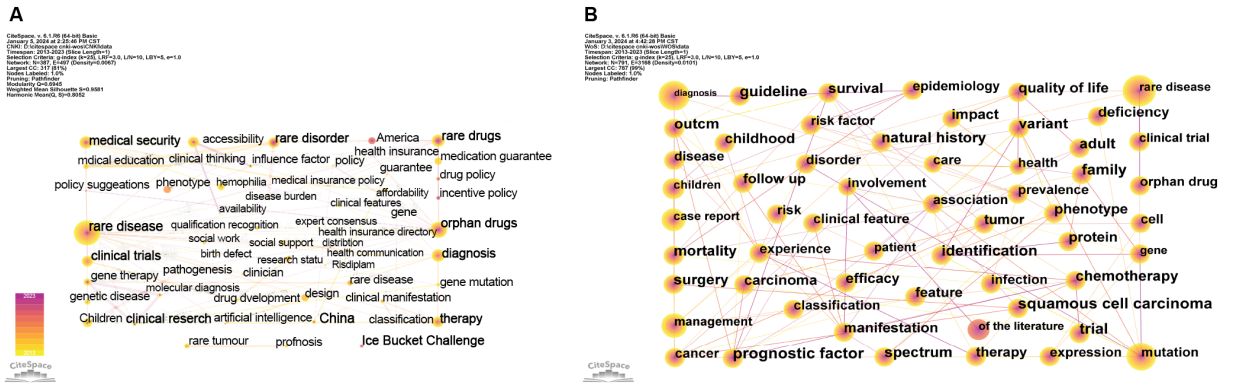

Figure 4. Network of primary keywords in publications. (A) Network of primary keywords in the CNKI database; **(B)** Network of primary keywords in the WoS core database. The circle indicates the keyword; the larger the size of the circle, the greater the frequency of the keyword.

3.3. Keywords

3.3.1. Co-occurrence of keywords

Keywords are a reflection of the core content of the literature. In this study, we selected co-occurrence network graphs of keywords with a frequency greater than 100 in the WoS database (Figure 4B). The top three keywords were "rare disease" (1,662), "diagnoses" (980), and "mutation" (833) (Table 3). "Management" (794) also had a high frequency. In the CNKI database (Figure 4A, Table 3), "rare disease" was also the most frequently appearing keyword (452), followed by "orphan drugs" (100) and "diagnosis" (38). A comparison of the two showed that "rare disease", "diagnosis", "therapy", and "Children" are among the

Table 3. The top 10 keywords ranked by frequency

Rank	Article counts	Centrality	Keywords of CNKI	Article counts	Centrality	Keywords of WoS
1	452	1.19	rare disease	1,662	0.01	rare disease
2	100	0.11	orphan drugs	980	0	diagnosis
3	38	0.05	diagnosis	833	0.01	mutation
4	35	0.02	therapy	794	0.01	management
5	31	0.01	medical security	619	0.01	disease
6	29	0.11	rare drugs	602	0.01	children
7	23	0.01	children	403	0.01	therapy
8	20	0.05	rare disorder	390	0	case report
9	19	0.01	clinical trial	380	0.01	expression
10	15	0.02	accessibility	377	0.01	cancer

top 10 keywords with the greatest frequency in both databases. This indicates that diagnosis, therapy, and children's rare diseases are common themes of concern in this field at both domestic and global levels, and they represent the research focus. However, among the high-frequency keywords in the domestic database, certain terms were related to orphan drugs and rare drugs, which were also treatment-related. In contrast, the international database focused more on gene-related directions such as "mutation" and "expression", as well as rare disease management. This reveals the different research perspectives between domestic and international studies in this field. Chinese research institutions may place a greater emphasis on drug development and application, whereas international research tends to explore the mechanisms underlying rare diseases and the social management of special populations.

3.3.2. Keyword clustering

Keyword clustering can reflect the different research focuses in a particular field. The smaller the clustering number, the more the keywords are included in that cluster. In the WoS database (Table 4), orphan drugs and whole-exome sequencing were the most prominent research directions. Case reports were the primary form of research output, indicating that with the improvement of medical standards and advancement of diagnostic methods, a greater number of rare disease cases are being reported. In the CNKI database (Table 4), orphan drugs remain a research focus. The comparison between the two databases reconfirmed the difference in research focus between domestic and international studies in the field of rare diseases. These research directions also indicate that international research on rare diseases has focused on findings at the molecular level, whereas Chinese journals continue to focus on treatment and medication, with less emphasis on the investigation of disease mechanisms. The development of the field of rare diseases in China is less advanced than the international community.

3.3.3. Keyword timeline graph

The timeline graph reflects the development of keywords within each cluster. In the WoS database (Figure 5B), the top ten keywords in terms of frequency have been appearing since 2013 and have maintained a high occurrence rate over the past 10 years. This indicates that the diagnosis, therapy, management, and genetic research of rare diseases constitute the foundation of this field. On this basis, medical genetics (2019), genomics (2018), and public health (2020) have advanced considerably, aiding rare disease research at a deeper investigative level. In the CNKI database (Figure 5A), high-frequency keywords such as "rare disease", "orphan drugs", and "rare drugs" related to medications have appeared since 2013 and have laid the foundation for subsequent research directions. Keywords such as "diagnosis", "therapy", and "clinical trials" only appeared first in 2015-2016, exhibiting a lag compared to international trends. Keywords related to rare disease management and social support in the field of social sciences also appeared relatively late, indicating that China still lacks sufficient social attention and policy support for rare disease populations. Based on observations of the keyword burst graph (Figure 6), the Ice Bucket Challenge, which went viral on social media for promoting awareness of amyotrophic lateral sclerosis (ALS), once created a wave of enthusiasm in China but disappeared within a year. This indicates that while marketing-style dissemination can increase social awareness of rare diseases, it does not significantly impact social security in rare disease populations. Currently, the research focus is concentrated on medical security, and governmental influence may significantly improve the lives of rare disease populations in China.

3.4. Authors

In the WoS database, Taruscio, Domenica; Boycott, Kym M; and Baynam, Gareth were the top three authors in terms of publication volume (Table 5). The three authors constituted the center in the collaboration network, and their collaboration has gained prominence in the past 5 years (Figure 7B). Before this period, the author collaboration network was not significant, with the majority of authors producing independent work.

Table 4. The information of clusters about keyword co-citation analysis

Database	Clusters	Label	Terms
WoS Database	0	rare disease	rare disease; target therapy; desmoplastic small round cell tumor; tyrosine kinase receptor; fetal growth restriction | squamous cell carcinoma; targeted therapy; penile cancer; systemic therapy; retroperitoneal sarcoma
	1	orphan drug	rare disease; orphan drug; spinal muscular atrophy; real-life outcome data; single-arm trial | rare diseases; comparative effectiveness; evidence generation; patient-oriented outcomes; evidence synthesis
	2	case report	case report; langerhans cell histiocytosis; igg4-related disease; central diabetes insipidus; bone marrow | magnetic resonance imaging; myeloid sarcoma; sacral spine; primary testicular lymphoma; testicular cancer
	3	whole-exome sequencing	rare disease; whole-exome sequencing; genic intolerance; health ethics; government regulation | mutation; variant; protein; tool; common disease
	4	aortic dilatation	rare disease; aortic dilatation; bicuspid aortic valve; aortic dissection; aortic coarctation | pulmonary hypertension; lung cancer; von recklinghausen; neurofibromatosis type; viral coinfection
	5	oxidative stress	rare disease; oxidative stress; lipid metabolism; mitochondrial dysfunction; brain iron accumulation | gene expression; growth; model; mutation; mice
CNKI database	0	rare disease	rare disease; orphan drugs; rare drugs; rare disorder; diagnosis
	1	orphan drugs	orphan drugs; medicine policy; comparative analysis; analysis; price negotiation
	2	rare drugs	rare drugs; America; pharmaceutical companies; research and development; indication expansion
	3	Pathology	Pathology; gene therapy; clinical studies; diagnosis; rare disease
	4	children	children; therapy; artificial intelligence; clinical manifestations; classification
	5	medical education	medical education; rare disorder; clinical thinking; computer technology; teaching quality
	6	medical security	medical security; Hemophilia; affordability; ethical principles; medical insurance policy
	7	accessibility	accessibility; influencing factors; availability; strategy research; evaluation index
	8	clinical trial	clinical trial; recruitment; drug development; rare lung tumors; priority review
	9	China	China; guarantee; national rare disease registration system; cooperation; Gauchers disease

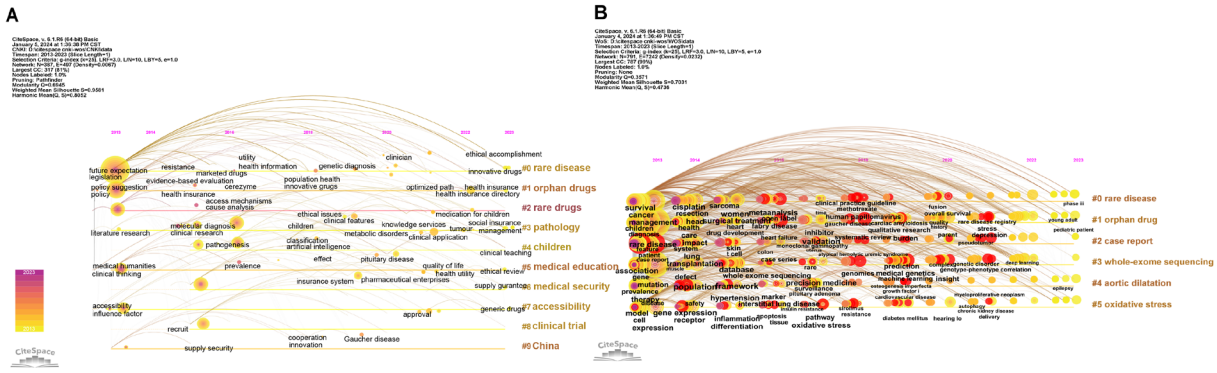

Figure 5. Timeline view of keywords. (A) Timeline view of keywords in the CNKI database. **(B)** Timeline view of keywords in the WoS core database. The circular nodes on the line represent the top three keywords with the greatest frequency of occurrence in this time slice. The timeline is shown at the top of the figure, and the year corresponding to the node is its publication time. The link between nodes represents the co-citation relationship.

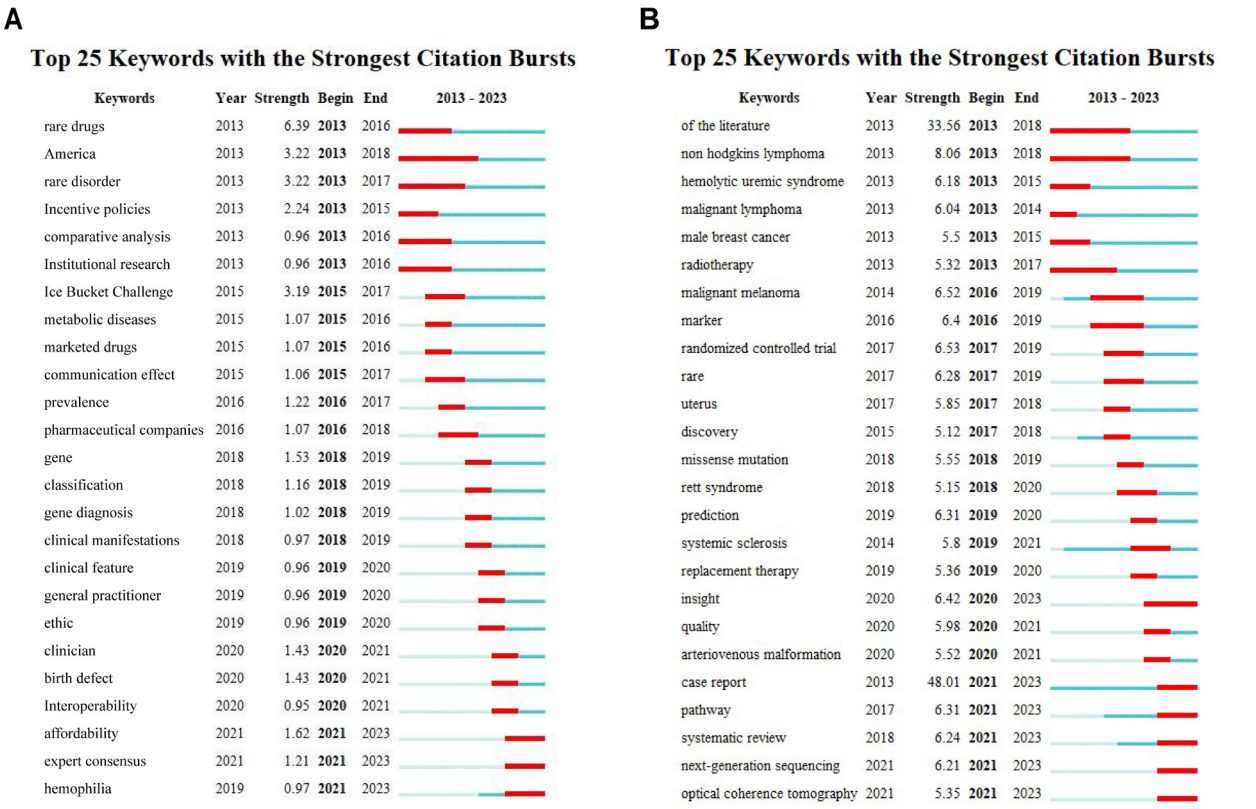

Figure 6. Twenty-five keywords with the strongest citation bursts. (A) The 25 keywords with the strongest citation bursts in CNKI; **(B)** The 25 keywords with the strongest citation bursts in the WoS core database. The blue line indicates the time axis, whereas the red segment on the blue time axis indicates burst detection, along with the start year, end year, and burst duration.

Table 5. The top 5 productive authors ranked by the numbers of publications

Article counts	Author	Rank	Article counts
26	Taruscio, Domenica	1	27
25	Boycott, Kym M	2	15
17	Baynam, Gareth	3	11
11	Robinson, Peter N	4	11
11	Lochmueller, Hanns	5	10
-	-	6	10

This indicates that contemporary research places greater emphasis on multi-team collaboration and multi-center studies. Taruscio, Domenica, Boycott, Kym M, and others have contributed to rare disease research for a long time, consistently producing research output, and they are important scholars in this field.

The author collaboration network reflected in the CNKI database is relatively close but shows clear stages (Figure 7A). In 2013-2014, Pei-Wen Wang and Jin-Ping Xie were the central authors in the network. The collaboration network centered around Shu-Yang Zhang, Bo Zhang, and Meng-Chun Gong lasted for a longer duration. The publication volume clearly indicated that this team has made outstanding contributions to the field of rare diseases (Table 5). However, their output decreased significantly in the past 3 years, and the formation of novel collaboration networks is not yet apparent.

3.5. Co-cited references

Highly cited references can indicate the 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. The top three co-cited references were "Estimating cumulative point prevalence of rare diseases: analysis of the Orphanet database (*14*)", "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology (*15*)", and "The mutational constraint spectrum quantified from variation in 141,456 humans (*16*)" (Table 6). The first-ranked reference was cited as many as 156 times. We used the Orphanet database to estimate the cumulative point prevalence of rare diseases. Of the 6,172 unique rare diseases, 71.9% were genetic, and 69.9% manifest in

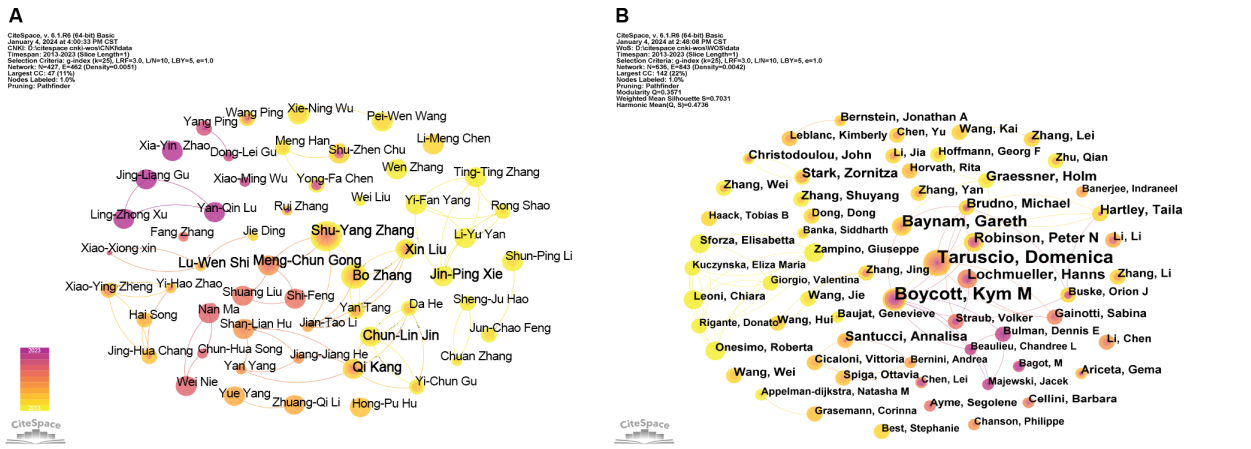

Figure 7. Co-author analysis. (A) Co-author network in the CNKI database; **(B)** Co-author network in the WoS core database. The circle indicates the author; the larger the size of the circle, the greater the number of publications from the author. The links between the nodes indicate cooperation between the authors.

Table 6. The top 10 cited references with the highest cited frequency

Freq	Burst	Centrality	Author	Source	DOI
170	28.18	0.02	Wakap SN	EUR J HUM GENET	10.1038/s41431-019-0508-0
110	34.05	0.01	Richards S	GENET MED	10.1038/gim.2015.30
86	13.74	0.01	Karczewski KJ	NATURE	10.1038/s41586-020-2308-7
83	20.45	0.01	Lek M	NATURE	10.1038/nature19057
60	3.62	0.02	Wright CF	NAT REV GENET	10.1038/nrg.2017.116
58	9.74	0.03	Ferreira CR	AM J MED GENET A	10.1002/ajmg.a.61124
57	0	0.02	Haendel M	NAT REV DRUG DISCOV	10.1038/d41573-019-00180-y
49	3	0.03	Boycott KM	AM J HUM GENET	10.1016/j.ajhg.2017.04.003
45	5.89	0.01	Landrum MJ	NUCLEIC ACIDS RES	10.1093/nar/gkx1153
43	17.32	0.02	Boycott KM	NAT REV GENET	10.1038/nrg3555
43	0	0.01	Rentzsch P	NUCLEIC ACIDS RES	10.1093/nar/gky1016

Table 7. The top 15 cited references with the strongest citation bursts

Freq	Burst	Centrality	Author	Source
110	34.05	0.01	Richards S	GENET MED
170	28.18	0.02	Wakap SN	EUR J HUM GENET
83	20.45	0.01	Lek M	NATURE
43	17.32	0.02	Boycott KM	NAT REV GENET
86	13.74	0.01	Karczewski KJ	NATURE
31	11.89	0.04	Kircher M	NAT GENET
26	10.43	0.01	Köhler S	NUCLEIC ACIDS RES
35	10.06	0.04	Richter T	VALUE HEALTH
23	9.91	0	Page MJ	PLOS MED
58	9.74	0.03	Ferreira CR	AM J MED GENET A
34	9.07	0.05	Köhler S	NUCLEIC ACIDS RES
34	9.06	0.02	Philippakis AA	HUM MUTAT
32	9.04	0.01	Zurynski Y	ORPHANET J RARE DIS
21	8.42	0.07	Girdea M	HUM MUTAT
21	8.42	0.02	Yang YP	NEW ENGL J MED

childhood. The second- and third-ranked references were both related to genomics.

Table 7 presents further analysis of hot topics and progress in the field of rare diseases over the past 10 years. This study selected the top 15 references with the strongest citation bursts. Among them, six articles were related to genomics (*15-20*), and three articles reported database-related content, including those related to the Orphanet (*14*) and HPO databases (*21, 22*). Two articles introduced disease-related platforms, namely The Matchmaker Exchange (*23*), a platform for rare disease gene discovery, and PhenoTips (*24*), a software for patient phenotype analysis for clinical and research purposes. The remaining articles cover topics

Table 8. The top 4 cited references with the highest centrality

Freq	Burst	Centrality	Author	Source	DOI
22	3.9	0.11	Benson MD	NEW ENGL J MED	10.1056/NEJMoa1716793
13	6.73	0.09	Bamshad MJ	NAT REV GENET	10.1038/nrg3031
8	0	0.09	Gurovich Y	NAT MED	10.1038/s41591-018-0279-0
21	8.42	0.07	Girdea M	HUM MUTAT	10.1002/humu.22347

such as writing standards for reviews (*25*), disease terms definition (*2*), social science research (*21*), and reviews (*26*) on rare diseases. Evidently, genomic research is a hot topic in the field of rare diseases. Contribution to disease diagnosis and treatment using big data and artificial intelligence is also a growing trend. With social progression, minority groups are gradually receiving more attention, leading to an increase in social science research on rare disease populations.

Centrality can indicate the importance and influence of literature in a specific field. In this study, the top four cited references with the greatest centrality were selected (Table 8). Among them, one article had a centrality ≥ 0.1. This article introduced the use of exome sequencing to identify disease-causing genes (*27*), reaffirming that genetics is a research hotspot and development trend in this field.

4. Discussion

4.1. Comparison of domestic and foreign databases

An analysis of the progressive knowledge graph in the field of rare diseases over the past 10 years shows that the field of rare disease research is entering a phase of rapid development. However, there are certain differences in the research data between domestic and foreign databases. In terms of the number of publications, the number of publications included in WoS over the past 10 years is ten times greater than that in CNKI. This could be explained by the fact that the publications were from global journals, whereas most journals in CNKI were from China. In terms of the growth rate, WoS has been growing rapidly since 2018, whereas CNKI showed a significant growth acceleration in 2022. This may be attributed to high-quality research from China likely being published internationally.

From the perspective of focus, the frequency ranking of keywords and clustering in CNKI indicate that domestic journals in China focus more on rare disease drugs, whereas foreign journals focus on genetics and therapies. China's policies may help explain this difference. European countries started using Managed Entry Agreements (MEAs) to manage rare disease drugs as early as 2013 or even before that. National authorities have the flexibility to adjust funding support, and pharmaceutical companies can ensure the normal market access of drugs considering cost-benefit conditions (*28*). In recent years, policies have further improved, and some countries have developed specific programs to evaluate the approval of rare disease drugs. For example, in Italy, the approval process for orphan drugs is stated to be completed within 100 days (*29*). In comparison, China's corresponding policies are less advanced. Following the announcement of the first batch of rare disease catalogs in 2018, in April 2019, *The Drug Administration Law of the People's Republic of China* was enacted to encourage development of innovative rare disease drugs and prioritize evaluation of approval (*30*). However, it did not specify a timeframe.

Policy orientation is closely related to academic development. Developed countries such as Europe and the United States have established comprehensive and mature systems for implementing research, approval, and market access of rare disease drugs. These countries are also the primary forces in research and production of rare disease drugs. Therefore, they conduct more genetic research, uncovering disease mechanisms and explore new signaling pathways and mechanisms of action for the development of new drugs and therapies. Conversely, in China, most rare disease drugs are imported, and issues related to drug accessibility and market access still require improvement, despite the introduction of multiple policies in recent years to encourage independent drug research and development. However, the process is lengthy, and significant achievements may only be reported years later.

4.2. Research hotspots in the field of rare diseases

4.2.1. Rise of social science research

Patients with rare diseases often bear a dual burden of physical and psychological challenges. In the era of the biological-psychological-social medicine model (*31*), the diagnosis and treatment of diseases are no longer limited to technical aspects; it is also important to focus on the psychological well-being and social identity of patients. This has led to the development of medical humanities research in the field of rare diseases. Among the top 15 highly cited references, Zurynski Y's cross-sectional study highlighted the reasons and consequences of delayed diagnosis in children with rare diseases. It emphasizes the need for healthcare professionals to provide psychological support to patients and for parents to prioritize genetic counseling and opt for rare disease-related education (*32*). Disease burden is also a prominent theme of research. In the

2019 US Rare Disease Economic Burden Assessment, excess expenditures were primarily attributed to hospital inpatient care, prescription drugs, and productivity losses in the labor market owing to absenteeism and early retirement (*33*).

In addition, the keyword "quality of life" appears over 200 times in the WoS database. A survey on the survival conditions of rare disease populations showed a positive correlation between informal social support and quality of life. Patients who received social assistance had better quality of life in the psychological and social domains than those who did not (*34*). "Management" and "healthcare coverage" are also frequent keywords. Countries worldwide are continuously updating their policies related to rare diseases (*35,36*). Processes to integrate rare disease populations into mainstream society are a global issue. In May 2021, the United Nations passed its first resolution on "Addressing the Challenges of Rare Disease Patients and Their Families," which covers various aspects, such as education, employment, poverty, gender inequality, and support for inclusion of patients with rare diseases in multiple societal dimensions (*37*). As social citizens, rare disease populations have the right to enjoy social welfare policies, equal access to education, and employment opportunities. This can help improve their community awareness and self-acceptance. At the societal level, it can alleviate socioeconomic and management burdens, indicating the progress of human civilization.

4.2.2. Application of artificial intelligence and big data

Owing to the small size of the rare disease population, misdiagnosis and mistreatment are common. Use of big data aids sharing and linkage of patient information worldwide. Using the powerful computational capabilities of artificial intelligence, the relationship between genes and disease onset can be determined, and potential drug targets can be explored. Therefore, establishing a rare disease information database is of great significance for diagnosis, treatment, and research. Orphanet is a widely used database with information on rare disease medical classifications, gene information, and epidemiological indicators, among other data. The most highly cited reference in this context uses accumulated data collected by Orphanet to estimate disease prevalence (*14*). The Matchmaker Exchange (MME), which mentioned in a high-burst cited reference, is a federated network connecting databases of genotypes and rare phenotypes using a common application programming interface (*23*). By 2022, which was 7 years since it was founded, the network had data from over 120,000 cases provided by more than 12,000 volunteers from 98 countries, with over 13,520 unique gene-to-gene matches established. New gene-disease connections are discovered every day (*38*). Over the past decade, artificial intelligence has been used in diagnosis and treatment of various diseases. For example, machine learning algorithms have been used to predict the clinical severity of progressive supranuclear palsy (PSP) by analyzing diffusion tensor imaging characteristics of the brain (*39*). Automatic assessment tools have been developed to predict speech disorders in patients with ALS based on speech acoustics and pronunciation samples (*40*). Big data and machine learning demonstrate the immense potential of artificial intelligence in the field of rare diseases. However, widespread implementation of these tools in clinical applications requires regulatory frameworks, evidence from clinical trials, and compliance with ethical guidelines (*41*).

Disease registration is also focused on both domestically and internationally. Owing to the challenges in collecting rare disease data, individualized tracking of each patient is typically performed using case registries. Collecting epidemiological data through registries helps monitor quality of management and patient prognosis processes while providing convenience for clinical research. Developed countries such as Europe and the United States have achieved a high level of refinement, regionalization, and networking in their registries. Examples include the European Rare Kidney Disease Registry (ERKReg) (*42*) and the European registry for patients with McArdle disease and other muscle glycogenoses (EUROMAC registry) (*43*) . In contrast, in China, rare disease registries have been developed relatively lately. The National Rare Disease Registry System (NRDRS), established in 2016, is expected to have a positive impact on epidemiological research within the country (*44*).

4.2.3. Genetic Research

Genomic research is one of the most important hotspots in the field of rare diseases. In keyword analysis, keyword clustering in CNKI includes the label of "gene therapy", whereas the high-frequency keywords in WoS include "mutation", "expression", and "gene", all of which appear more than 300 times. Among highly cited references in WoS, approximately half of the articles are related to genomics. The highest-ranked highly cited reference is a standard and guideline on interpretation of sequence variations. It uses various types of variant evidence, including population data, computational data, functional data, and segregation data, to classify variations into five standards: "pathogenic", "likely pathogenic", "uncertain significance", "likely benign," and "benign" (*15*). This article serves as a foundational declaration in the field of genomics. However, owing to the significant proportion of genetic diseases among rare diseases, development of this guideline also contributes to the standardized representation of gene mutations in rare disease genomics research, which is significant in determining gene-disease relationships.

Genomics is not only used to explore the etiology of rare diseases but also plays a role in development of therapeutic drugs and treatments as research progresses. Luxturna, a gene therapy drug using adeno-associated virus (AAV) vectors, is administered through subretinal injection to treat inherited retinal dystrophy caused by RPE65 protein deficiency. Patisiran is an oligonucleotide-based therapy for familial transthyretin amyloidosis (FTA) that inhibits misfolding of transthyretin protein, thereby impeding disease progression. Rare disease drugs based on gene therapies such as CAR-T and retroviral vectors have also been approved and launched in foreign markets (*45*). As more mechanisms are discovered, the scope of gene therapy will continue to expand, bringing new hope for treatment of rare diseases. However, adoption of gene therapy in China remains limited, and further research and development are needed before it can be widely used in clinical practice.

4.3. Key scholars and institutions in the field of rare diseases

In the CNKI database, the core author Shu-Yang Zhang stands out with a significantly high publication volume. As the President of Peking Union Medical College Hospital (PUMCH) and Vice President of Peking Union Medical College, the collaborative network centered around Shu-Yang Zhang has consistently produced output over the past decade. Research institutions associated with Shu-Yang Zhang, namely PUMCH and Peking Union Medical College, have also achieved remarkable results in the field of rare disease research in China. Shu-Yang Zhang is a leading scholar in the field of rare diseases in China, with most of his published articles focusing on rare cardiovascular diseases or rare disease policies in China. Examples include "Progress in the Diagnosis and Treatment of Rare Cardiovascular Diseases (*46*)" and "Construction and Application of China's National Rare Disease Registration System (*47*) ". The articles published by Shu-Yang Zhang are mostly systematic reviews and expert consensus articles. PUMCH, as the host institution of the National Key Laboratory for Severe and Rare Diseases, encompasses various disciplines such as oncology, surgery, obstetrics, and gynecology. Its achievements span multiple types of publications, including case reports, systematic reviews, and clinical studies.

In the WoS database, Taruscio D and Boycott KM are the authors with the greatest number of publications. The collaborative network centered around these authors is the only evident collaboration network in the WoS database. Taruscio D is affiliated with the Istituto Superiore di Sanita in Rome, Italy, whereas Boycott KM is affiliated with the Children's Hospital of Eastern Ontario in Canada. Both researchers have interests in genetics and experimental medicine. However, Taruscio D also focuses on public health management and healthcare services related to rare diseases, whereas Boycott KM has a stronger emphasis on genomics.

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 Diseases (*48*)" and "Rare-disease genetics in the era of next-generation sequencing: discovery to translation (*18*)" . However, Taruscio D does not have any highly cited articles. This indicates that Boycott KM is a key scholar in the field of international rare disease research and an important member of the International Rare Diseases Research Consortium (IRDiRC).

From an institutional perspective, Chinese research institutions have demonstrated significant achievements in scientific research globally. However, collaboration with international institutions is yet to be established. Domestic and international institutions and scholars should strengthen communication and cooperation, learn from each others' experience, and collectively promote the development of the field of rare diseases.

5. Conclusion

This study used CiteSpace to analyze outstanding academic achievements in the field of rare diseases over the past decade. By visualizing and mapping key institutions, authors, keywords, and high-citation information, it identified research hotspots and development trends from different perspectives, providing insights for global advancement of the field of rare diseases.

However, there are certain limitations to this study. First, the literature selection process was manual, which may have introduced subjectivity in the inclusion of literature, organization of literature information, and merging of keywords, potentially leading to the loss of some relevant articles. Secondly, CiteSpace uses its in-built algorithms for keyword merging and clustering, which may not fully represent all research content within a particular cluster. Additionally, research on specific rare diseases often does not directly include the term "rare diseases," which implies that we may have excluded some studies focusing on individual rare diseases. Furthermore, owing to limitations in the CiteSpace software and CNKI database, the direct export of citation data was not possible, resulting in a lack of co-citation analysis for Chinese literature in this study.

Funding: This work was supported by Shanghai Municipal Philosophy and Social Sciences Planning Project: Research on the Multidimensional Construction of a Multi-level Healthcare System for Rare Diseases in Shanghai under the New Situation (grant number: 2022BGL002); 2022 Annual Talent Development Program of the Shanghai Municipal Health Commission: Research on Enhancing the Prevention, Treatment, and Assurance Capacity of Rare Diseases in Shanghai

under the New Development Situation (grant number: 2022YQ058).

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

References

1. Schieppati A, Henter JI, Daina E, Aperia A. Why rare diseases are an important medical and social issue. Lancet. 2008; 371:2039-2041.
2. Richter T, Nestler-Parr S, Babela R, Khan ZM, Tesoro T, Molsen E, Hughes DA; International Society for Pharmacoeconomics and Outcomes Research Rare Disease Special Interest Group. Rare disease terminology and definitions-a systematic global review: Report of the ISPOR Rare Disease Special Interest Group. Value Health. 2015; 18:906-914.
3. Joint Meeting of the Chairpersons of Rare Disease Societies. A report on the definition of rare diseases in China (2021). *http://www.diagnoschina.com/research/detail/nid-83*. (accessed November 10, 2024). (in Chinese)
4. Chung CCY; Hong Kong Genome Project; Chu ATW, Chung BHY. Rare disease emerging as a global public health priority. Front Public Health. 2022; 10:1028545.
5. Groft SC, Posada M, Taruscio D. Progress, challenges and global approaches to rare diseases. Acta Paediatr. 2021; 110:2711-2716.
6. Delaye J, Cacciatore P, Kole A. Valuing the "burden" and impact of rare diseases: A scoping review. Front Pharmacol. 2022; 13:914338.
7. Fu MY, Guan XD, Wei GX, Xin XX, Shi LW. Medical service utilisation, economic burden and health status of patients with rare diseases in China. Chinese Pharmaceutical Journal. 2018; 27:361-369.
8. 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 March 1, 2024). (in Chinese)
9. Lu Y, Han J. The definition of rare disease in China and its prospects. Intractable Rare Dis Res. 2022; 11:29-30.
10. 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 March 1, 2024). (in Chinese)
11. Dharssi S, Wong-Rieger D, Harold M, Terry S. Review of 11 national policies for rare diseases in the context of key patient needs. Orphanet J Rare Dis. 2017; 12:63.
12. Li Y, Zhang Y, Zeng K, Zhang S. Comparison of different literature information analysis tools. Chinese Journal of Medical Library and Information Science. 2015; 24:41-47. (in Chinese)
13. Synnestvedt MB, Chen C, Holmes JH. CiteSpace II: visualization and knowledge discovery in bibliographic databases. AMIA Annu Symp Proc. 2005; 2005:724-728.
14. 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.
15. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL; ACMG Laboratory Quality Assurance Committee. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015; 17:405-424.
16. Karczewski KJ, Francioli LC, Tiao G, *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020; 581:434-443.
17. Lek M, Karczewski KJ, Minikel EV, *et al.* Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016; 536:285-291.
18. 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.
19. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. Nat Genet. 2014; 46:310-315.
20. Yang Y, Muzny DM, Reid JG, *et al.* Clinical whole-exome sequencing for the diagnosis of mendelian disorders. N Engl J Med. 2013; 369:1502-1511.
21. Köhler S, Vasilevsky NA, Engelstad M, *et al.* The human phenotype ontology in 2017. Nucleic Acids Res. 2017; 45:D865-D876.
22. Köhler S, Doelken SC, Mungall CJ, *et al.* The human phenotype ontology project: Linking molecular biology and disease through phenotype data. Nucleic Acids Res. 2014; 42:D966-D974.
23. Philippakis AA, Azzariti DR, Beltran S, *et al.* The matchmaker exchange: A platform for rare disease gene discovery. Hum Mutat. 2015; 36:915-921.
24. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, Chitayat D, Faghfoury H, Meyn MS, Ray PN, So J, Stavropoulos DJ, Brudno M. PhenoTips: Patient phenotyping software for clinical and research use. Hum Mutat. 2013; 34:1057-1065.
25. Page MJ, McKenzie JE, Bossuyt PM, *et al.* The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. PLoS Med. 2021; 18:e1003583.
26. Ferreira CR. The burden of rare diseases. Am J Med Genet A. 2019; 179:885-892.
27. Benson MD, Waddington-Cruz M, Berk JL, *et al.* Inotersen treatment for patients with hereditary transthyretin amyloidosis. N Engl J Med. 2018; 379:22-31.
28. Morel T, Arickx F, Befrits G, Siviero P, van der Meijden C, Xoxi E, Simoens S. Reconciling uncertainty of costs and outcomes with the need for access to orphan medicinal products: a comparative study of managed entry agreements across seven European countries. Orphanet J Rare Dis. 2013; 8:198.
29. 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.
30. National Medical Products Administration. Drug Administration Law of the People's Republic of China. *https://www.nmpa.gov.cn/xxgk/fgwj/flxzhfg/20190827083801685.html*. (accessed March 1, 2024). (in Chinese)
31. Song P, Tang W. Emphasizing humanities in medical education: Promoting the integration of medical scientific spirit and medical humanistic spirit. Biosci Trends. 2017;

11:128-133.
32. Zurynski Y, Deverell M, Dalkeith T, Johnson S, Christodoulou J, Leonard H, Elliott EJ; APSU Rare Diseases Impacts on Families Study Australian children living with rare diseases: experiences of diagnosis and perceived consequences of diagnostic delays. Orphanet J Rare Dis. 2017; 12:68.
33. Yang G, Cintina I, Pariser A, Oehrlein E, Sullivan J, Kennedy A. The national economic burden of rare disease in the United States in 2019. Orphanet J Rare Dis. 2022; 17:163.
34. Gao J, Ma Z. Study on the relationship between social support and quality of life among patients with rare diseases. Chinese Journal of Health Policy. 2020; 13:38-44. (in Chinese)
35. Ying Z, Gong L, Li C. An update on China's national policies regarding rare diseases. Intractable Rare Dis Res. 2021; 10:148-153.
36. Moliner AM, Waligora J. The European Union policy in the field of rare diseases. Adv Exp Med Biol. 2017; 1031:561-587.
37. Rare Diseases International. UN resolution on persons living with a rare disease. *https://www.rarediseasesinternational.org/un-resolution/* (accessed March 1, 2024)
38. Boycott KM, Azzariti DR, Hamosh A, Rehm HL. Seven years since the launch of the Matchmaker Exchange: The evolution of genomic matchmaking. Hum Mutat. 2022; 43:659-667.
39. Chen YL, Zhao XA, Ng SH, Lu CS, Lin YC, Cheng JS, Tsai CC, Wang JJ. Prediction of the clinical severity of progressive supranuclear palsy by diffusion tensor imaging. J Clin Med. 2019; 9:40.
40. Wang J, Kothalkar PV, Kim M, Bandini A, Cao B, Yunusova Y, Campbell TF, Heitzman D, Green JR. Automatic prediction of intelligible speaking rate for individuals with ALS from speech acoustic and articulatory samples. Int J Speech Lang Pathol. 2018; 20:669-679.
41. Visibelli A, Roncaglia B, Spiga O, Santucci A. The impact of artificial intelligence in the odyssey of rare diseases. Biomedicines. 2023; 11:887.
42. Bassanese G, Wlodkowski T, Servais A, *et al.* The European Rare Kidney Disease Registry (ERKReg): Objectives, design and initial results. Orphanet J Rare Dis. 2021; 16:251.
43. Pinós T, Andreu AL, Bruno C, *et al.* Creation and implementation of a European registry for patients with McArdle disease and other muscle glycogenoses (EUROMAC registry). Orphanet J Rare Dis. 2020; 15:187.
44. Guo J, Liu P, Chen L, Lv H, Li J, Yu W, Xu K, Zhu Y, Wu Z, Tian Z, Jin Y, Yang R, Gu W, Zhang S Administrative Group of National Rare Diseases Registry System of China. National Rare Diseases Registry System (NRDRS): China's first nation-wide rare diseases demographic analyses. Orphanet J Rare Dis. 2021; 16:515.
45. Papaioannou I, Owen JS, Yáñez-Muñoz RJ. Clinical applications of gene therapy for rare diseases: A review. Int J Exp Pathol. 2023; 104:154-176.
46. Zhang SY, Tian Z, Zhang SY, *et al*. Advances in the Diagnosis and Treatment of Rare Cardiovascular Diseases. Journal of Rare Diseases. 2023; 2:1-5. (in Chinese)
47. Guo J, Liu P, Jing ZC, Liu J, Cheng J, Ding J, Gu W, Chen L, Zhu Y, Zhang S. Construction and application of national rare diseases registry system of China. Journal of Rare Diseases. 2022; 1:7-12.
48. 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.

Received November 19, 2024; Revised January 13, 2025; Accepted February 6, 2025.

[§]These authors contributed equally to this work.
**Address correspondence to:*
Qi Kang, Shanghai Health Development Research Center (Shanghai Medical Information Research Center), No.1477 West Beijing Road, Shanghai, China.
E-mail: kangqi@shdrc.org

Peihao Yin, Putuo Hospital, Shanghai University of Traditional Chinese Medicine, No.164 Lanxi Road, Putuo District, Shanghai, China.
E-mail: yinpeihao@shutcm.edu.cn

Released online in J-STAGE as advance publication February 12, 2025.