

Worldwide Trends in Registering Real-World Studies at ClinicalTrials.gov: A Cross-Sectional Analysis

Yuanxiao Li^{1,*}, Ying Tian^{2,*}, Shufen Pei³, Baoyuan Xie², Xiaonan Xu¹, Bin Wang¹

¹Department of Pediatric Gastroenterology, Lanzhou University Second Hospital, Lanzhou, People's Republic of China; ²Department of Clinical Medicine, Lanzhou University Second Hospital, Lanzhou, People's Republic of China; ³Department of Clinical Medicine, North Sichuan Medical College, Nanchong, People's Republic of China; ⁴Department of Infectious Diseases, Second Affiliated Hospital, Zhejiang University School of Medicine, Hangzhou, People's Republic of China

*These authors contributed equally to this work

Correspondence: Bin Wang, 88 Jie Fang Lu, Shangcheng District, Hangzhou, Zhejiang, 310009, People's Republic of China, Email wangbin2022@zju.edu.cn

Objective: The purpose of this study was to characterize real-world studies (RWSs) registered at ClinicalTrials.gov to help investigators better conduct relevant research in clinical practice.

Methods: A retrospective analysis of 944 studies was performed on February 28, 2023.

Results: A total of 944 studies were included. The included studies involved a total of 48 countries. China was the leading country in terms of the total number of registered studies (37.9%, 358), followed by the United States (19.7%, 186). Regarding intervention type, 42.4% (400) of the studies involved drugs, and only 9.1% (86) of the studies involved devices. Only 8.5% (80) of the studies mentioned both the detailed study design type and data source in the "Brief Summary". A total of 49.4% (466) of studies had a sample size of 500 participants and above. Overall, 63% (595) of the studies were single-center studies. A total of 213 conditions were covered in the included studies. One-third of the studies (32.7%, 309) involved neoplasms (or tumors). China and the United States were very different regarding the study of different conditions.

Conclusion: Although the pandemic has provided new opportunities for RWSs, the rigor of scientific research still needs to be emphasized. Special attention needs to be given to the correct and comprehensive description of the study design in the Brief Summary of registered studies, thereby promoting communication and understanding. In addition, deficiencies in ClinicalTrials.gov registration data remain prominent.

Keywords: real-world data, real-world evidence, real-world study, study design, ClinicalTrials.gov

Introduction

Real-world data (RWD) are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.¹ Real-world evidence (RWE) is clinical evidence regarding the usage and potential benefits or risks of a medical product derived from the analysis of RWD.¹ A real-world study (RWS) collects RWD in a real-world environment and obtains RWE of the use value and potential benefits or risks of medical products through analysis.² The coronavirus disease 2019 (COVID-19) pandemic has led to the disruption of traditional clinical trials and the activation of a large number of clinical studies. In recent years, the Food and Drug Administration (FDA) has issued numerous RWE-related guidelines to support regulatory decisions.³ A number of previous studies have assessed the characteristics of COVID-19 studies registered in the American Clinical Trials Registry (ClinicalTrials.gov).^{4,5} Previous studies have reviewed in detail the types of clinical research for COVID-19, particularly with regard to issues of data collection and sharing in terms of real-world data.⁶ However, the characteristics of RWSs registered worldwide for different diseases are still unknown. Efforts have been made to standardize RWD/RWE for some time. In guiding researchers to design and implement RWD research, some standards have been put forth, such as guidelines,^{7,8} processes,^{9–11} and templates.^{12,13} Although RWSs differ greatly from traditional clinical trials in terms of research characteristics, researchers still adopt the habit of writing traditional randomized controlled trial (RCT) research protocols. Since there is still a large gap in the

writing of standardized RWS research protocols, it is necessary to analyze the problems in RWS registry research to provide insights for the improvement of follow-up guidelines. Therefore, the purpose of this study was to comprehensively evaluate RWSs registered at ClinicalTrials.gov to help investigators better conduct relevant research in clinical practice.

Methods

Search Strategy

On February 28, 2023, we retrieved data using the ClinicalTrials.gov application programming interface (API) (version 1.01.05). The API contained 328 research fields (see [Supplementary File](#)). After filtering out fields with a large number of missing values that were not important, we screened only 97 valuable research fields (see [Supplementary File](#)). ClinicalTrials.gov limits API queries to 1000 records and 20 fields per use. Using the code loop, all data can be obtained within minutes. The search expression we used to build the query Uniform Resource Locator (URL) was “real-world data” OR “real-world study” OR “real-world evidence” OR “real-world research”. Python (version 3.9.12, Python Software Foundation, Fredericksburg, Virginia, USA) was used to send requests and parse returned query data. Retrieved data were manually checked to exclude clinical trials. The entire process was completed independently by two authors, and finally, the results were compared. The Python code for scraping data from ClinicalTrials.gov using its API is provided in our online [Supplementary Material](#) (see [Supplementary File](#)). This code is publicly available and can be used to make data requests with any search term.

Data Summarization and Visualization

Python was used to perform statistical analysis. Descriptive statistics were used to analyze the data. Sankey diagrams were used to show the flow between two or more categories, in which the width of each individual element was proportional to the flow rate. To ensure that only the most relevant flows were included in the Sankey diagram, we applied a threshold screening approach. This approach allowed us to visualize the most relevant flows and avoid cluttering the diagram with nonsignificant flows. We showed Sankey diagrams for conditions and countries, as well as Sankey diagrams for multicenter research numbers and countries. For data requested from the API, the field “LocationCountry” recorded the country in which each research center was located in a long string. To avoid double counting, only the country of the study sponsor was included as the country of registration for each study in this analysis. We used the same method for counts from studies involving multiple countries.

Results

Selection of Eligible Studies

A total of 960 RWSs were registered at ClinicalTrials.gov. Sixteen studies were excluded due to ineligibility. Finally, a total of 944 studies were included. The flow diagram of the study screening process is shown in [Figure 1](#).

General Characteristics

[Table 1](#) presents the general characteristics of the included studies. We divided the studies into two groups by status, including completed and other statuses. In 63.2% (597) of the studies, the observational model was designed as a cohort. A total of 46.8% (442) of the studies were prospective. Regarding intervention type, 42.4% (400) of the studies involved drugs, and only 9.1% (86) of the studies involved devices. In 40.7% (384) of the studies, the lead sponsor type was industry. More than one-third (37.6%, 355) of the studies were in the recruiting stage, and 27.4% (259) of the studies were completed; therefore, approximately one-third of the studies had not yet recruited participants. Only 15.1% (143) of studies were supervised by data monitoring committees. A total of 87.3% (824) of the studies were observational. The median (interquartile range) estimated sample size was 481 (158, 2000). In 14% (132) of the studies, the sample size was less than 100 participants. A total of 49.4% (466) of studies had a sample size of 500 participants and above. Overall, 63% (595) of the studies were single-center studies. Only 8.5% (80) of the studies mentioned both the detailed study design type and data source in the “Brief Summary”. A total of 22% (208) of studies defined their specific study types or data sources.

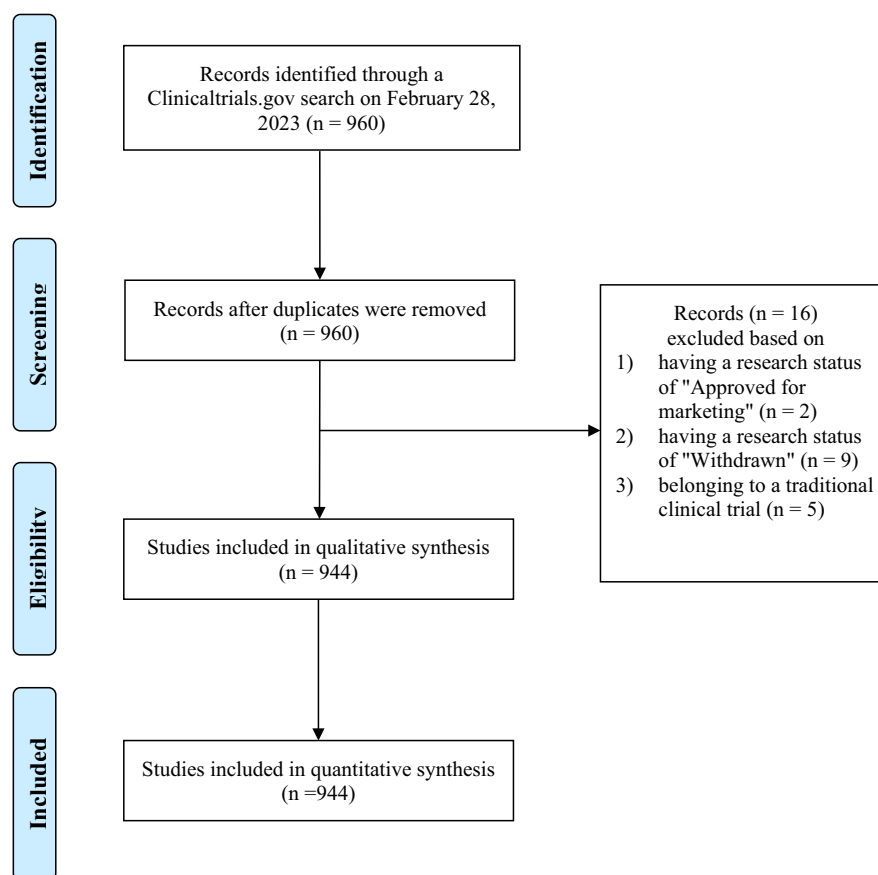


Figure 1 Study selection flow diagram.

Study Conditions

“Condition” and “ConditionMeshTerm” were included in the study fields provided by the ClinicalTrials.gov API. We manually compared the two fields and kept the main standardized condition names. A total of 213 conditions were covered in the included studies. Gastrointestinal neoplasms were the leading condition in the registered studies (8.6%, 81), followed by lung neoplasms (8.4%, 79) and breast neoplasms (6%, 57). Overall, one-third of the studies (32.7%, 309) involved neoplasms (or tumors). The top 20 conditions are shown in Figure 2. The top 10 conditions in China and the United States are shown in Figure 3. Among the registered studies in the United States that included tumors, 22.2%

Table 1 General Characteristics of Real-World Registration Studies

Item	Category	Overall (n=944)	Completed (n=259)	Other Status (n=685)
Design Observational Model, n (%)	Case-Control	29 (3.1)	9 (3.5)	20 (2.9)
	Case Only	91 (9.6)	27 (10.4)	64 (9.3)
	Cohort	597 (63.2)	179 (69.1)	418 (61.0)
	Ecologic or Community	4 (0.4)	1 (0.4)	3 (0.4)
	Case-Crossover	3 (0.3)	–	3 (0.4)
	Family-Based	1 (0.1)	–	1 (0.1)
	Other	98 (10.4)	19 (7.3)	79 (11.5)
	Missing	121 (12.8)	24 (9.3)	97 (14.2)

(Continued)

Table I (Continued).

Item	Category	Overall (n=944)	Completed (n=259)	Other Status (n=685)
Design Time Perspective, n (%)	Cross-Sectional	25 (2.6)	10 (3.9)	15 (2.2)
	Prospective	442 (46.8)	84 (32.4)	358 (52.3)
	Retrospective	290 (30.7)	125 (48.3)	165 (24.1)
	Other	65 (6.9)	15 (5.8)	50 (7.3)
	Missing	122 (12.9)	25 (9.7)	97 (14.2)
Enrollment Type, n (%)	Actual	333 (35.3)	255 (98.5)	78 (11.4)
	Anticipated	611 (64.7)	4 (1.5)	607 (88.6)
Age, n (%)	Adult	18 (1.9)	5 (1.9)	13 (1.9)
	Adult/Older Adult	714 (75.6)	186 (71.8)	528 (77.1)
	Child	21 (2.2)	8 (3.1)	13 (1.9)
	Child/Adult	12 (1.3)	5 (1.9)	7 (1.0)
	Child/Adult/Older Adult	169 (17.9)	52 (20.1)	117 (17.1)
	Older Adult	10 (1.1)	3 (1.2)	7 (1.0)
Sex, n (%)	All	863 (91.4)	242 (93.4)	621 (90.7)
	Female	67 (7.1)	10 (3.9)	57 (8.3)
	Male	12 (1.3)	6 (2.3)	6 (0.9)
	Missing	2 (0.2)	1 (0.4)	1 (0.1)
Study Type, n (%)	Interventional	120 (12.7)	23 (8.9)	97 (14.2)
	Observational	824 (87.3)	236 (91.1)	588 (85.8)
Intervention Type, n (%)	Behavioral	17 (1.8)	6 (2.3)	11 (1.6)
	Biological	7 (0.7)	1 (0.4)	6 (0.9)
	Combination Product	78 (8.3)	19 (7.3)	59 (8.6)
	Device	86 (9.1)	24 (9.3)	62 (9.1)
	Diagnostic Test	14 (1.5)	1 (0.4)	13 (1.9)
	Dietary Supplement	2 (0.2)	1 (0.4)	1 (0.1)
	Drug	400 (42.4)	135 (52.1)	265 (38.7)
	Procedure	38 (4.0)	7 (2.7)	31 (4.5)
	Radiation	8 (0.8)	2 (0.8)	6 (0.9)
	Genetic	5 (0.5)	–	5 (0.7)
	Other	99 (10.5)	32 (12.4)	67 (9.8)
	Missing	190 (20.1)	31 (12.0)	159 (23.2)

(Continued)

Table 1 (Continued).

Item	Category	Overall (n=944)	Completed (n=259)	Other Status (n=685)
Lead Sponsor Class, n (%)	Industry	384 (40.7)	158 (61.0)	226 (33.0)
	Network	11 (1.2)	1 (0.4)	10 (1.5)
	Federal	1 (0.1)	–	1 (0.1)
	Individual	1 (0.1)	–	1 (0.1)
	Other Government	17 (1.8)	1 (0.4)	16 (2.3)
	Other	530 (56.1)	99 (38.2)	431 (62.9)
Overall Status, n (%)	Completed	259 (27.4)	259 (100.0)	–
	Active, not recruiting	95 (10.1)	–	95 (13.9)
	Enrolling by invitation	23 (2.4)	–	23 (3.4)
	Not yet recruiting	129 (13.7)	–	129 (18.8)
	Recruiting	355 (37.6)	–	355 (51.8)
	Terminated	12 (1.3)	–	12 (1.8)
	Unknown status	71 (7.5)	–	71 (10.4)
Accepts Healthy Volunteers, n (%)	Yes	51 (5.4)	7 (2.7)	44 (6.4)
	No	806 (85.4)	233 (90.0)	573 (83.6)
	Missing	87 (9.2)	19 (7.3)	68 (9.9)
Oversight has Data Monitoring Committee, n (%)	Yes	143 (15.1)	21 (8.1)	122 (17.8)
	No	544 (57.6)	179 (69.1)	365 (53.3)
	Missing	257 (27.2)	59 (22.8)	198 (28.9)
Sampling Method, n (%)	Probability Sample	215 (22.8)	49 (18.9)	166 (24.2)
	Nonprobability Sample	609 (64.5)	187 (72.2)	422 (61.6)
	Missing	120 (12.7)	23 (8.9)	97 (14.2)
Study Sample Size	Median [Q1, Q3]	480.5 [157.5, 2000.0]	350.0 [159.0, 2184.0]	500.0 [156.0, 1967.0]
	Missing	0	–	–
Study Sample Size Classification, n (%)	0–49	55 (5.8)	12 (4.6)	43 (6.3)
	50–99	77 (8.2)	27 (10.4)	50 (7.3)
	100–199	133 (14.1)	38 (14.7)	95 (13.9)
	200–299	98 (10.4)	34 (13.1)	64 (9.3)
	300–499	114 (12.1)	33 (12.7)	81 (11.8)
	500–999	124 (13.1)	26 (10.0)	98 (14.3)
	1000–4999	184 (19.5)	43 (16.6)	141 (20.6)
	5000–9999	44 (4.7)	12 (4.6)	32 (4.7)
	10,000–2,133,900	114 (12.1)	33 (12.7)	81 (11.8)
	Missing	1 (0.1)	1 (0.4)	–
Number of Research Centers	Mean (SD)	9.5 (41.2)	13.7 (56.3)	7.8 (33.2)
	Missing, n	46	3	43

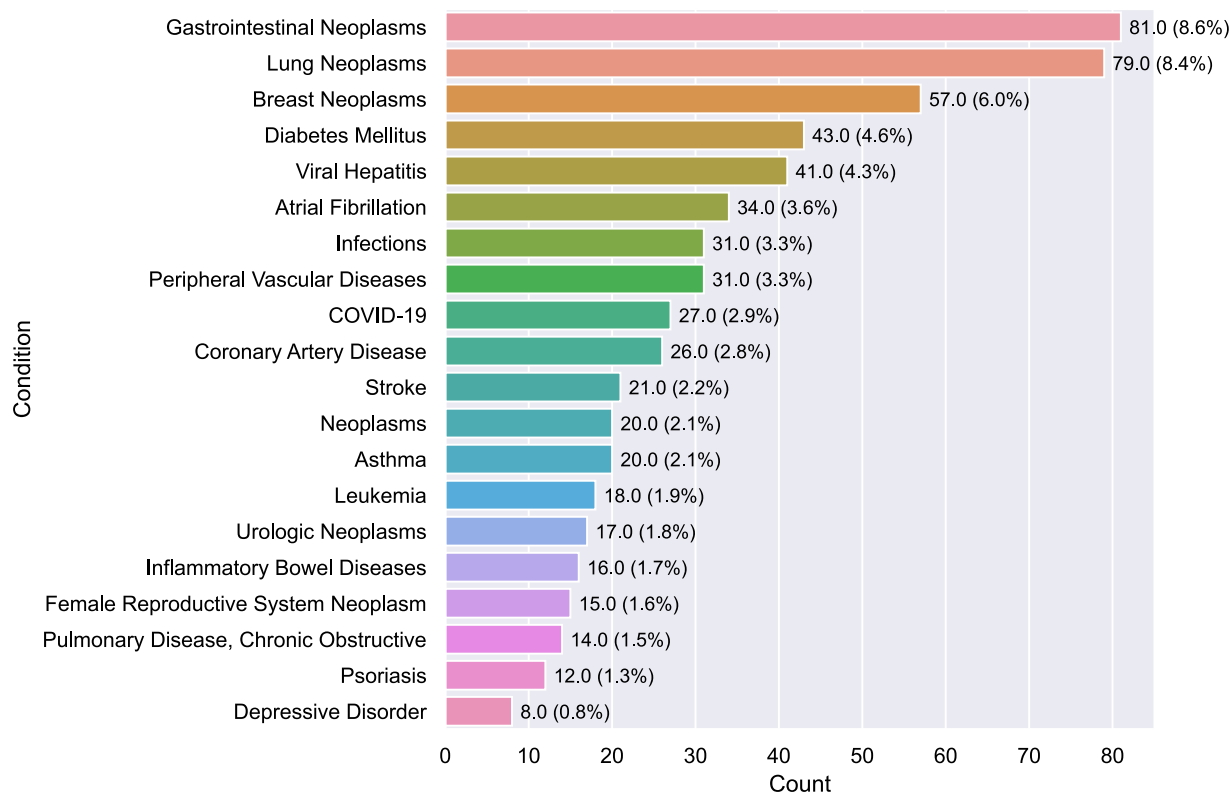


Figure 2 Number of registered studies by condition. Only the top 20 conditions are shown.

(6/27) of the United States studies did not specify particular tumor types, whereas only 3.8% (7/182) of the studies in China did not restrict specific tumor types. Out of all the disease registries that contained information on tumors, 18.5% (5/27) of the research data in the United States came from tumor registry databases. In contrast, only 1.1% (2/182) of research data in China came from tumor registry databases.

Global Distribution

The included studies involved a total of 48 countries. China was the leading country in terms of the total number of registered studies (37.9%, 358), followed by the United States (19.7%, 186). The sum of the number of studies in China and the United States was 57.6%, which was more than half of the total number of studies. The detailed global distribution of the number of included studies is shown in [Figure 4](#). The top 20 countries are shown in [Figure 5](#).

Trends in Study Registration

We counted the number of new studies registered on the ClinicalTrials.gov website and displayed it in a line graph. [Figure 6](#) contains two subgraphs, one for the overall trend and the other grouped by country. In the country grouping, we separated out China and the United States and merged all remaining countries into an “other countries” category. From 2020 to 2022, the total number of registered studies was 60% (568).

Relationship Between Countries and Conditions

We used a Sankey diagram to explore countries that contributed to major condition studies. We excluded flows that represented less than four of the total number of studies flowing from a given country or to a given condition. The filtering thresholds were manually defined based on the complexity of the Sankey diagram. The Sankey diagram is shown in [Figure 7](#).

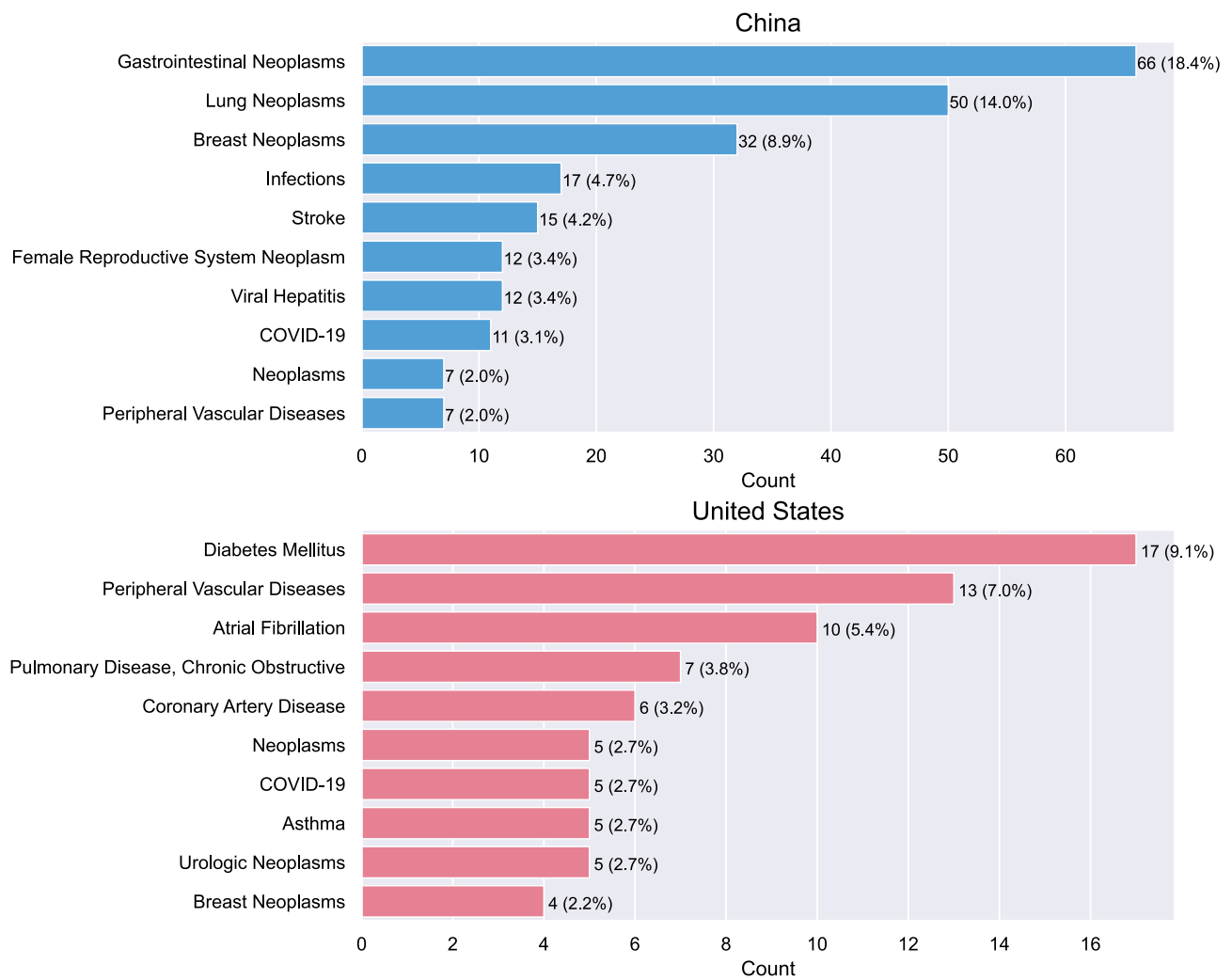


Figure 3 Number of registered studies in China and the United States by condition. Only the top 10 conditions are shown.

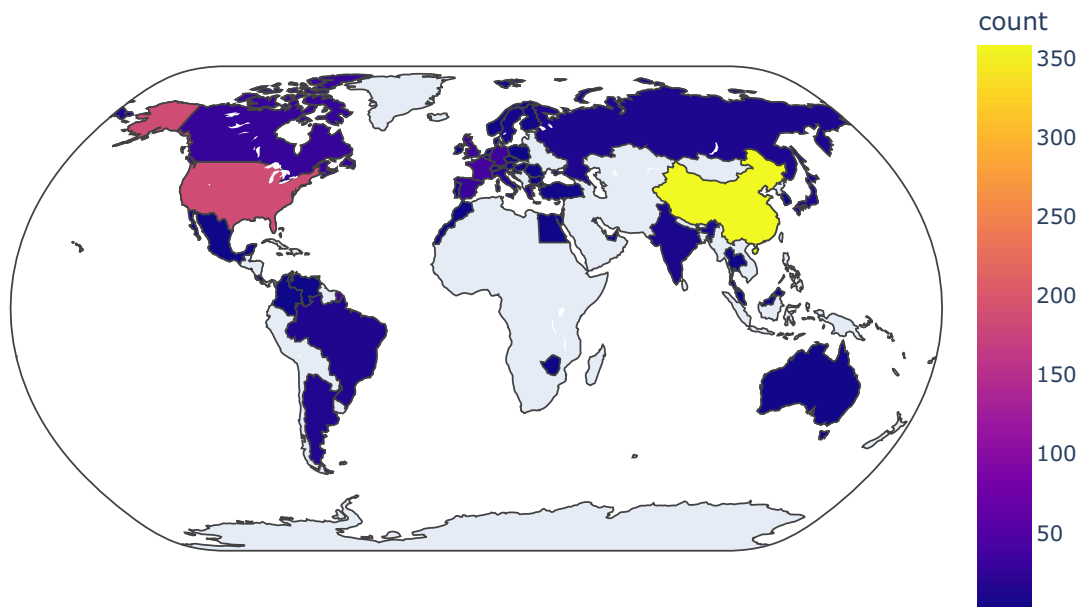


Figure 4 Number of real-world registration studies in each area.

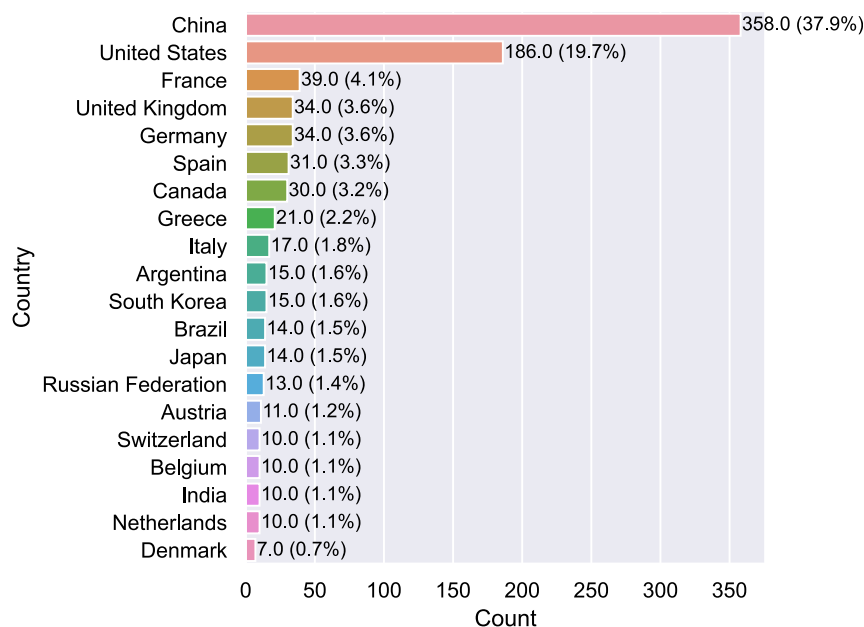


Figure 5 Number of registered studies in the top 20 countries.

Relationship Between Countries and Multicenter Research Numbers

We excluded flows that represented less than ten of the total number of studies flowing from a given country or to a given multicenter research number. Only 3% (29) of the studies had 50 or more research centers. Therefore, the category “ ≥ 50 ” in the study research number classification is not shown in the Sankey diagram (Figure 8).

Discussion

The passage of the 21st Century Cures Act in 2016 provided a milestone for the FDA in the promotion of RWE. This can also be seen from the line chart in Figure 6, which shows 2016 as an important time inflection point. The COVID-19 outbreak that began in 2020 has boosted researchers' interest and acceptance of RWSs. As seen in Figure 3, China and the United States are very different regarding the study of different conditions. The number of registered RWSs shows great differences in tumor diseases. By 2019, a total of 574 cancer registries had been established in China, covering a population of 438 million (accounting for 31.5% of the total population of China).¹⁴ The United States has several national cancer registries that collect data on cancer incidence, mortality, and survival rates.¹³ Some of the major cancer registries in the United States include the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) program¹⁵ and the National Program of Cancer Registries (NPCR). The reason behind this may be related to the differences in disease distributions in the two countries. The incidence rate and mortality rate of most digestive system malignancies in China are higher than those in the United States.¹⁴ Another important reason is that China's national tumor registration platform has not yet been made public, which makes RWE-based oncology research largely dependent on routine diagnosis and treatment data from electronic medical records (EMRs). China's EMR system cannot link the data of multiple medical institutions, unlike the American electronic health record (EHR) system. The number of cancer-related registration studies in China far exceeded that in the United States, mainly due to differences in disease distribution and the acquisition of data sources. Although the number of cancer-related registration studies in the United States is relatively small, it is easier to cover multiple tumor types and convene multiple research centers with the help of tumor registry platforms. The Sankey diagram from country to condition can provide researchers with ideas for selecting diseases when registering multinational multicenter studies.

When talking about RWE or RWD, it is easy to become confused by the terms that describe research design. While the term RWS is not mentioned in the FDA guidelines, the term is derived from the China National Medical Products Administration (NMPA) guidelines on RWE.² These three terms are not entirely newly proposed concepts, as data

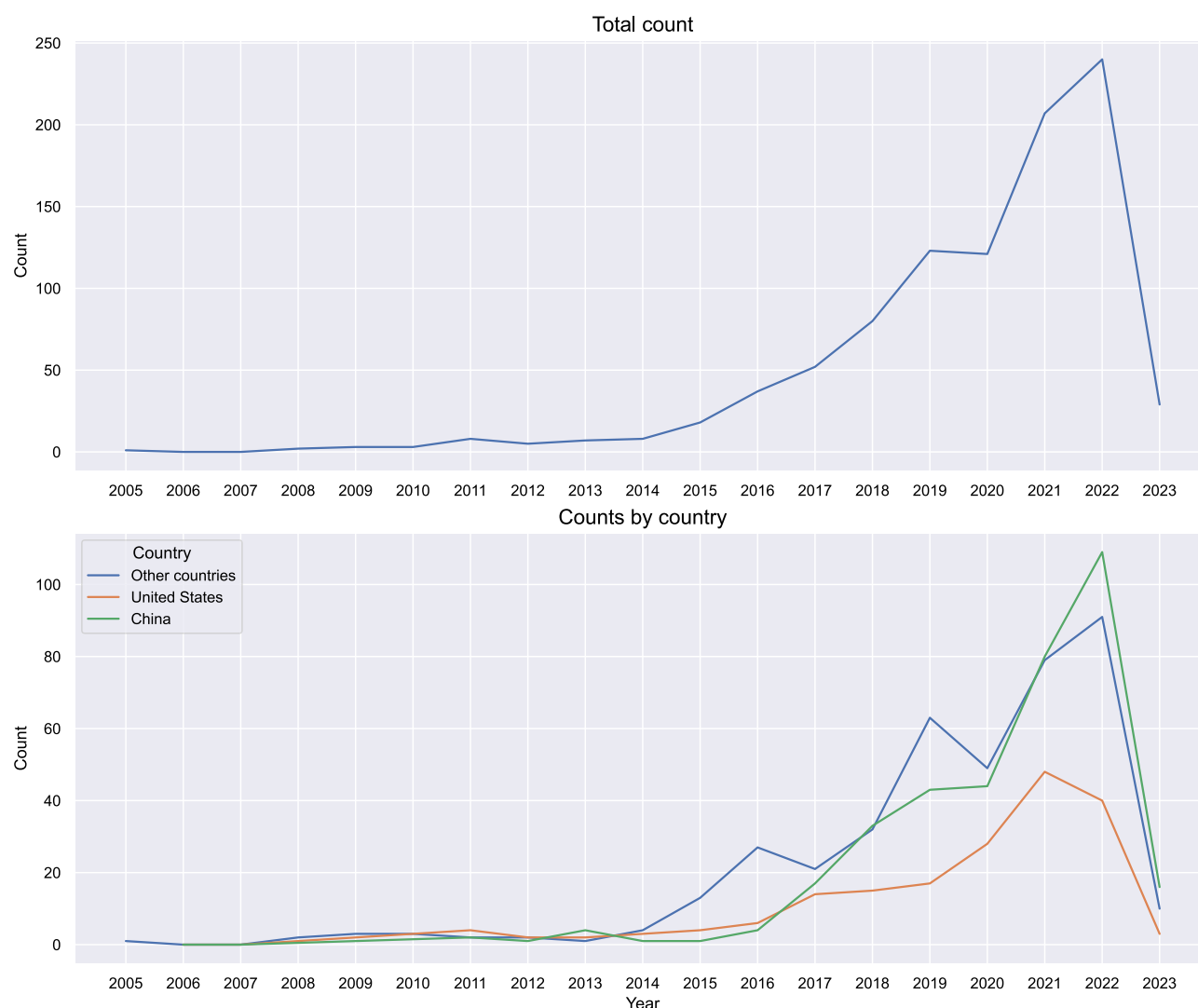


Figure 6 Number of new studies registered at ClinicalTrials.gov each year and by country.

Notes: The 2023 numbers might be low because the search was performed on February 28, 2023, rather than the reduction being a true representation of the data.

sources and study design types have not fundamentally changed. Furthermore, these terms do not define the study design. The recommended strategy is to define the interventional or noninterventional study design, the primary collection or secondary use of data in the study design, the characteristics of the comparison groups, and the assessment of prognostic certainty for the corresponding causal associations.¹⁶ The dichotomy of RCTs and observational studies does not capture the full picture of the study design. Depending on the definition of the term, it is important to note that RWD differs from data explicitly collected for research purposes in a health care setting. The degree of reliance on RWD varies with different types of study designs. Therefore, when an RCT is incorporated into RWD, RWE can also be generated. The review by Baumfeld Andre et al outlined the hybrid trial methodology in detail to encompass studies with RCT characteristics and RWD.¹⁷ The study by Taur et al introduced the observational design of the RWS and enumerated the advantages and disadvantages of common study types.¹⁸ From this perspective, for studies registered at ClinicalTrials.gov, the description of the study design in the study field “Brief Summary” is very confusing. Therefore, special attention needs to be given to the correct and comprehensive description of the study design in the Brief Summary, thereby promoting communication and understanding.

Compared to the strict research registration process of RCTs, there are currently no necessary requirements for the registration of RWSs. When interpreting the results, it must be noted that due to the lagged nature of self-reporting on

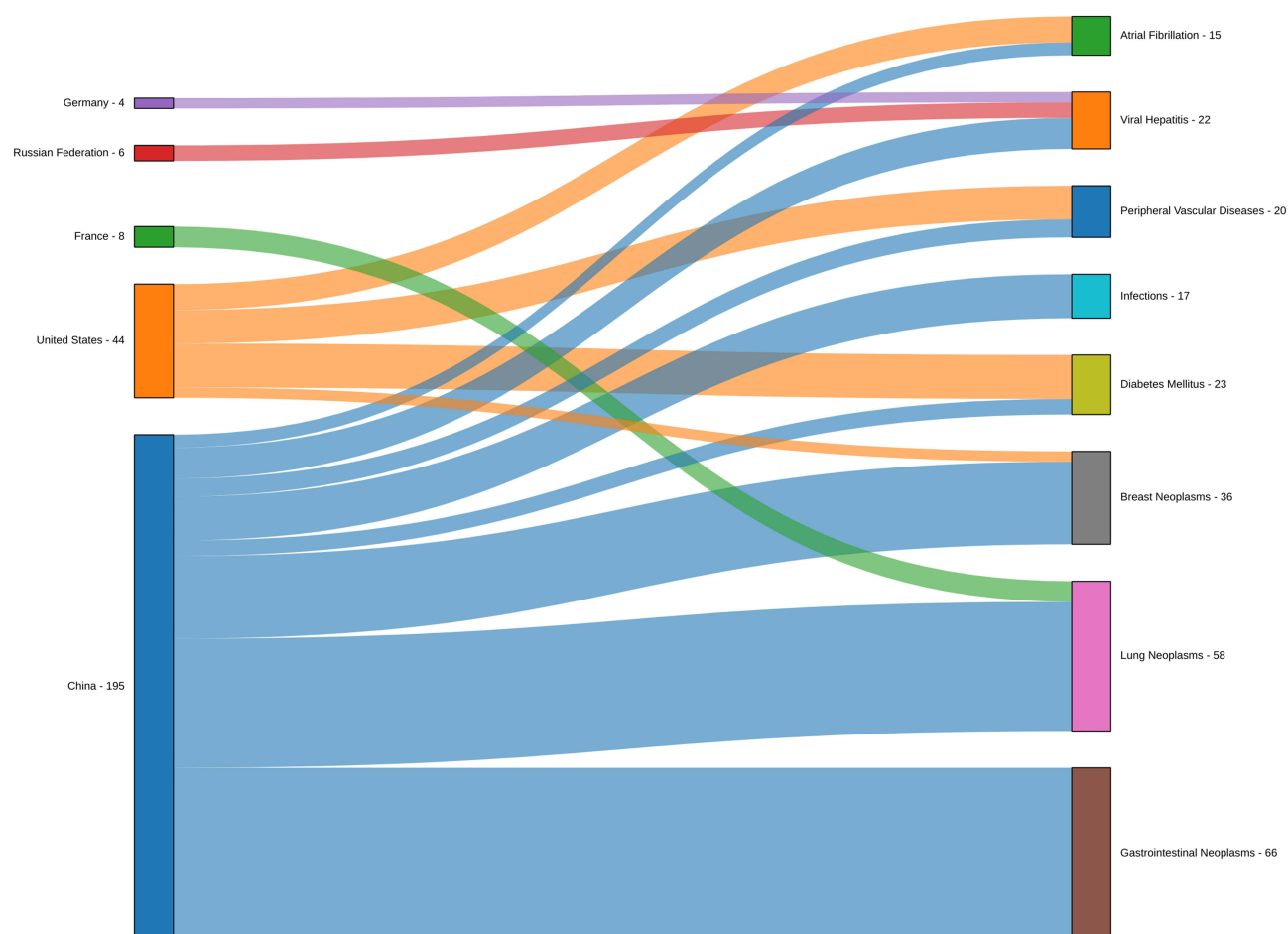


Figure 7 Sankey diagram from country to condition. The numbers after the country and condition names indicate the total number of outflow and inflow studies, respectively. For example, “China - 195” means that a total of 195 studies from China flowed to different condition nodes on the right side of the figure. Note that since the remaining nodes are filtered out, these numbers only represent the statistical association between the nodes displayed in the Sankey diagram.

ClinicalTrials.gov, the characteristics of some studies may be biased from the actual conditions of the groups that registered the studies. Some common characteristics, such as study execution status or enrollment numbers, are neglected when research teams fail to make timely updates to registries. Previous studies have pointed out the shortcomings of registered studies on ClinicalTrials.gov.¹⁹ It has been reported that in RCTs,²⁰ the discrepancy between study registration results and published results is still relatively high due to reasons such as the research team neglecting to update. Similar studies have also raised the need for improvements in the quality of registry data and the reporting of studies for pragmatic trials.²¹ Our research similarly reveals numerous deficiencies in RWS registries.

Regulators and academics are working hard to promote the writing of RWS proposals. The 2013 Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT 2013) provide a good reference for the writing of clinical trial protocols.²² The main applicable target of SPIRIT 2013 is RCTs. RWSs have great variability in design and analysis choices, and the currently used SPIRIT 2013 does not reflect the characteristics of RWSs very well. The National Evaluation System for Health Technology Coordinating Center in the United States released two frameworks related to data quality and research methods in 2020 as part of its work to enable and support the use of RWE to better understand certain medical devices.²³ This methodological framework outlines the structure of a device study protocol and the importance of prespecification to support the rigorous design and execution of RWSs. A novel RWE framework tool was developed and pilot tested by researchers from the pharmaceutical industry to support the interactive visualization of study designs.²⁴ A structured template, Structured template and reporting tool for real world evidence (STaRT-RWE), developed by a public–private consortium serves as a guiding tool for designing and conducting reproducible RWSs,

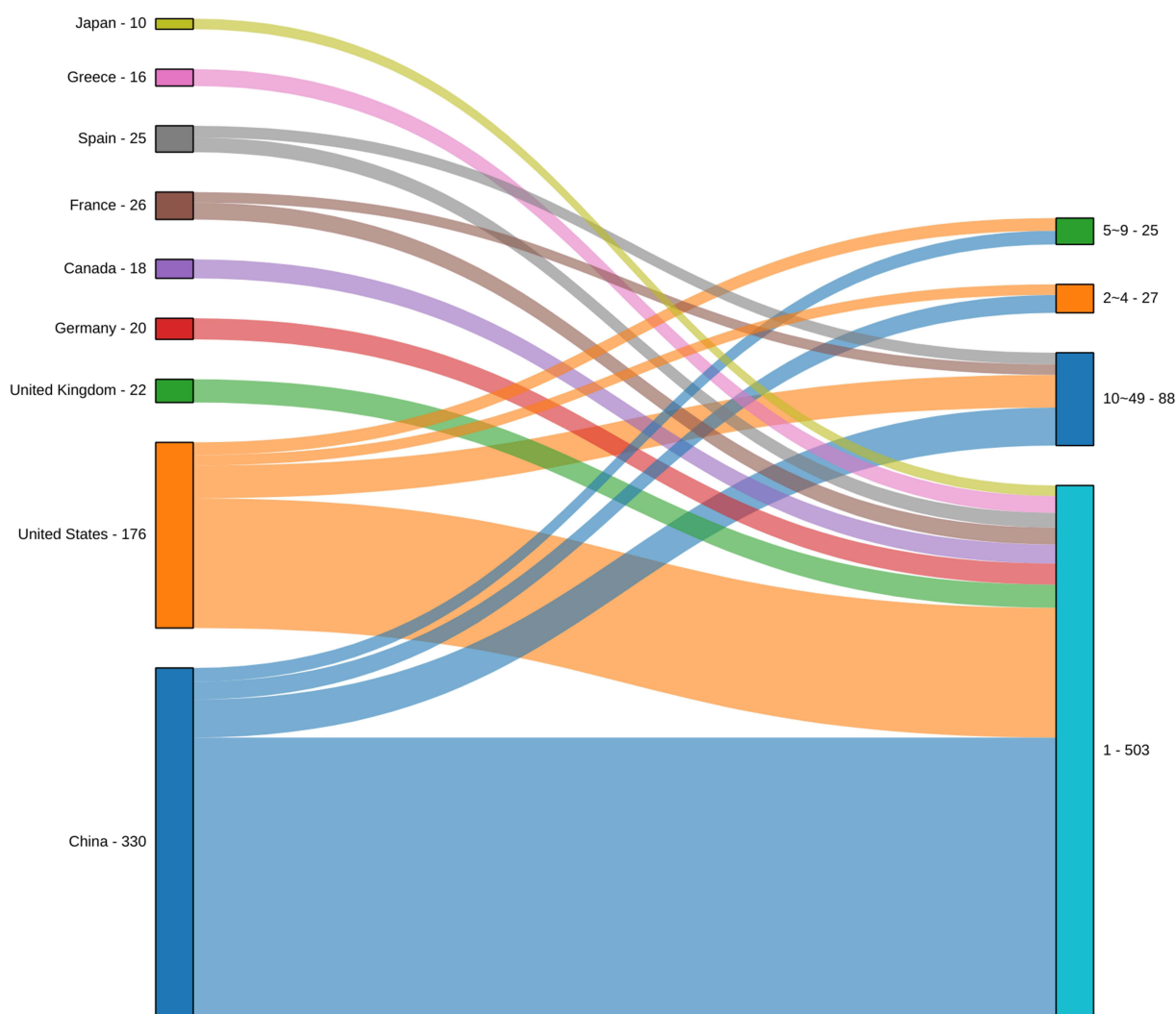


Figure 8 Sankey diagram from country to research center classification. Numerical research center variables are discretized into categorical variables with different intervals. The study centers were classified into five categories: "1", "2-4", "5-9", "10-49", and "≥50". The category "≥50" is not displayed in the plot, as lines with node counts less than 10 are filtered out.

facilitating reproducibility, validity assessment, and evidence synthesis.²⁵ This research provided a STaRT-RWE example library containing 4 common use cases, which is also of great reference value for the writing of research proposals.²⁵ An extended research framework examining long-term safety and efficacy using RWD was recently reported in a study accredited by the International Society for Pharmacoepidemiology.²⁶ Although the COVID-19 pandemic has provided new opportunities for RWSs, the rigor of scientific research still needs to be emphasized.^{27,28} Compared with the retraction of previous clinical studies, the fierce publishing competition during the COVID-19 pandemic has produced a series of side effects.²⁹⁻³¹ The pandemic has also led to a degree of abuse in promoting the widespread use of RWD.³²

The three main components of RWSs (the triad of research question, design, and data) are important when evaluating research validity.³³ Therefore, a visual framework is recommended for demonstrating study designs and data observability.³⁴ In addition, there have been many efforts to improve the transparency and reporting standards of RWSs.³⁵⁻⁴⁰ Regarding the challenge of integrating data from multiple sources, the FDA Sentinel Innovation Center recently presented a roadmap and outlined plans and visions for the future.^{41,42} The Center for Drug Evaluation of the NMPA recently drafted the Guidelines for Design and Protocol Frameworks for Real-World Drug Studies (draft open for comments).⁴³ The guidelines mention that the source of research data should be clarified, including the research center from which they came, the start and end time of data collection, and the system and record format of data storage. If the

data are derived from previous research, the form of recording and storage of the original data should be described to ensure the traceability of the research data.⁴³

Our study has the following limitations. First, the registry studies we included were only from ClinicalTrials.gov. Second, we did not search for published RWSs more broadly to aggregate this information for a more in-depth analysis. Finally, this study is of interest to researchers who are new to RWSs. Therefore, we did not explain the details of the various study designs from a statistical point of view. We focused on providing readers with an analysis of research hotspots in RWSs, as well as reducing misunderstandings and summarizing more comprehensive research design tools and resources.

Conclusion

Although the pandemic has provided new opportunities for RWSs, the rigor of scientific research still needs to be emphasized. Special attention needs to be given to the correct and comprehensive description of the study design in the Brief Summary of registered studies, thereby promoting communication and understanding. In addition, deficiencies in ClinicalTrials.gov registration data remain prominent. Although most of the registered studies describing the 3 most commonly mentioned diseases were from China, the reasons behind this were related to the source of the data and the distribution of the diseases.

Abbreviations

API, application programming interface; COVID-19, coronavirus disease 2019; EHRs, electronic health records; EMRs, electronic medical records; FDA, United States Food and Drug Administration; NMPA, the China National Medical Products Administration; NPCR, National Program of Cancer Registries; RCT, randomized controlled trial; RWD, real-world data; RWE, real-world evidence; RWS, real-world study; SEER, Surveillance, Epidemiology, and End Results; STaRT-RWE, Structured template and reporting tool for real world evidence; URL, Uniform Resource Locator; SPIRIT, Standard Protocol Items: Recommendations for Interventional Trials.

Data Sharing Statement

All data can be obtained using the ClinicalTrials.gov API (<https://clinicaltrials.gov/api/gui/home>). The Python code for scraping data from ClinicalTrials.gov using its API is provided in our online [Supplementary Material](#). This code is publicly available and can be used to make data requests with any search term. Please see the online [Supplementary Materials](#) for a table containing all ClinicalTrials.gov fields.

Funding

This study was supported by the Lanzhou Chengguan District Science and Technology Planning Project (No. 2020SHFZ0041), the 2022 Education and Teaching Reform Research Project of the Second Clinical Medical College of Lanzhou University (No. DELC-202241) and the Cuiying Scientific Training Program for Undergraduates of Lanzhou University Second Hospital (No. CYXZ2022-21).

Disclosure

All authors declared no potential conflicts of interest in this work.

References

1. Sherman RE, Anderson SA, Dal Pan GJ, et al. Real-world evidence - what is it and what can it tell us? *N Engl J Med*. 2016;375(23):2293–2297. doi:10.1056/NEJMsbl609216
2. NMPA. CfDEo. Real-world data used to generate real-world evidences (Interim); 2021. Available from: <http://zhengbiaoke-com.oss-cn-beijing.aliyuncs.com/2104/2104-773f72d9-7fc4-4f02-ae13-ad4aeafc2f98.pdf>. Accessed February 11, 2022.
3. Food and Drug Administration. Real-world evidence. Available from: <https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence>. Accessed July 12, 2022.
4. Wang B, Lai J, Yan X, Jin F, Yao C. Implications of the lack of a unified research project framework: an investigation into the registration of clinical trials of COVID-19. *Curr Med Res Opin*. 2020;36(7):1131–1135. doi:10.1080/03007995.2020.1771294
5. Wang B, Lai J, Yan X, et al. COVID-19 clinical trials registered worldwide for drug intervention: an overview and characteristic analysis. *Drug Des Devel Ther*. 2020;14:5097–5108. doi:10.2147/DDDT.S281700

6. Dron L, Kalatharan V, Gupta A, et al. Data capture and sharing in the COVID-19 pandemic: a cause for concern. *Lancet Digit Health*. 2022;4(10):e748–e756. doi:10.1016/S2589-7500(22)00147-9
7. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1033–1039. doi:10.1002/pds.4297
8. Food and Drug Administration (FDA). Use of electronic health record data in clinical investigations guidance for industry; 2018. Available from: <https://www.fda.gov/regulatory-information/search-fda-guidance-documents/use-electronic-health-record-data-clinical-investigations-guidance-industry>. Accessed August 24, 2021.
9. Gatto NM, Campbell UB, Rubinstein E, et al. The structured process to identify fit-for-purpose data: a data feasibility assessment framework. *Clin Pharmacol Ther*. 2022;111(1):122–134. doi:10.1002/cpt.2466
10. Gatto NM, Reynolds RF, Campbell UB. A structured preapproval and postapproval comparative study design framework to generate valid and transparent real-world evidence for regulatory decisions. *Clin Pharmacol Ther*. 2019;106(1):103–115. doi:10.1002/cpt.1480
11. Maruszczuk K, Aiyegbusi OL, Cardoso VR, et al. Implementation of patient-reported outcome measures in real-world evidence studies: analysis of ClinicalTrials.gov records (1999–2021). *Contemp Clin Trials*. 2022;120:106882. doi:10.1016/j.cct.2022.106882
12. Schneeweiss S, Rassen JA, Brown JS, et al. Graphical depiction of longitudinal study designs in health care databases. *Ann Intern Med*. 2019;170(6):398–406. doi:10.7326/M18-3079
13. Penberthy LT, Rivera DR, Lund JL, Bruno MA, Meyer AM. An overview of real-world data sources for oncology and considerations for research. *CA Cancer J Clin*. 2022;72(3):287–300. doi:10.3322/caac.21714
14. Wang Z, Zhou C, Feng X, Mo M, Shen J, Zheng Y. Comparison of cancer incidence and mortality between China and the United States. *Precis Cancer Med*. 2021;4:1.
15. Hayat MJ, Howlader N, Reichman ME, Edwards BK. Cancer statistics, trends, and multiple primary cancer analyses from the Surveillance, Epidemiology, and End Results (SEER) program. *Oncologist*. 2007;12(1):20–37. doi:10.1634/theoncologist.12-1-20
16. Concato J, Corrigán-Curay J. Real-world evidence - where are we now? *N Engl J Med*. 2022;386(18):1680–1682. doi:10.1056/NEJMp2200089
17. Baumfeld Andre E, Reynolds R, Caubel P, Azoulay L, Dreyer NA. Trial designs using real-world data: the changing landscape of the regulatory approval process. *Pharmacoepidemiol Drug Saf*. 2020;29(10):1201–1212. doi:10.1002/pds.4932
18. Taur SR. Observational designs for real-world evidence studies. *Perspect Clin Res*. 2022;13(1):12–16. doi:10.4103/picr.picr_217_21
19. Zarin DA, Tse T, Williams RJ, Califf RM, Ide NC. The ClinicalTrials.gov results database—update and key issues. *N Engl J Med*. 2011;364(9):852–860. doi:10.1056/NEJMs1012065
20. Pranic S, Marusic A. Changes to registration elements and results in a cohort of ClinicalTrials.gov trials were not reflected in published articles. *J Clin Epidemiol*. 2016;70:26–37. doi:10.1016/j.jclinepi.2015.07.007
21. Nicholls SG, Carroll K, Hey SP, et al. A review of pragmatic trials found a high degree of diversity in design and scope, deficiencies in reporting and trial registry data, and poor indexing. *J Clin Epidemiol*. 2021;137:45–57. doi:10.1016/j.jclinepi.2021.03.021
22. Chan AW, Tetzlaff JM, Altman DG, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158(3):200–207. doi:10.7326/0003-4819-158-3-201302050-00583
23. National Evaluation System for health Technology Coordinating Center for Drug Evaluation N. NESTcc research methods and data quality frameworks; 2020. Available from: <https://nestcc.org/data-quality-and-methods/>. Accessed July 14, 2022.
24. Xia AD, Schaefer CP, Szende A, Jahn E, Hirst MJ. RWE Framework: An interactive visual tool to support a real-world evidence study design. *Drugs Real World Outcomes*. 2019;6(4):193–203. doi:10.1007/s40801-019-00167-6
25. Wang SV, Pinheiro S, Hua W, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. *BMJ*. 2021;372:m4856. doi:10.1136/bmj.m4856
26. Burcu M, Manzano-Salgado CB, Butler AM, Christian JB. A framework for extension studies using real-world data to examine long-term safety and effectiveness. *Ther Innov Regul Sci*. 2022;56(1):15–22. doi:10.1007/s43441-021-00322-8
27. Kim AHJ, Sparks JA, Liew JW, et al. A rush to judgment? Rapid reporting and dissemination of results and its consequences regarding the use of hydroxychloroquine for COVID-19. *Ann Intern Med*. 2020;172(12):819–821. doi:10.7326/M20-1223
28. Pillar C, Travis J. Authors, elite journals under fire after major retractions. *Science*. 2020;368(6496):1167–1168. doi:10.1126/science.368.6496.1167
29. Wang B, Lai J, Yan X, Jin F, Yao C. Exploring the characteristics, global distribution and reasons for retraction of published articles involving human research participants: a literature survey. *Eur J Intern Med*. 2020;78:145–146. doi:10.1016/j.ejim.2020.03.016
30. Shah K, Charan J, Sinha A, Saxena D. Retraction rates of research articles addressing COVID-19 pandemic: is it the evolving COVID epidemiology or scientific misconduct? *Indian J Community Med*. 2021;46(2):352–354. doi:10.4103/ijcm.IJCM_732_20
31. Soltani P, Patini R. Retracted COVID-19 articles: a side-effect of the hot race to publication. *Scientometrics*. 2020;125(1):819–822. doi:10.1007/s11192-020-03661-9
32. Dolgin E. Core concept: the pandemic is prompting widespread use-and misuse-of real-world data. *Proc Natl Acad Sci U S A*. 2020;117(45):27754–27758. doi:10.1073/pnas.2020930117
33. Wang SV, Schneeweiss S. Assessing and interpreting real-world evidence studies: introductory points for new reviewers. *Clin Pharmacol Ther*. 2022;111(1):145–149. doi:10.1002/cpt.2398
34. Wang SV, Schneeweiss S. A framework for visualizing study designs and data observability in electronic health record data. *Clin Epidemiol*. 2022;14:601–608. doi:10.2147/CLEP.S358583
35. Orsini LS, Berger M, Crown W, et al. Improving transparency to build trust in real-world secondary data studies for hypothesis testing—why, what, and how: recommendations and a road map from the real-world evidence transparency initiative. *Value Health*. 2020;23(9):1128–1136. doi:10.1016/j.jval.2020.04.002
36. Langan SM, Schmidt SA, Wing K, et al. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*. 2018;363:k3532. doi:10.1136/bmj.k3532
37. Wang SV, Schneeweiss S, Berger ML, et al. Reporting to improve reproducibility and facilitate validity assessment for healthcare database studies V1.0. *Pharmacoepidemiol Drug Saf*. 2017;26(9):1018–1032. doi:10.1002/pds.4295

38. Berger ML, Sox H, Willke RJ, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: recommendations from the joint ISPOR-ISPE special task force on real-world evidence in health care decision making. *Value Health*. 2017;20(8):1003–1008. doi:10.1016/j.jval.2017.08.3019
39. Wang SV, Verpillat P, Rassen JA, Patrick A, Garry EM, Bartels DB. Transparency and reproducibility of observational cohort studies using large healthcare databases. *Clin Pharmacol Ther*. 2016;99(3):325–332. doi:10.1002/cpt.329
40. Benchimol EI, Moher D, Ehrenstein V, Langan SM. Retraction of COVID-19 pharmacoepidemiology research could have been avoided by effective use of reporting guidelines. *Clin Epidemiol*. 2020;12:1403–1420. doi:10.2147/CLEP.S288677
41. Desai RJ, Matheny ME, Johnson K, et al. Broadening the reach of the FDA sentinel system: a roadmap for integrating electronic health record data in a causal analysis framework. *NPJ Digit Med*. 2021;4(1):170. doi:10.1038/s41746-021-00542-0
42. Center. FSI. Innovation Center (IC) master plan; 2021. Available from: <https://www.sentinelinitiative.org/news-events/publications-presentations/innovation-center-ic-master-plan>. Accessed July 15, 2022.
43. NMPA. CfDEo. Notice on public comments on the “guidelines for the design and protocol framework of drug real-world study (draft for comment)”; 2022. Available from: <https://www.cde.org.cn/main/news/viewInfoCommon/ea778658adc3d1ae3ffe3f1cc0522e5e>. Accessed July 15, 2022.

International Journal of General Medicine

Dovepress

Publish your work in this journal

The International Journal of General Medicine is an international, peer-reviewed open-access journal that focuses on general and internal medicine, pathogenesis, epidemiology, diagnosis, monitoring and treatment protocols. The journal is characterized by the rapid reporting of reviews, original research and clinical studies across all disease areas. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/international-journal-of-general-medicine-journal>