An Overview of Real-World Data Sources for Real-World Studies in Oncology and Rare Diseases in China

RWD162

Yaqi Jia[1], Siyan Zhan[2,3], Jeff Jianfei Guo[4], Xian Cao[1], Yan-Jun Zhang[1]

1. Medical Affairs, Takeda (China) International Trading Company, Beijing, China. *2. Department of Epidemiology and Biostatistics, School of Public Health, Peking University, Beijing, China*

3. Research Center of Clinical Epidemiology, Peking University Third Hospital, Beijing, China. *4. James L. Winkle College of Pharmacy, University of Cincinnati Academic Health Center, Cincinnati, OH, United States.*

OBJECTIVES

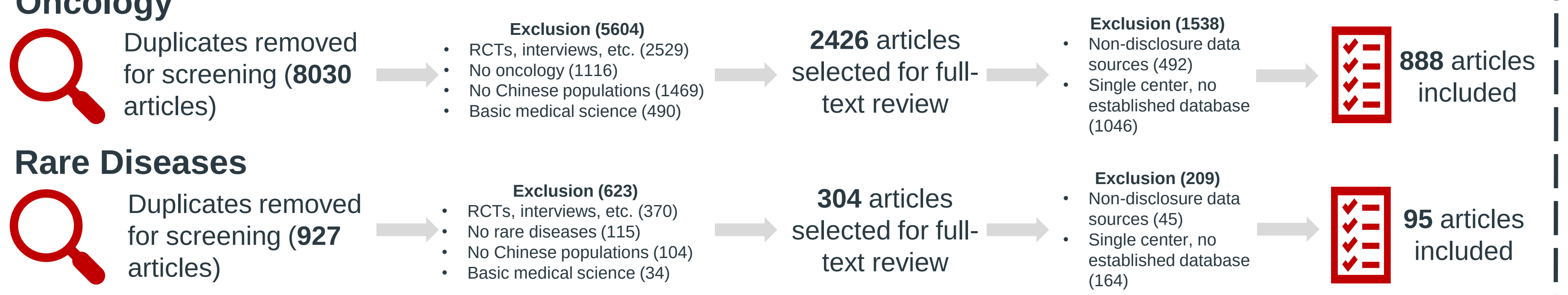
As legislation and incentives to advance the sharing of health data in China increase over time, we aimed to investigate the trends and potentials of real-world data (RWD) sources in oncology and rare diseases in China.

METHODS

- **Publications:** Real-world studies (RWS) covering Chinese population**, f**rom **PubMed** and **Embase** in English, **China National Knowledge Infrastructure (CNKI)** and **Wanfang** in Chinese, with restrictions on publication dates from 1st January 2010 to 31st August 2022
- **Keywords development:** Diseases and therapeutics, study designs, data sources, publication dates, countries of study populations, and exclusion criteria related keywords
- **Data extraction process:** Duplicates removal → two rounds of article screening (i.e., title/abstract and full-text screening) → final data extraction (i.e., study identifiers, study features, and data source features)

Inclusion criteria	Exclusion criteria
• RWS in oncology or rare diseases	• Randomized controlled trials (RCTs)
• Covered Chinese populations from China	• Interviews, reviews, case reports/series, commentaries, expert consensus, editorials, education programs, or studies in basic medical science
• Used secondary databases as data sources	• Used non-disclosure data sources
	• Single center studies without using established databases as data sources

RESULTS



Oncology

Duplicates removed for screening (**8030** articles) → **Exclusion (5604)**
- RCTs, interviews, etc. (2529)
- No oncology (1116)
- No Chinese populations (1469)
- Basic medical science (490)

→ **2426** articles selected for full-text review → **Exclusion (1538)**
- Non-disclosure data sources (492)
- Single center, no established database (1046)

→ **888** articles included

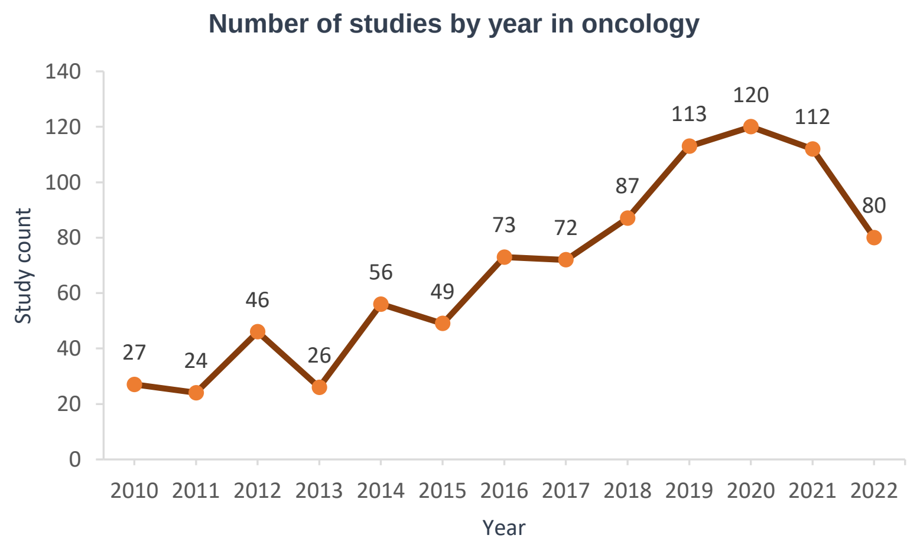
Rare Diseases

Duplicates removed for screening (**927** articles) → **Exclusion (623)**
- RCTs, interviews, etc. (370)
- No rare diseases (115)
- No Chinese populations (104)
- Basic medical science (34)

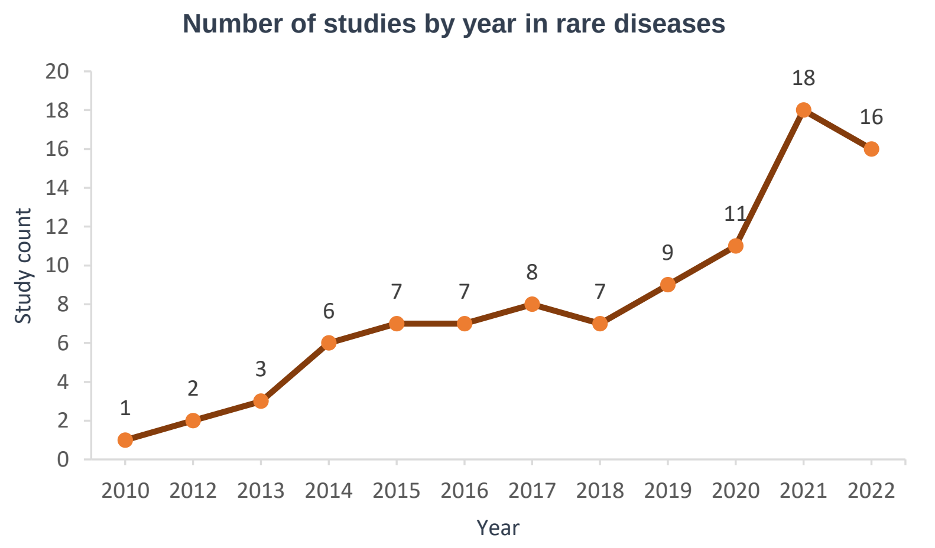
→ **304** articles selected for full-text review → **Exclusion (209)**
- Non-disclosure data sources (45)
- Single center, no established database (164)

→ **95** articles included

1 **Finding 1:** There has been an increasing trend in the number of studies in both therapeutic areas (TAs) since 2010. Particularly, the increasing rate became faster since/after 2017.

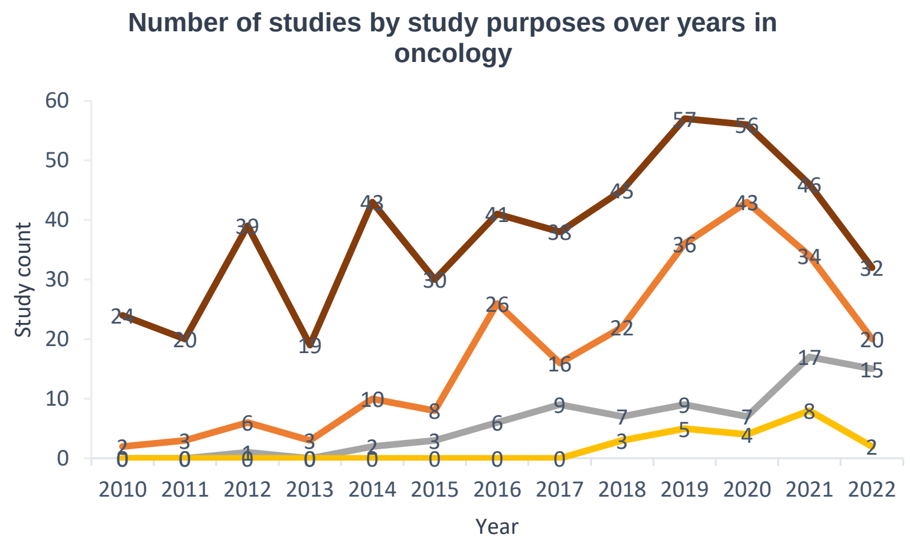


Number of studies by year in oncology

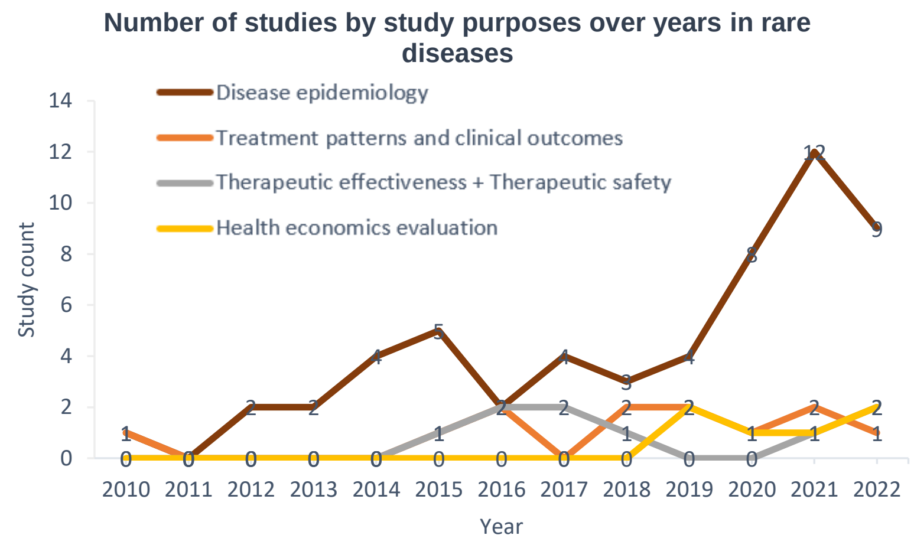


Number of studies by year in rare diseases

2 **Finding 2:** Study purposes gradually became diversified since 2010 in both TAs. The number of studies for treatment patterns and clinical outcomes, effectiveness and safety evaluations increased over years. Health economics evaluations emerged since 2018/19.



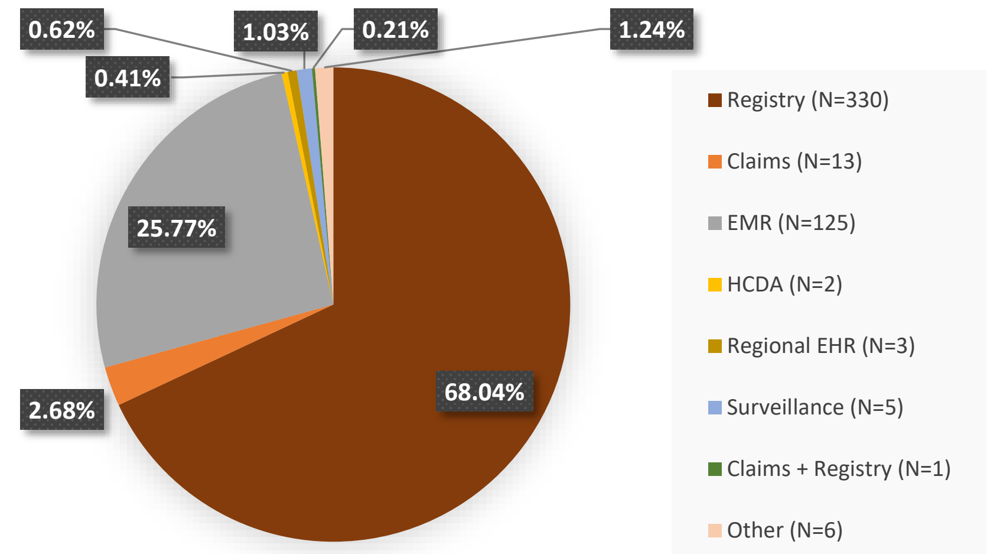
Number of studies by study purposes over years in oncology



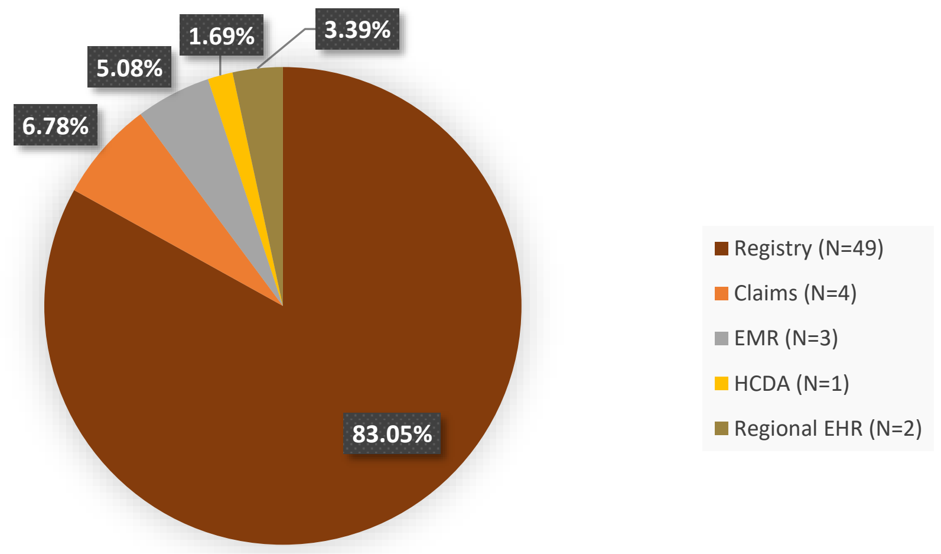
Number of studies by study purposes over years in rare diseases

Studies with "other" purposes (N=103 in oncology and N=19 in rare diseases) were not shown in figures. "Other" purposes were those about diagnostic approach optimization, database establishment, database quality control, etc.

3 **Finding 3:** Databases were mostly registries in both TAs. Regional EHRs and HCDAs were also identified. Data source types were more diversified in oncology than in rare diseases.



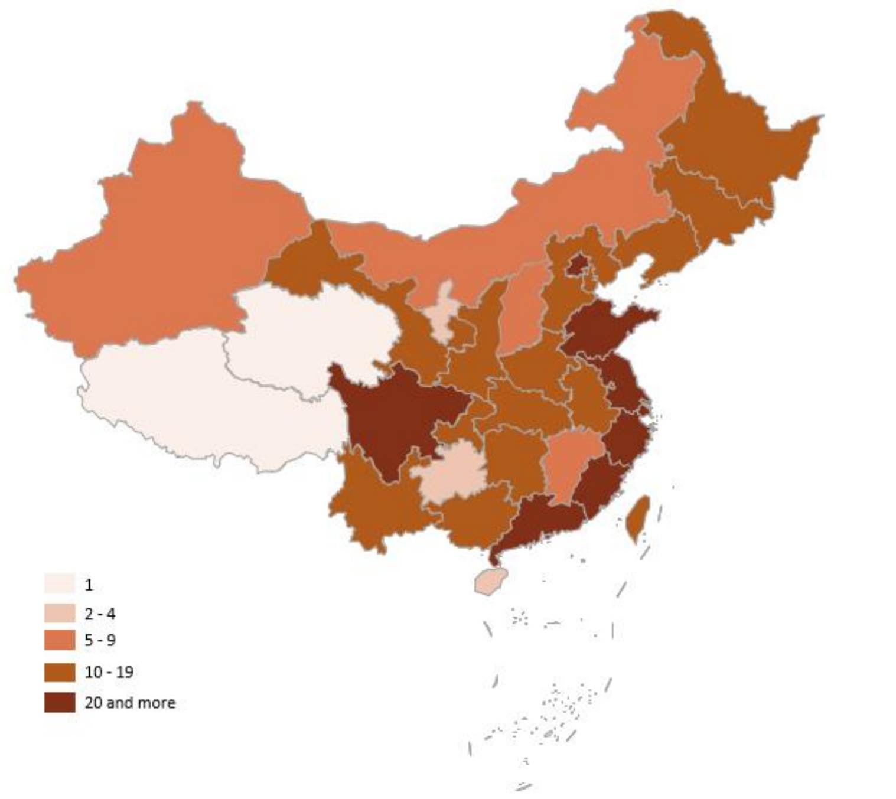
Number of databases by data source types in oncology



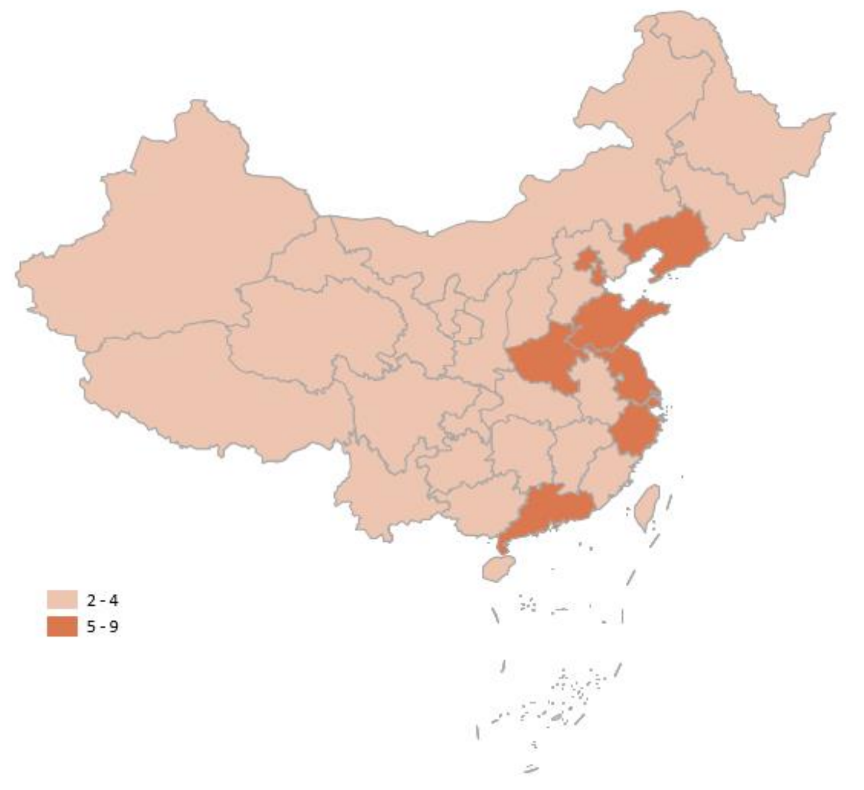
Number of databases by data source types in rare diseases

EMR: Electronic medical record – the non-closed-loop medical records of clinical diagnoses, treatments, and medical services for outpatients and inpatients; Regional EHR: Regional electronic health record – the integrated closed-loop data of the multi-source health records in the region; HCDA: Health care data aggregator – also called health technology data company, which performed data aggregation using a specified model for either nonprofit purposes (e.g., quality improvement), or commercial purposes (e.g., drug development), or both.

4 **Finding 4:** Databases of oncology covered more regions than those of rare diseases. Provinces in eastern and central China (particularly coastal provinces with advanced economy) seemed more popular to be covered in a database's network in both TAs.

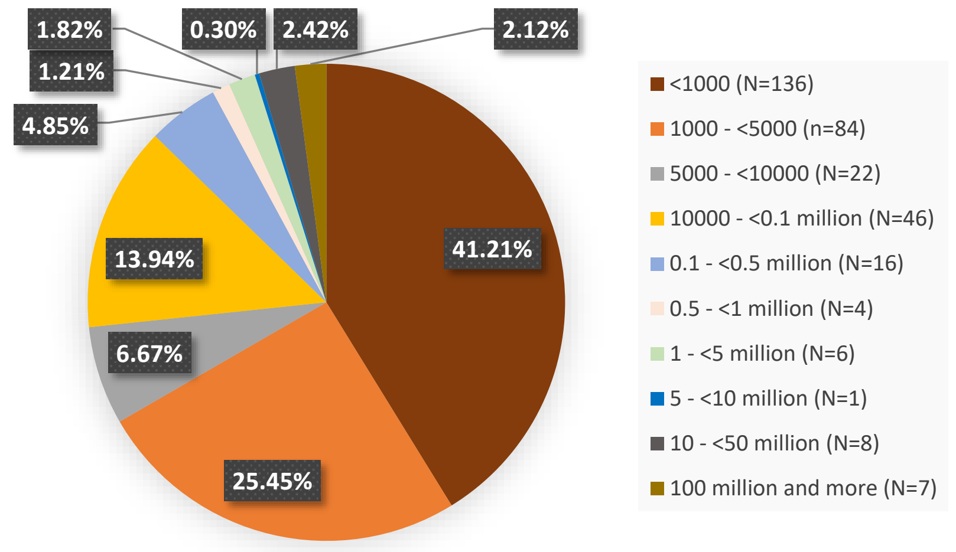


Number of databases by geographical distribution in oncology



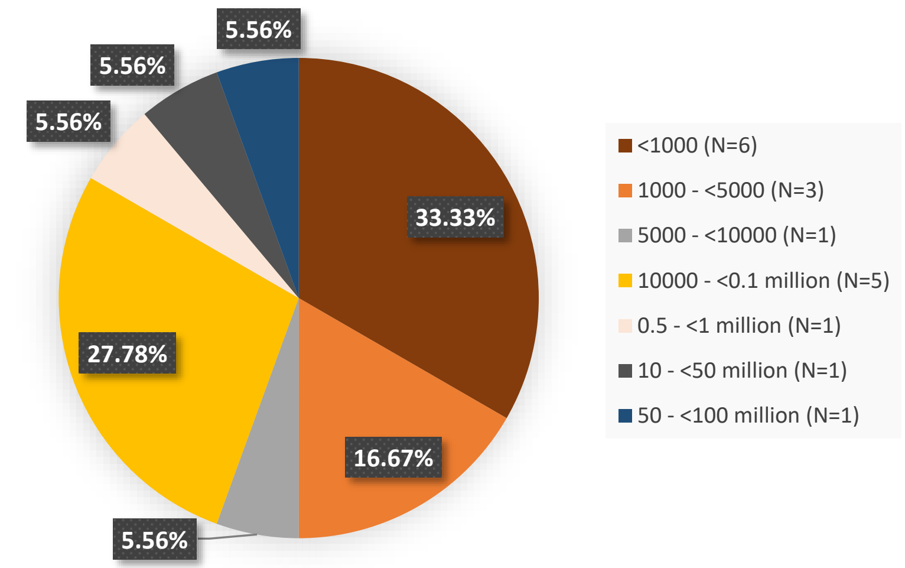
Number of databases by geographical distribution in rare diseases

5 **Finding 5:** In available records of both TAs, databases were mainly multicenter; and databases covering ≥100 hospitals were identified. In oncology, three databases even involved ≥500 hospitals. Over half of the databases in both TAs contained ≥1000 participants. The maximum sample size ranges were ≥100 million in oncology and 50–<100 million in rare diseases.



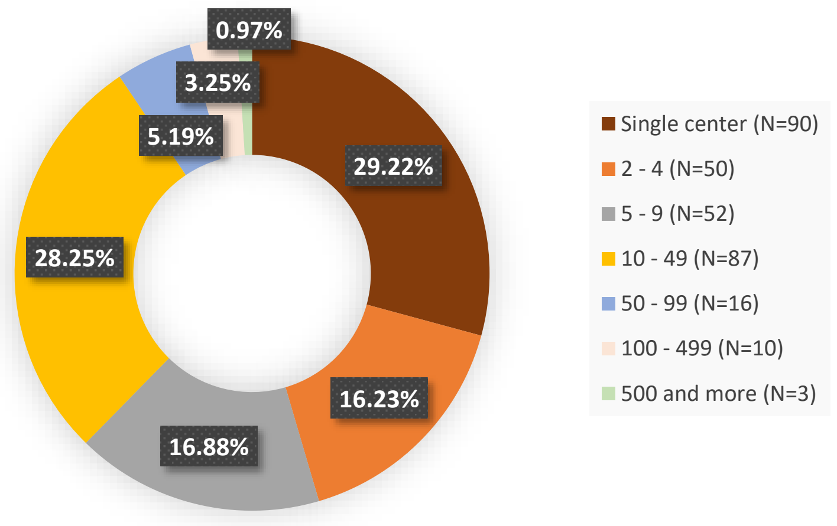
Number of databases by sample size in oncology

Sample sizes of 155 databases remained unknown.



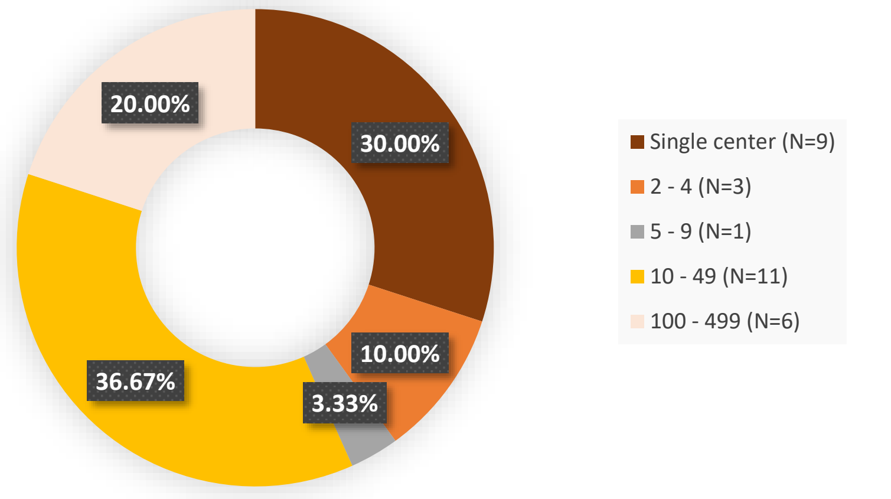
Number of databases by sample size in rare diseases

Sample sizes of 41 databases remained unknown.



Number of databases by hospital number in oncology

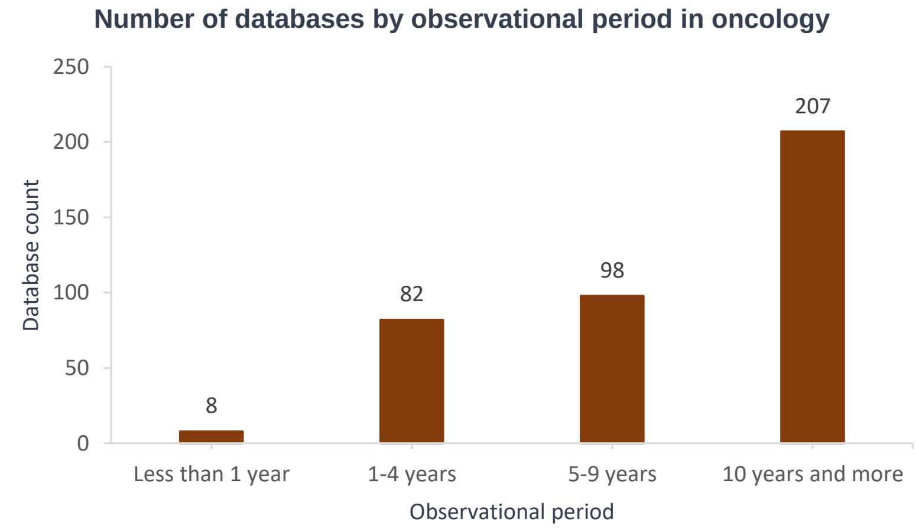
The number of hospitals engaged in 177 databases remained unknown.



Number of databases by hospital number in rare diseases

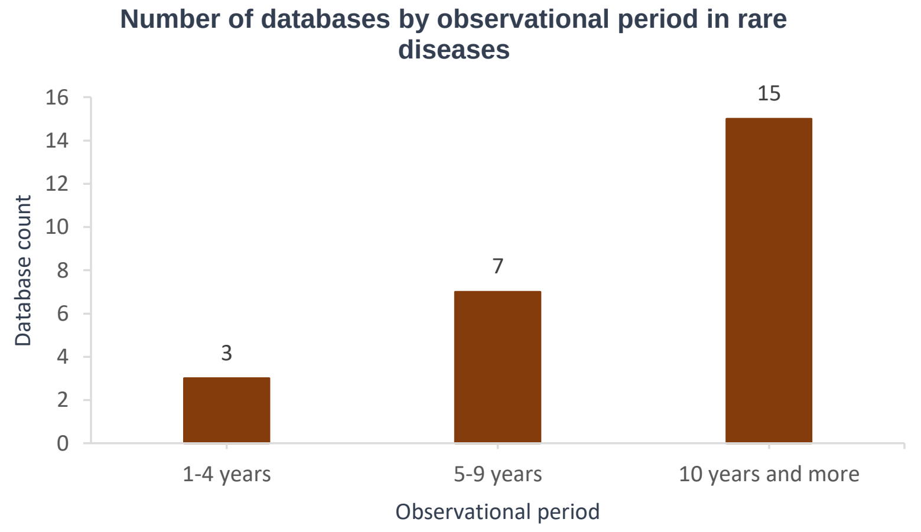
The number of hospitals engaged in 29 databases remained unknown.

6 **Finding 6:** In available records of both TAs, over half of the databases had ≥10 years of follow-up. 2% of databases in oncology were followed within one year.



Number of databases by observational period in oncology

Observational period of 90 databases remained unknown.



Number of databases by observational period in rare diseases

Observational period of 34 databases remained unknown.

CONCLUSIONS

- The number of RWS has a **rapid increase**, and the landscape of RWD sources is **broad and expanding** and includes **diversified** data source types over the past two decades in China.
- The importance of collecting real-world evidence in a **broad population** has been **well-noted**.
- Database features **differed distinctly** across databases; and the development of regional EHRs, claims databases and HCDAs are still at the **preliminary stage**.
- Data application requires consideration in evolving **academic guidance** and **quality assurance** of RWD.
- **Consistent regulatory procedures** for database access and **partnership across healthcare eco-system** are essential for promoting database aggregation and unlocking values of real-world evidence.

ISPOR 2023 Conference | Boston, United States

Acknowledgements: Editorial assistance was provided by Wentian Lu and Yuanming Mao of IQVIA Greater China and this study was funded by Takeda (China) International Trading Company.
Conflicts of interest: Yaqi Jia, Xian Cao, and Yan-Jun Zhang are employees of Takeda (China) International Trading Company. Siyan Zhan and Jeff Jianfei Guo have no conflicts of interest related to this study to disclose.

Please contact us for more information: yaqi.jia@takeda.com