Advancing the Development of Real-world Data for Cancer Care in China: Challenges and Opportunities

Wentian Lu,[1] Danyang Gui,[1] Shuo Yang,[1] Jing Li,[1] Long Pang,[1] Can Chen,[1] Yulan Guo,[2] Cheng Tong,[2] Yang Xie,[1] Zheng Yin[1]

1. IQVIA Greater China, Real World Solutions, Shanghai, China

2. Beijing Yiyong Technology Co., Ltd., Beijing, China

IQVIA
Creating a healthier world

OBJECTIVES

The utilization of secondary database in China has significantly advanced oncology research. We aimed to investigate the trends, current utilization and potentials of databases in oncology in China.
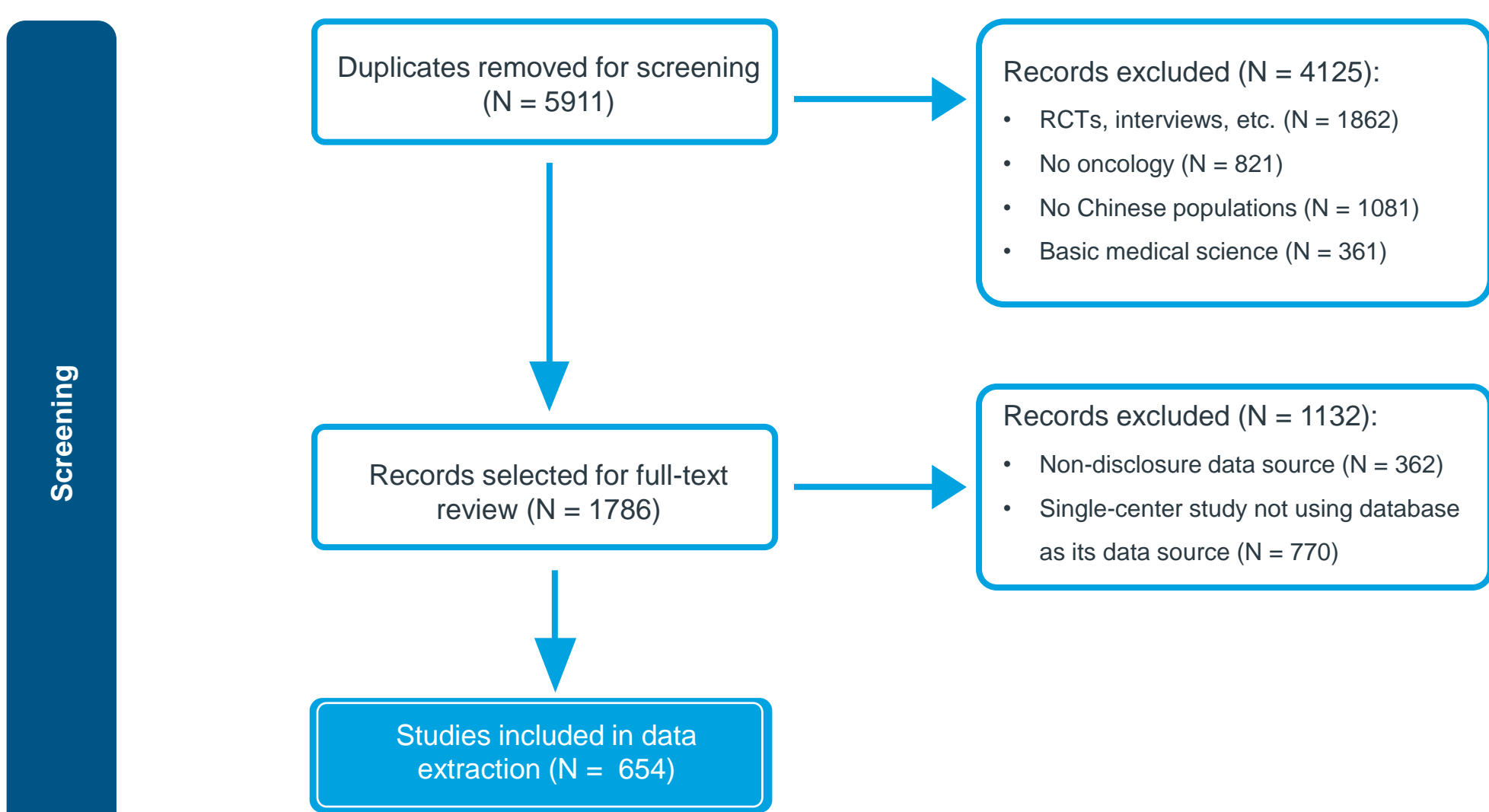
METHODS

- **Publications:** Real-world studies (RWS) on oncology covering Chinese population, from **PubMed** and **Embase** in English, **China National Knowledge Infrastructure (CNKI)** and **Wanfang** in Chinese, with restrictions on publication dates from **1st January 2020** to **31st December 2024**

- **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)

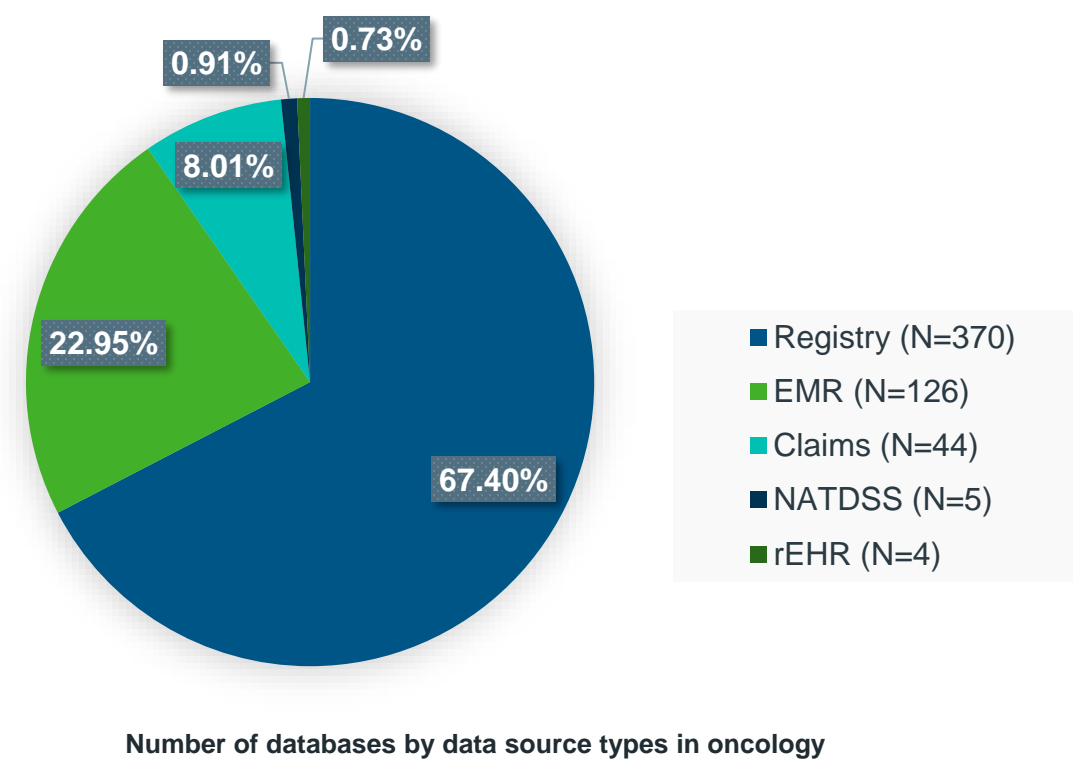| Inclusion criteria	Exclusion criteria
✓ RWS in oncology	✗ Randomized controlled trials (RCTs)
✓ Covered Chinese populations from China	✗ Interviews, reviews, case reports/series, commentaries, expert consensus, editorials, education programs, or studies in basic medical science
✓ Used secondary databases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

- **Qualitative interview**: Database coverage, data collection, variable capture, and research partnership

RESULTS

- A total of **5911** studies were retrieved from PubMed, Embase, CNKI, and Wanfang after deduplication. After two rounds of screening, **654** records were eligible for data extraction.

Screening

Duplicates removed for screening (N = 5911) → Records excluded (N = 4125):
- RCTs, interviews, etc. (N = 1862)
- No oncology (N = 821)
- No Chinese populations (N = 1081)
- Basic medical science (N = 361)

Records selected for full-text review (N = 1786) → Records excluded (N = 1132):
- Non-disclosure data source (N = 362)
- Single-center study not using database as its data source (N = 770)

Studies included in data extraction (N = 654)

Summary statistics 1: Data source types were diversified, containing 67.4% (n=370) used registries, 23.0% (n=126) used EMR, 8.0% (n=44) used claims, 0.9% (n=5) NATDSS and 0.7% (n=4) used rEHR.


Number of databases by data source types in oncology

- 0.73%
- 0.91%
- 8.01%
- 22.95%
- 67.40%

Legend:
- Registry (N=370)
- EMR (N=126)
- Claims (N=44)
- NATDSS (N=5)
- rEHR (N=4)

EMR: Electronic medical record – the **non-closed-loop** medical records of clinical diagnoses, treatments, and medical services for outpatients and inpatients

rEHR: Regional electronic health record – the **integrated closed-loop** data of the multi-source health records in the region

NATDSS: National Anti-Tumor Drugs Surveillance System – the **only national** EMR-based oncology database in mainland China, which was established in 2018 by the National Cancer Center (NCC).
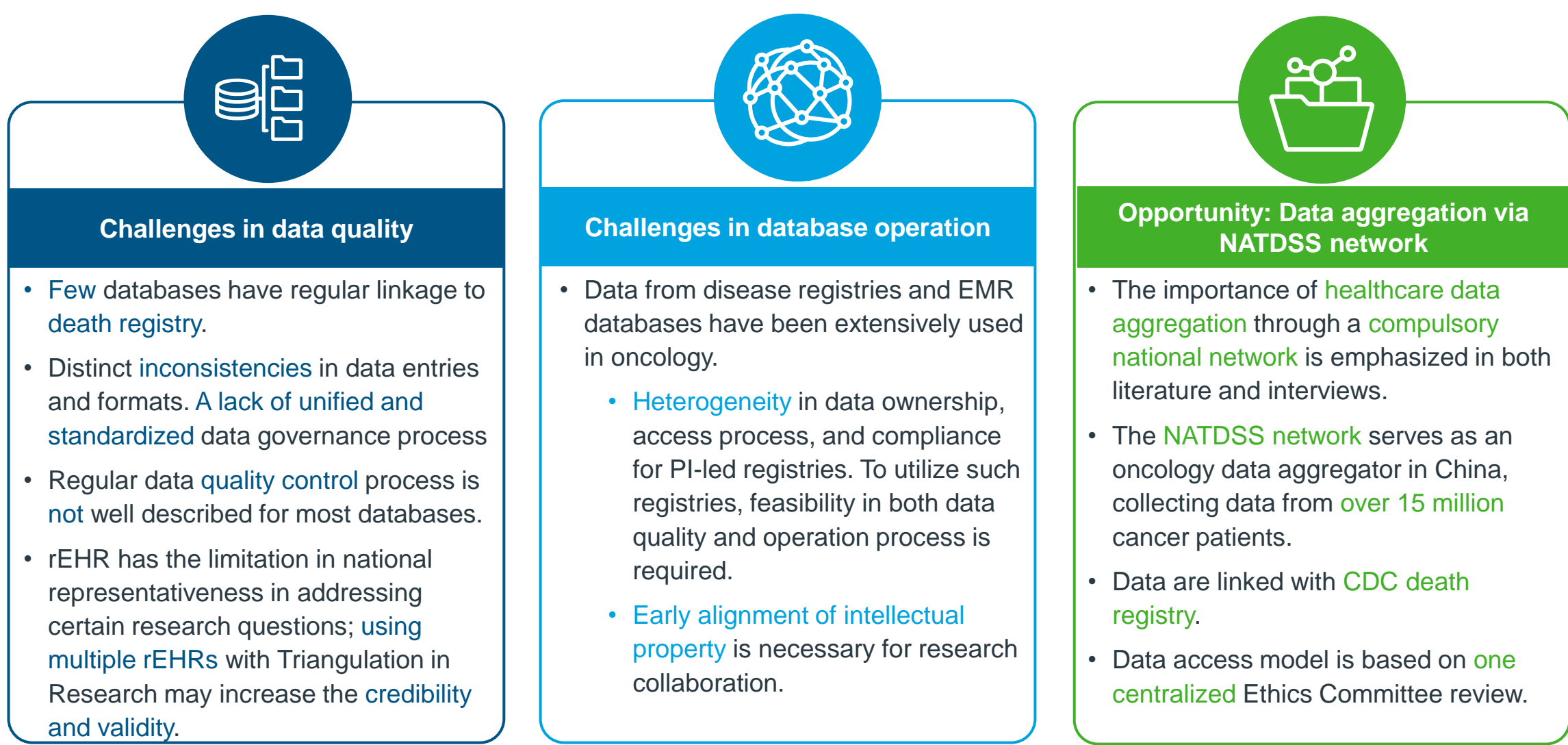
Summary statistics 2: Purposes of database studies on oncology in China were also diversified. Databases have been mainly used to describe disease epidemiology and natural history (43.4%; 284/654), clinical effectiveness and drug safety (12.3%; 80/654), treatment pattern and outcomes (31.1%; 204/654) and health economics (4.4%; 29/654).
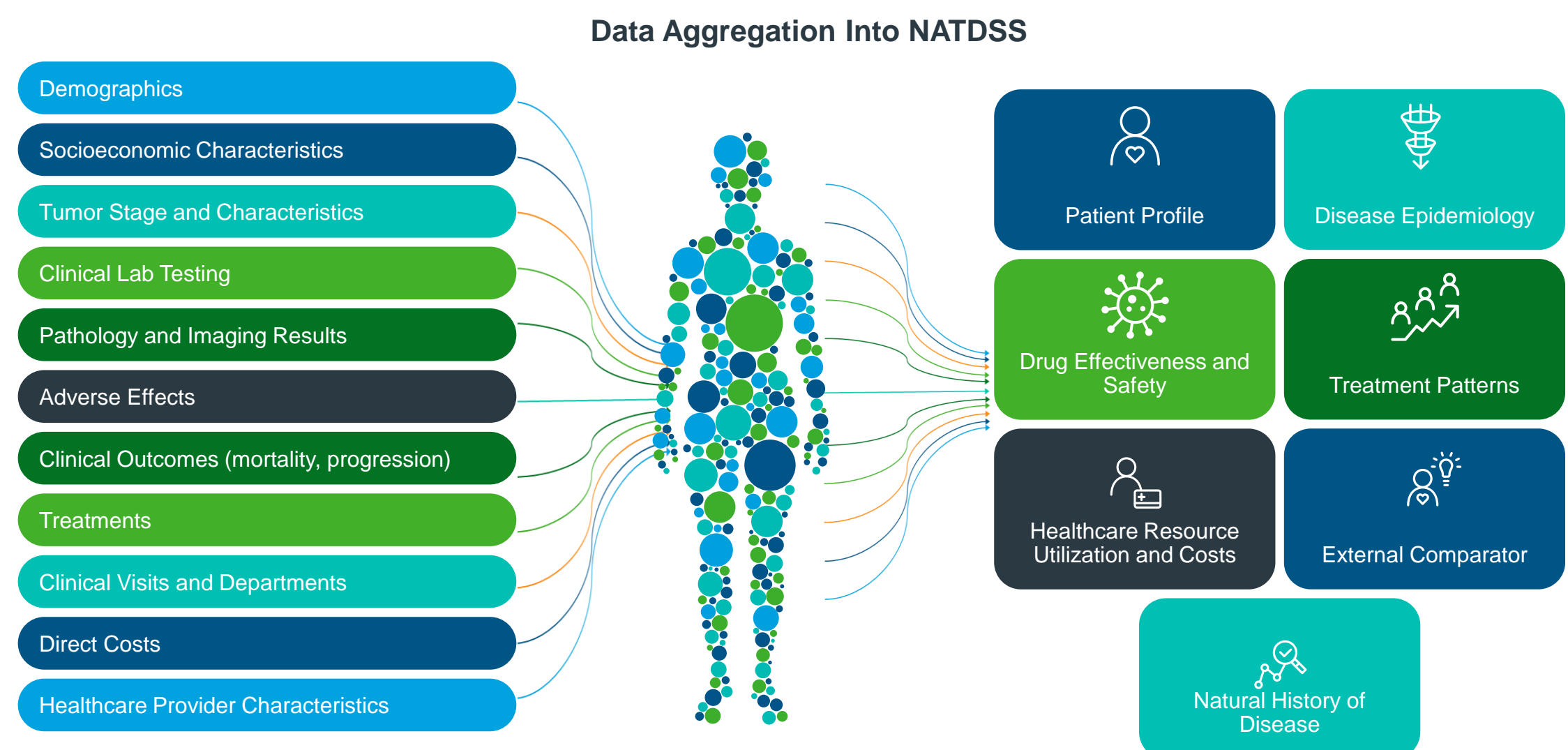
Distribution of databases by study purposes in oncology

- HEALTH ECONOMICS (N=29): 4.40%
- CLINICAL EFFECTIVENESS AND DRUG SAFETY (N=80): 12.30%
- TREATMENT PATTERN AND OUTCOMES (N=204): 31.10%
- DISEASE EPIDEMIOLOGY AND NATURAL HISTORY (N=284): 43.40%
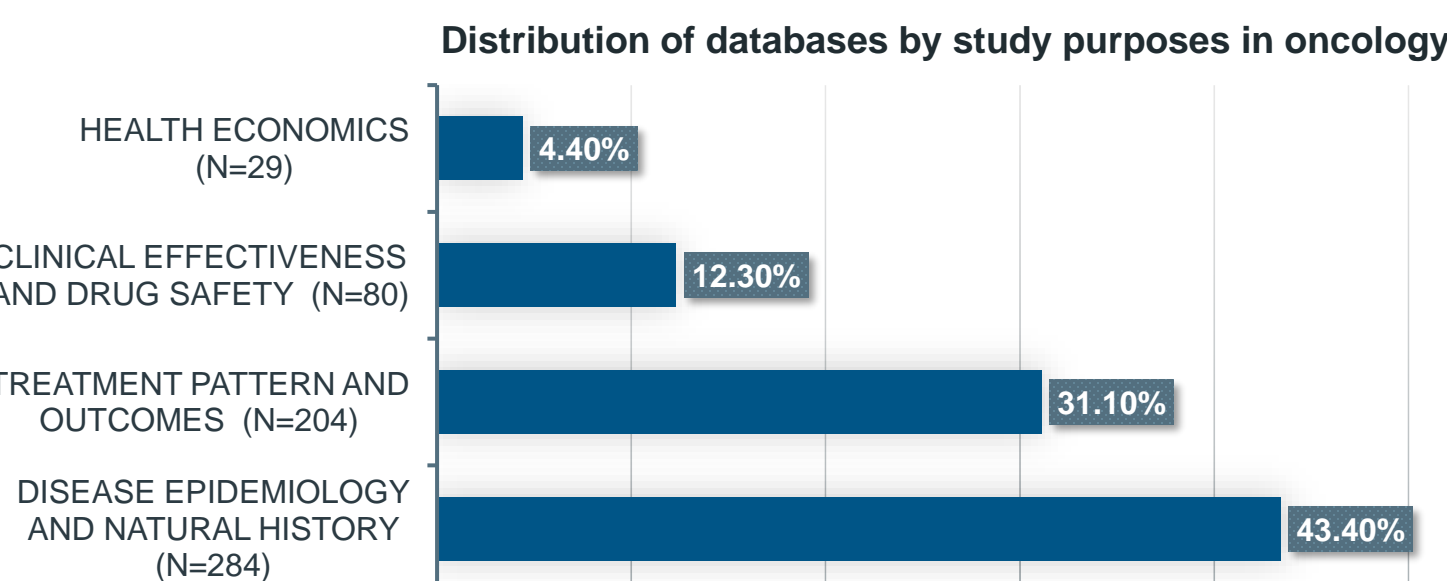
Note: Studies with "other" purposes (8.8% of 654 studies) were not shown in chart. "Other" purposes were those about diagnostic approach optimization, history of database establishment, database quality control, etc.

Challenges and opportunities for real-world databases in cancer care:
A synthesis based on publication review and qualitative interview

Challenges in data quality
- Few databases have regular linkage to death registry.
- Distinct inconsistencies in data entries and formats. A lack of unified and standardized data governance process
- Regular data quality control process is not well described for most databases.
- rEHR has the limitation in national representativeness in addressing certain research questions; using multiple rEHRs with Triangulation in Research may increase the credibility and validity.

Challenges in database operation
- Data from disease registries and EMR databases have been extensively used in oncology.
  - Heterogeneity in data ownership, access process, and compliance for PI-led registries. To utilize such registries, feasibility in both data quality and operation process is required.
  - Early alignment of intellectual property is necessary for research collaboration.

Opportunity: Data aggregation via NATDSS network
- The importance of healthcare data aggregation through a compulsory national network is emphasized in both literature and interviews.
- The NATDSS network serves as an oncology data aggregator in China, collecting data from over 15 million cancer patients.
- Data are linked with CDC death registry.
- Data access model is based on one centralized Ethics Committee review.

Promoting the **continued development and broader application** of the **NATDSS network** as an oncology data aggregator for research is encouraged. This network has the potential to offer a **nationally representative** population of cancer patients and **comprehensive** clinical endpoint evaluations from over **2,000 hospital information systems in 31 provinces**, facilitating real-world evidence generation in China.

Data Aggregation Into NATDSS

- Demographics
- Socioeconomic Characteristics
- Tumor Stage and Characteristics
- Clinical Lab Testing
- Pathology and Imaging Results
- Adverse Effects
- Clinical Outcomes (mortality, progression)
- Treatments
- Clinical Visits and Departments
- Direct Costs
- Healthcare Provider Characteristics

- Patient Profile
- Disease Epidemiology
- Drug Effectiveness and Safety
- Treatment Patterns
- Healthcare Resource Utilization and Costs
- External Comparator
- Natural History of Disease

Building on the NATDSS network, **strategic partnerships among stakeholders**, including government authorities, industry, academia, and the healthcare system could be more effectively implemented. This would be advantageous for supporting **clinical research, drug management** and **health economics evaluation,** establishing **disease-specific cohorts,** and **enhancing patient outcomes** in China.

Lung cancer
- 1 million +

Breast cancer
- 0.8 million +

Gastric cancer
- 0.6 million +

Liver cancer
- 0.5 million +

Adjuvant treatment strategy evolution and risk stratification for hormone receptor-positive, human epidermal growth factor receptor-2 negative early breast cancer in China
Ying Fan, Danyang Ji, Mingxia Jiang, Yujing Tan, Yang Yang, Tianyi Li, Xiao Ma, Binghe Xu ✉ Author Notes
The Oncologist, Volume 29, Issue 9, September 2024, Pages e1104–e1112,
https://doi.org/10.1093/oncolo/oyae095

The current landscape of gastric cancer and gastroesophageal junction cancer diagnosis and treatment in China: a comprehensive nationwide cohort analysis
Yang Chen, Keren Jia, Yi Xie, Jiajia Yuan, Dan Liu, Lei Jiang, Haoxin Peng, Jia Zhong, Jian Li, Xiaotian Zhang & Lin Shen ✉
Journal of Hematology & Oncology 18, Article number: 42 (2025) | Cite this article

Front. Oncol. 29 November 2022
Sec. Thoracic Oncology
Volume 12 - 2022 | https://doi.org/10.3389/fonc.2022.974227

Immune checkpoint inhibitors alone or in combination with chemotherapy for treatment of advanced non-small cell lung cancer after first-line platinum-based chemotherapy: A propensity score matching analysis

BREAST CANCER–LOCAL/REGIONAL ADJUVANT
e12014
Publication Only
Optimal neoadjuvant treatment and prognostic factors in patients with HR-positive/HER2-positive early or locally advanced breast cancer: A national real-world study in China.
Zhenzhen Liu, Jiuyan Zhu, Chengcheng Wang, Zhenbo Liu, Xiaohan Chen, Lanshang Li, Xuefu Sun, Chongjian Zhang, Jianghua Qiao, Min Yao, Department of Breast Disease, Henan Breast Cancer Center, Affiliated Cancer Hospital of Zhengzhou University, Zhengzhou, China

CONCLUSIONS

- The scope of real-world databases in oncology is **extensive and growing**, encompassing a variety of data sources in China.

- **Data aggregation** facilitates **comprehensive analysis** of patient demographics, drug treatments, biomarkers, and survival outcomes, thereby improving the understanding of cancer epidemiology and treatment efficacy within the Chinese population.

- Encouraging the **continued development and broader application** of the **NATDSS network** as an oncology data aggregator for research would contribute to improved **clinical research, drug management, health economics evaluation** and **patient outcomes** in China.