Editorial: Registries and Population Databases in Clinical Research and Practice

Dinah V. Parums, MD PhD

Science Editor, Medical Science Monitor, International Scientific Information, Inc., Melville, NY, U.S.A.
e-mail: dinah.v.parums@isi-science.com

Abstract   Patient registries include data on patient diagnosis, demographics, treatment, and outcomes and are now fundamental to the provision of successful global health systems. Patient registries include mainly local, regional, and national patient data on general or specific patient groups. Global registries currently exist mainly for rare diseases. Some of the most studied registries include the national Surveillance, Epidemiology, and End Results (SEER) program and the hospital-based Medical Information Mart for Intensive Care (MIMIC-III) dataset. The limitations of registry databases have included lack of feedback from clinical studies to the clinical center, the lack of patient involvement, and limited findings on patient-reported outcomes (PROs). In September 2020, the European Medicines Agency (EMA) published its draft guidelines on registry-based clinical studies. Guidelines for the development and analysis of registry data will improve the quality and registry-based studies and increase the role of registry data to support clinical trials. This Editorial aims to present the current status of registries and population databases in clinical research and practice.

Keywords: Editorial • Registries • Population • Guidelines as Topic • Epidemiology

The availability of large-scale patient data collections, or registries, including patient diagnosis, demographics, treatment, and outcomes, is now fundamental to the provision of successful global health systems [1,2]. Patient registries are also known as clinical audit, patient databases, population databases, and quality improvement programs, and may include local, regional, national, and global patient data for general or specific patient groups [1,2]. Globally, there are now hundreds of registries. Both the United States (US) and Sweden have more than 100 federally approved registries that report health quality metrics for all age groups [3,4]. Patient registries facilitate public health reports, retrospective and prospective research studies, professional medical development, and provide evidence to support healthcare improvement and funding [1,2]. In the UK, data from registries have recently supported advances in orthopedic practice and joint replacement, and patient management for cancer [5,6]. The US Centers for Disease Control and Prevention (CDC) publishes the Morbidity and Mortality Weekly Report (MMWR) on public health investigations, reports, and evidence-based clinical recommendations in its role in monitoring public health and disease epidemiology and mortality data in the US [7]. The MMWR has also provided a rationale for developing or studying patient registry data in clinical studies.

Registry data are used to support clinical trial data before regulatory approvals [1,8]. The National Institutes of Health publishes a list of current US-based patient registries that support clinical trials [3]. In September 2020, the European Medicines Agency (EMA) published its draft guidelines on registry-based clinical studies [9]. The 2020 EMA guidelines recommend the valuation of data quality, data analysis, and patient confidentiality [9].

The Surveillance, Epidemiology, and End Results (SEER) program began in 1973, initially included nine cancer registries in the USA, and is funded by the National Cancer Institute (NCI) [10]. SEER collects clinical and demographic patient data on cancer incidence, treatment, and patient survival, currently from 18 selected cancer registries throughout the USA, and covers 28% of the US population [10]. Also, SEER-Medicare linked databases include cancer data for patients age 65 years and older [11]. The National Program of Cancer Registries (NPCR) was funded by the Centers for Disease Control and Prevention (CDC) in 1992 to support central cancer registries and covers 96% of the US population, including 45 US states, the District of Columbia, the US Pacific Island Jurisdictions, and Puerto Rico [12]. SEER population studies have been published for decades. However, the United States Cancer Statistics (USCS) database of cancer registries combines SEER with the NPCR registries and has received less research attention because it has only recently become available [13]. The USCS database provides oncology data geographically coded at the local level to facilitate clinical oncology planning and evaluation [14]. Based on data from the USCS database, the CDC, NCI, and the North American Association of Central Cancer Registries (NAACCR)

publish official federal statistics on cancer incidence from registries that meet the publication standards of the USCS [14]. The USCS database is now available inside National Centers for Health Statistics (NCHS) and Census Research Data Centers (RDCs) to validated researchers [14].

Studies using the USCS database have highlighted some of the limitations of SEER data [15]. In 1999, a study to compare the cancer incidence rates from SEER and NPCR data showed that US cancer incidence rates for specific sites varied between registries [15]. Also, SEER data under-represented cancer incidence rates for all sites combined, over-represented the incidence of breast cancer, and under-represented the incidence of colorectal cancer [15,16]. More recent comparative studies have shown that SEER under-represented cancer-related mortality in several cancer sites between geographic areas [17]. Therefore, national population registries that collect regional data can have limitations, which may be due to different regional socioeconomic factors, demographic factors, and variations in health care over time [15-17].

Local and regional public health databases have provided epidemiological evidence to guide health provision and funding for previously high-risk population groups. For example, data analysis from the California Birth Statistical Master Files (BSMF) database from 2007 to 2016 identified ethnic and socioeconomic risk factors for low birth weight infants and maternal mortality and provided evidence for the adverse effects on these outcomes from maternal obesity and smoking [18]. In the US, access to individual hospital registries has become possible for research purposes. For example, the Medical Information Mart for Intensive Care (MIMIC-III) dataset version 1.4. MIMIC-III integrates clinical data of 53,423 hospital admissions of adult patients to the intensive care unit (ICU) of the Deaconess Medical Center, MA, USA, between 2001 and 2012 [19]. External access to the MIMIC-III database requires approval from the Massachusetts Institute of Technology (MIT)

and Beth Israel Deaconess Medical Center and the completion of a course on how to search and analyse the data from MIMIC-III [19].

The limitations of registry databases and their use should be recognized. Feedback of clinical data and research findings from registry studies to the participating clinical centers may lag behind patient care [1,2]. Also, some registries rely on non-digital patient records and data analysis and use manual data entry, resulting in errors or missing content [1,2]. Most registries lack patient input into their design and management, and demographic and socioeconomic data that affect health outcomes may not be included [20]. Because data in patient registries are anonymized, patients cannot access information that may support self-management or allows participation in shared clinical decisions [20].

However, some recent changes are beginning to address these limitations. National clinical audits in the UK, Sweden, and the Netherlands now include patient-reported outcome (PRO) data with clinical registry data [21]. In the US, the ImproveCareNow network for inflammatory bowel disease involves patients and their families, care teams, and clinical scientists in the design, governance, and management [22]. The Swedish Rheumatology Quality Registry allows patients to track symptoms at home to identify early signs of increased disease activity, resulting in improved patient management and clinical outcomes [23].

Conclusions

Patient registries include local, regional, and national patient data on general or specific patient groups and diseases. Global registries currently exist mainly for rare diseases. Guidelines for the development and analysis of registry data continue to improve the quality of registry-based studies, including registry data to support clinical trials.

References:

1. Gliklich RE, Dreyer NA, Leavy MB. Editors. Registries for evaluating patient outcomes: A user's guide. 3rd edition. Rockville (MD): Agency for Healthcare Research and Quality (US). April 1, 2014. Patient Registries. https://www.ncbi.nlm.nih.gov/books/NBK208643/

2. Nelson EC, Dixon-Woods M, Batalden PB, et al. Patient focused registries can improve health, care, and science. BMJ. 2016;354:3319

3. National Institutes of Health (NIH). Clinical Research Trials and You. List of Registries. https://www.nih.gov/health-information/nih-clinical-research-trials-you/list-registries

4. Örnerheim M. Policymaking through healthcare registries in Sweden. Health Promot Int. 2018; 33(2):356-65

5. Smith AJ, Dieppe P, Vernon K, et al. National Joint Registry of England and Wales. Failure rates of stemmed metal-on-metal hip replacements: analysis of data from the National Joint Registry of England and Wales. Lancet 2012;379:1199-204

6. Cornish JA, Tekkis PP, Tan E, et al. The national bowel cancer audit project: the impact of organisational structure on outcome in operative bowel cancer within the United Kingdom. Surg Oncol 2011;20:e72-77

7. Centers for Disease Control and Prevention (CDC). Morbidity and Mortality Weekly Report (MMWR). 2021. https://www.ncbi.nlm.nih.gov/pmc/journals/2817

8. Olmo CA, McGettigan P, Kurz X. Barriers and opportunities for use of patient registries in medicines regulation. Clin Pharmacol Ther. 2019;106(1):39-42

9. European Medicines Agency (EMA). Draft guidelines on registry-based clinical studies. Sept 2020. https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en.pdf

10. National Cancer Institute (NCI). Surveillance, Epidemiology, and End Results (SEER) Program. 2021. http://seer.cancer.gov/registries/

11. Warren JL, Klabunde CN, Schrag D, et al. Overview of the SEER-Medicare data: Content, research applications, and generalizability to the United States elderly population. Med Care. 2002;40(8 Suppl.):IV-3-18

12. Center for Disease Control and Prevention (CDC). National Program of Cancer Registries (NPCR) Standards.
http://www.cdc.gov/cancer/npcr/standards.htm

13. Center for Disease Control and Prevention (CDC). United States Cancer Statistics (USCS). The Official Federal Cancer Statistics. 2021.
https://www.cdc.gov/cancer/uscs/index.htm

14. Centers for Disease Control and Prevention (CDC). National Program of Cancer Registries (NPCR). 2021.
https://www.cdc.gov/cancer/npcr/index.htm

15. Wingo PA, Jamison PM, Hiatt RA, et al. Building the infrastructure for nation-wide cancer surveillance and control – a comparison between the National Program of Cancer Registries (NPCR) and the Surveillance, Epidemiology, and End Results (SEER) Program (United States). Cancer Causes Control. 2003;14(2):175-93

16. Henley SJ, King JB, German RR, et al. Centers for Disease Control and Prevention (CDC). Surveillance of screening-detected cancers (colon and rectum, breast, and cervix) – United States, 2004-2006. MMWR Surveill Summ. 2010;59(9):1-25

17. Merrill RM, Dearden KA. How representative are the surveillance, epidemiology, and end results (SEER) program cancer data of the United States? Cancer Causes Control. 2004;15(10):1027-34

18. Ratnasiri AWG, Lee HC, Lakshminrusimha S, et al. Trends in maternal pre-pregnancy body mass index (BMI) and its association with birth and maternal outcomes in California, 2007-2016: A retrospective cohort study. PLoS One. 2019;14(9):e0222458

19. Johnson AE, Pollard TJ, Shen L, et al. MIMIC-III, a freely accessible critical care database. Sci Data. 2016;3:160035

20. Reuben DB, Tinetti ME. Goal-oriented patient care – an alternative health outcomes paradigm. N Engl J Med. 2012;366:777-79

21. Black N. Patient reported outcome measures could help transform healthcare. BMJ. 2013; 346:167

22. Crandall W, Kappelman MD, Colletti RB, et al. ImproveCareNow: The development of a pediatric inflammatory bowel disease improvement network. Inflamm Bowel Dis 2011;17:450-57

23. Eriksson JK, Askling J, Arkema EV. The Swedish Rheumatology Quality Register: Optimisation of rheumatic disease assessments using register-enriched data. Clin Exp Rheumatol 2014; 32(Suppl.85):S147-49