REVIEW   **Open Access**

Scoping review of the recommendations and guidance for improving the quality of rare disease registries

JE Tarride[1,2,3], A. Okoh[1], K. Aryal[1], C. Prada[1], Deborah Milinkovic[2*], A. Keepanasseril[1] and A. Iorio[1]

Abstract

Background Rare disease registries (RDRs) are valuable tools for improving clinical care and advancing research. However, they often vary qualitatively, structurally, and operationally in ways that can determine their potential utility as a source of evidence to support decision-making regarding the approval and funding of new treatments for rare diseases.

Objectives The goal of this research project was to review the literature on rare disease registries and identify best practices to improve the quality of RDRs.

Methods In this scoping review, we searched MEDLINE and EMBASE as well as the websites of regulatory bodies and health technology assessment agencies from 2010 to April 2023 for literature offering guidance or recommendations to ensure, improve, or maintain quality RDRs.

Results The search yielded 1,175 unique references, of which 64 met the inclusion criteria. The characteristics of RDRs deemed to be relevant to their quality align with three main domains and several sub-domains considered to be best practices for quality RDRs: (1) governance (registry purpose and description; governance structure; stakeholder engagement; sustainability; ethics/legal/privacy; data governance; documentation; and training and support); (2) data (standardized disease classification; common data elements; data dictionary; data collection; data quality and assurance; and data analysis and reporting); and (3) information technology (IT) infrastructure (physical and virtual infrastructure; and software infrastructure guided by FAIR principles (Findability; Accessibility; Interoperability; and Reusability).

Conclusions Although RDRs face numerous challenges due to their small and dispersed populations, RDRs can generate quality data to support healthcare decision-making through the use of standards and principles on strong governance, quality data practices, and IT infrastructure.

Keywords Rare diseases, Registries, Quality standards, Guidance

*Correspondence:
Deborah Milinkovic
milinkda@mcmaster.ca
[1]Department of Health Research Methods, Evidence and Impact, Faculty of Health Sciences, McMaster University, Hamilton, Canada
[2]Centre for Health Economics and Policy Analysis (CHEPA), McMaster University, Hamilton, Canada
[3]Programs for the Assessment of Technologies in Health (PATH), The Research Institute of St. Joe's Hamilton, St. Joseph's Healthcare Hamilton, Hamilton, ON, Canada

Introduction

Randomized clinical trials (RCTs) for many years have been the main source of clinical evidence for regulatory and reimbursement decisions of healthcare technologies. However, as regulators and health technology assessment (HTA) agencies move towards a life cycle approach [1, 2], there is an opportunity to broaden the evidence base and enhance decision-making through the integration of real-world evidence (RWE) into decision making. Based on real world data (RWD), RWE allows decision-makers to better understand how health technologies are being used, how they perform, and whether they are cost-effective in real-world healthcare settings. It is therefore not surprising that several frameworks have been developed over the past years to guide the use and reporting of RWD for decision making [3–9]. For common diseases, RWE is often provided by post-marketing phase IV clinical trials, administrative databases, or electronic medical records. In the case of rare diseases (RDs) characterized by small populations (e.g., fewer than one in 2,000 as per the Canadian or European definitions, or fewer than 200,000 in the US) [10], both traditional trials and common sources of RWD may be not providing sufficient evidence. For example, it may also not be feasible or ethical to conduct clinical trials for RDs [11–13]. Therefore, high-quality rare disease registries (RDRs) can play an important role in HTA, health policy, and clinical decision-making for RDs [14, 15]. RDRs can improve our knowledge of RD conditions, support clinical research, improve patient care, and inform overall healthcare planning. However, RDRs are often diverse in nature, supported by different data governance and funding models, and may lack standardized data collection methods [16]. As such, HTA agencies may be reluctant to use RDR data to inform funding decisions on treatments for rare diseases [17, 18].

To support acceptance of registry data by HTA bodies wishing to use registry data, the European Network for Health Technology Assessment (EUnetHTA) Joint Action 3 led the development of the "Registry Evaluation and Quality Standards Tool" (REQueST) [19] based on the Methodological Guidance on the Efficient and Rational Governance of Registries (PARENT) guidelines [20] and a series of HTA consultations [17, 20]. Although not specific to RDRs, the REQueST tool includes 23 criteria to support the assessment of whether registries meet the needs of the regulatory and HTA bodies including eight criteria describing the methodology used (type of registry; use for registry-based studies and previous publications; geographical and organizational setting; duration; size; inclusion/exclusion criteria; follow-up and confounders), 11 criteria that are essential standards for good practices and data quality (registry aims and methodology; governance; informed consent; data dictionary;

minimal data set; standard definitions; terminology and specifications; data collection; quality assurance; data cleaning; missing data; financing; protection; and security and safeguards) and three criteria that deal with information that may be required when evaluating a registry for a particular purpose (e.g., interoperability and readiness for data linkage; data sources; and ethics). The REQueST tool was piloted with two established European registries and results indicated that both registries performed well, with more than 70% of the domains rated satisfactory and none of the domains failed. However, results indicated that more information was required in terms of governance structure (e.g., the role of industry), data quality checks, and interoperability [17].

The REQueST tool was also used by the Canadian Agency for Drugs and Technologies in Health (CADTH) to describe 25 RDRs based on publicly available information reported by the RDRs [21]. Within the study limitations (e.g., an assessment with the REQueST tool should be completed by registry data holders and not based on public information), the results indicated that most Canadian RDRs scored well for the 8 methodological criteria, although no RDRs provided public information on methods used to measure and control confounding. While information on the RDR purpose, governance, and informed consent was publicly available for almost all RDRs, there was considerable variation in the amount of publicly available information on the other REQueST criteria for the 25 Canadian RDRs, thus prompting a call for the establishment of Canadian standards for RDRs [21]. Therefore, to support decision making around the approval or funding of treatments for RDs in Canada and elsewhere, the objective of this study was to identify best practices to improve the quality of RDRs.

Methods

A scoping review was conducted to meet the study objectives, as scoping review designs are particularly appropriate to answer broad research questions [22]. The scoping review included four steps: (1) developing the literature search strategy; (2) study selection; (3) data charting; and (4) summarizing and reporting the results.

Search strategy

The search strategy was developed by a librarian from CADTH. The search strategy (Appendix 1) included several search terms (e.g., rare disease, registry, recommendations, guidance, standards). Databases searched were MEDLINE and EMBASE and the search was restricted to articles published in English from 2010 to April 2023. The year 2010 was chosen as the cut-off point because 2010 corresponds to the guidance on RDRs published by the European Rare Disease Task Force initially published in 2009 and updated in 2011 [23]. Grey literature was

searched from websites of regulatory bodies (e.g., European Medicines Agency, Food and Drug Administration, Health Canada) and HTA authorities (e.g., National Institute for Health and Care Excellence, CADTH).

Study selection

Screening for articles that met the inclusion was conducted using Rayyan [24]. Titles and abstracts were screened against the inclusion and exclusion criteria (Level I screening). Full texts of the publications that passed the Level I screening were retrieved before being screened for final inclusion and exclusion (Level II screening). The literature was screened by two pairs of independent reviewers (KA & CP; AK & AO) at each stage of the Level I and Level II screenings. Conflicts within each pair of reviewers were resolved through discussion. When consensus could not be reached an additional reviewer was consulted (JET). The same process was used for screening the grey literature.

Inclusion and exclusion criteria

Literature was included if it was reporting on standards, processes, guidance, or recommendations for improving the quality of RDRs. Exclusion criteria included: (1) non-English literature; (2) conference proceedings and letters; and (3) papers presenting clinical data based on an existing RDR without reporting on standards, guidance, or considerations relevant to RDR quality. The references cited in the included papers were also scanned to identify any relevant literature, including non-RDR guidance cited in the RDR literature.

Data charting

Based on the preliminary scoping of the literature, the following data were selected for abstraction: publication details and specific guidance related to RDRs' governance; patient engagement and consent; diversity and equity issues; funding model and sustainability; ethical/legal/regulatory requirements; data quality and management; data elements; standardization; data linkage; data validity and audit; IT infrastructure; and barriers and facilitators for improving the quality of RDRs. Data was abstracted into a Microsoft Excel spreadsheet.

Data summary and synthesis

Once the data were abstracted, summaries were created by the team and the information was synthesized in terms of best practices for improving the quality of RDRs.

Results
Results of the search strategy

Out of 1,135 unique citations identified by the search, 93 were assessed for eligibility based on a full-text review, and 47 studies were included for data abstraction. For the grey literature, 35 documents were identified, 18 were assessed for eligibility based on full-text review and 6 documents were included for data abstraction. In addition, 11 documents were identified by reviewing the references cited in the included papers, for a total of 64 documents included in our scoping review. Figure 1 presents the PRISMA diagram summarizing the screening process and key reasons for exclusion. Appendix 2 presents the list of the 64 documents identified through the literature review and used to develop the framework.

Conceptual framework

Upon review of the evidence and authors' discussion, the literature was synthesized according to three key quality domains (governance, data, and information technology) and several sub-domains: eight for governance (registry purpose and description; governance structure; stakeholder engagement; sustainability; ethics/legal/privacy; data governance; documentation and training and support), six for data (standardized disease classification; common data elements; data dictionary; data collection; data quality and assurance; and data analysis and reporting), and two for IT infrastructure (physical/virtual infrastructure; and software infrastructure).

Domain 1: RDR governance

Governance was the most discussed domain (48 of 64 sources), which was not surprising given that governance is foundational for quality and trust. Governance refers to the formalized structure that guides the RDR leadership and high-level decision-making required to achieve RDR's objectives and long-term operational sustainability [13, 16, 25]. The following describes the guidance reported in the literature for each of the 8 governance sub-domains, while Table 1 summarizes the key guidance.

Sub-domain 1 — registry purpose and description

The critical first step in any registry description is to state its purpose and objectives since they establish the framework for all activities that follow (e.g., data collection, inclusion, and exclusion criteria). A comprehensive description of the registry, available through the registry website or publications, allows other stakeholders, including potential researchers or regulatory or HTA users, to understand and appraise the registry's quality and potential usefulness. In addition to the RDR purpose and objectives, common attributes reported in the literature to describe a registry include registry design, timeframe, population characteristics, settings, geographical coverage area, type of data captured and data sources, data quality procedures, data access policies, ethics approvals and dissemination activities [17, 20, 23, 25–42].
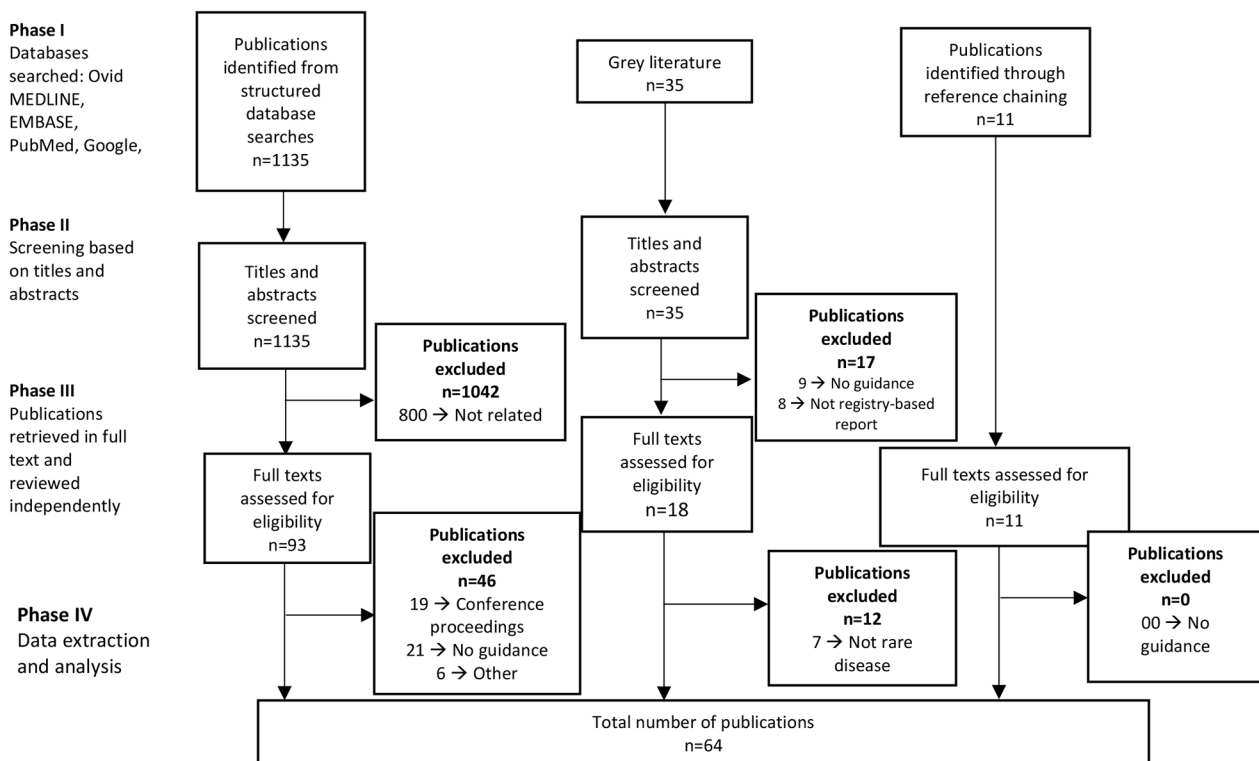
Fig. 1 PRISMA diagram

Sub-domain 2 — governance structure

The governance structure reflects the nature and extent of registry operations [13]. As for many organizations, the adoption of an organigram (a visual representation of the registry governance structure) helps clarify roles and responsibilities, reporting and decision-making flow, and how the different roles interconnect [29, 43]. Examples of key roles and expertise reported in the RDR literature include: registry lead(s); project manager(s) and management team with financial and leadership experience; information technology experts; data entry personal; and team members with specific expertise (e.g., ethics, legal, statistics, population-based research) [25, 31, 37, 38, 42, 44]. A central contact point for stakeholders is advisable [18, 38, 45].

Depending on the size and scope of the registry, a governing body, spanning from an independent Board of Directors to a Steering Committee comprised of various internal and external experts, has been recommended [35, 36]. The role of the governing body is to direct daily operations and ensure compliance with applicable laws and regulations, directly or through small targeted work groups [42, 45, 46]. In addition, independent advisory boards can provide technical guidance and scientific independence [13, 23, 47]. Patient representation on governing bodies or committees facilitates patient centeredness and engagement in the decision-making process [17, 20, 25]. As for most public and private organizations, board and committee members should declare conflicts of interest to enhance transparency [17, 35, 48].

Sub-domain 3 — stakeholder engagement

Multi-stakeholder engagement (e.g., clinicians, patients and their families, patients organizations, provider organizations, regulators, payers, drug companies) is suggested to facilitate long-term sustainability of RDRs [23, 25, 37, 38, 41, 46, 49–53]. The integration of a broad group of stakeholders can also facilitate quality improvements [23, 30, 50, 54]. Patient advocacy groups, for example, can enhance the accuracy and completeness of patient data [49]. However, the literature points out that decision-making may be challenging with a large number of stakeholders [25].

Sub-domain 4 — sustainability

Durable and long-term sustainability is dependent upon funding and is key to ensuring RDR quality [25, 36]. Compared to non-RD registries that have large populations to draw upon, RDRs are constrained by small and dispersed patient populations and limited funding opportunities. These constraints can inhibit data accuracy, patient follow-up, standardization, and result in knowledge gaps [26, 29, 30, 52, 54]. Multiple funding sources (e.g., public or private organizations, public-private partnerships,

Table 1 Literature guidance on RDR governance

Sub-domain

Registry purpose and description

• Registry purpose/objectives and description (e.g., design, timeframe, population characteristics [e.g., inclusion/exclusion criteria, sample size, representativeness], settings, geographical coverage area, health interventions considered in the registry, type of data captured and data sources, data quality procedures, data access policies, ethics approval, dissemination activities) [17, 20, 23, 25–42].

Governance structure

• The governance structure reflects the nature and extent of registry operation [13].

• Organigram outlining the roles and responsibilities within the registry [18, 23, 25, 29, 31, 37, 38, 42–45].

• Board of Directors or Steering Committee performing an oversight function [35, 36, 42, 45, 46, 48].

• Independent advisory board providing technical guidance and scientific independence (where appropriate) [13, 23, 47].

• Inclusion of patients in the decision-making process [17, 20, 25].

Stakeholder engagement

• Broad range of stakeholders across health facilities, regulators, patients, industry partners, etc [23, 25, 30, 37, 38, 41, 46, 49–54].

Sustainability

• Access to multiple funding sources to facilitate long-term sustainability [17, 18, 25, 35, 44, 45, 49].

• Long-term comprehensive financial plan [25–27, 29, 30, 36, 39, 47, 50, 52, 54, 55].

Ethics, Legal, Privacy

• Compliance with the relevant (international, national, or local) ethical, legal, and privacy regulations [23, 25, 32, 33, 35, 36, 39, 42, 49, 50].

• Informed consent for data collection, storage, use and reuse, data linkages, and right to withdrawal of consent [17, 20, 23, 27–30, 34, 38, 45, 50, 56].

• Approaches used to encourage participation, if any, e.g., incentives [26].

Data governance

• Clear definition and documentation of data ownership rights [18, 29, 35, 36, 38, 45, 46, 48, 59].

• Adoption of a managed access approach for third-party data requests [16, 26].

Documentation

• Standard Operating Procedures (SOPs) manual [25, 30, 35, 38, 61].

• Regular reports on its activities and repository of registry-based publications [17, 25, 48, 62].

• Standardized forms (e.g., informed consent, Investigator Declaration Forms) [33, 40].

• Ethical and other regulatory approvals for registry-based studies [12, 32, 33, 47].

Training and support

• Training resources for registry staff, data providers, and new users [25, 38, 45, 63].

• Support team or help desk [59].

non-profit foundations, patient groups, professional societies) may contribute to the long-term sustainability of RDRs [25, 35, 44, 49], but for transparency, it is important that all funding sources are publicly disclosed [17, 18, 45, 50]. Sustainability also necessitates a long-term comprehensive financial plan including future-oriented exit strategies and succession planning [39, 47, 50]. A registry's utility, effectiveness, efficiency, and agility are also important to ensure its long-term sustainability [27, 55].

Sub-domain 5 – ethics, legal, privacy

RDRs must comply with ethical, legal, and privacy regulations [25, 32, 33, 35, 36, 39, 42] for the collection, storage, and use and re-use of patient health data for RDR activities [23, 27–30, 34, 38, 45] or for regulatory purposes [12, 45]. The collection of informed consent necessitates that participants understand the risks and benefits that might accrue to them specifically, who might have access to their data, how their data will be used and re-used including potential linkages to other registries or future research activities, and a participant's right to withdraw their consent at any time [17, 20, 45, 50, 56]. Since the withdrawal of consent impacts both the current data holdings and past, present, and future research analyses, precise language in the consent about the withdrawal consent (e.g. what happens to the data of an individual withdrawing consent and how it impacts the analyses) will help mitigate potential misunderstanding and future conflicts [13]. Approaches used to encourage participation, if any (e.g., incentives) should be documented [26]. For international registries or RDRs intending to link to international registries, multiple statutes may apply (e.g. EU General Data Protection Regulation (GDPR), Canada's Personal Information Protection and Electronic Documents Act (PIPEDA), various in the U.S.) [57, 58]. For this reason, it is often recommended that RDRs strive to comply with international, national, and local ethical, legal, and privacy regulations as appropriate [23, 25, 36, 49, 50].

Sub-domain 6 - data governance

Data ownership and data custodianship are at the forefront of data governance [16, 35, 36, 38, 48, 59]. Data ownership refers to the possession, responsibility, and control of data including the right to assign access privileges to others [60]. Patient participants may grant the registry authorization to access and use their data for research [16, 29, 38]. However, more than one entity (e.g., patients, clinicians, hospitals, funders) could have a claim to the aggregate data in the registry [16, 18, 38, 46]; therefore, data ownership must be clearly defined. Data custodianship is the responsibility of the registry organization, which includes monitoring and managing registry use, data access policies, and data sharing agreements [18, 45, 46]. A protocol for third-party data requests, such as the administration of these requests through a data access committee, will ensure that requests are appropriately assessed and responded to in a timely manner [16]. Full disclosure of the registry's fee structure (e.g., fees for

ad-hoc requests versus subscription fee models) will mitigate potential miscommunication or misinterpretation of data access requirements [26].

Sub-domain 7 — documentation
Documentation is essential to maintaining a quality registry because it facilitates shared understanding and transparency around the registry activities. A Standard Operating Procedures (SOPs) manual that is updated regularly provides step-by-step guidance on the registry's routine activities including performance targets [25, 30, 35, 38, 61]. Regular provision of activity reports (e.g., annual reports) and a repository of registry-based publications increase the transparency of the RDR processes and activities [17, 25, 48, 62]. Similarly essential is the documentation of ethical and regulatory approvals for registry-based studies [12, 32, 33, 47] and the adoption of standardized templates and forms (e.g., informed consent) that reflect the registry's objective and use standardized language [33, 40, 61]. The adoption of an Investigator and User Declaration Form or similar document

Table 2 Literature guidance on RDR data

Sub-domain

Disease classification standardization
 • Adoption of standardized disease classifications to facilitate interoperability and registry linkages [31, 32, 50, 65–67, 77].

Common data elements
 • Minimum or core group of data elements which can be expanded to meet the informational needs or objectives of the registry [37, 41, 50, 53, 56, 69, 76].
 • Process describing how to define and organize the core data elements [13, 25, 70, 71, 74, 75, 77].

Data dictionary
 • Data dictionary provides clear instructions for data entry [17, 20, 25].
 • Complete alignment between the data dictionary and the data being collected [55].

Data collection
 • Data collection procedure and standardized data collection forms [18, 35, 51].
 • Adequate staff training in data collection procedures to minimize measurement errors [35].
 • Data collection methods are flexible enough to promote equitable access to different demographics [13, 32, 37, 39, 40].
 • Adequate funding to facilitate data collection across a patient's lifespan [78].

Data quality and assurance
 • Data quality and assurance plan [12, 25, 32, 61, 64, 69].
 • Performing routine data quality checks and data cleaning [26, 31, 35, 45, 55].
 • Provision of regular feedback to data providers on data quality activities and registry-based research findings [25, 45, 50].

Data analysis and reporting
 • Study protocols inclusive of statistical analysis plans for registry-based studies [12, 13, 25, 45].
 • Adoption of reporting guidelines (e.g., STROBE, PCORI) to improve reporting transparency and accuracy [13].

will affirm compliance with regulatory and operational processes [32]. The literature also recommends publishing study protocols and registering registry-based studies in a public register [18, 45].

Sub-domain 8 – training and support
Training is essential for registry staff, data providers, and new users to ensure consistency and quality [25, 38, 45, 63]. A training manual, "how-to videos", and a comprehensive training plan that are updated regularly facilitate consistent training protocols [25, 37, 54]. A registry might also benefit from designated data entry personnel who can systematically monitor and evaluate data quality [38]. A Support Team or Help Desk is also beneficial to the operations of the registry [59].

Domain – data
Data was the second most discussed domain (45 of 64 sources). Data refers to the structures, policies, and processes required to ensure a RDR can maintain a high-quality database [13]. A high-quality database is characterized by completeness, accuracy, usefulness, and representativeness [13, 25, 64], which is paramount for meeting the needs of decision-makers.

Table 2 summarizes the guidance reported in the literature, which is described in more detail below in terms of 6 sub-domains (standardized disease classification; common data elements; data dictionary; data collection; data quality and assurance; data analysis and reporting).

Sub-domain 1 – standardized disease classifications
Standardized disease classifications such as the Orphanet Rare Disease Ontology (ORDO), Human Phenotype Ontology (HPO), ORPHA-codes or the International Classification of Disease ICD-9, ICD-10, or ICD-11 or some combination [31, 32, 50, 65, 66] have been proposed for data collection to ensure future interoperability and registry linkages. Being able to link to other registries facilitates knowledge creation, decision making, and improvements in clinical care that may not otherwise be possible for small RD patient populations [33, 36, 49, 51, 67]. When RDRs transfer or merge their data to or with other entities, the documentation of the process used to validate the data transfer ensures quality and consistency [68]. Although linking to international RDRs expands population reach, it poses some additional challenges (e.g., the regulatory environment) that need to be identified early in the registry design process [44]. The use of international standards and ontology codes apply in this context.

Sub-domain 2 – common data elements
Registries have to consider the informational needs of the registry against the needs of their other stakeholders and

the available resources [25]. A minimum set of common data elements collected across the RDR sites (e.g., administrative data, socio-demographics, diagnosis, disease history, treatments, clinical and safety outcomes) that could be expanded upon to meet the specific needs of the registry is usually identified [37, 41, 50, 53, 56, 69]. Ideally, these common data elements would be harmonized across all registries that represent the same rare disease when applicable [12]. However, the main challenge around common data elements is reaching a consensus regarding the choice, organization, and definition of the various elements [25, 70, 71]. Beyond simply determining the composition of the common data elements, other challenges include data coding standards (e.g., integer, float, string, date, derived data, and file names) [13, 72, 73], standardized data constructs, vocabulary and terminology [28, 33, 37, 65, 71, 74], defined variable interpretation to avoid inconsistency (e.g., sex – genotypic sex or declared sex) [18, 75] and ontology harmonization to facilitate convergence from different terms or languages [56, 65, 76, 77]. The latter necessitates consistent agreed-upon disease classification standards [23, 50, 77].

Sub-domain 3 – data dictionary

A detailed data dictionary is an essential tool for quality data collection [17, 20, 25]. A data dictionary provides clear instructions for data entry and analysis by defining all data elements and their purpose as well as the coding values including permissible values, representation class, data type, and format [17, 20]. Complete alignment between the variables described in the data dictionary and those captured by the registry's interface is expected [55].

Sub-domain 4 – data collection

Procedures for documenting the entire data collection process including adverse event monitoring, baseline and follow-up data, causality assessment, and reporting timelines are recommended to improve data accuracy [18, 51]. Standardized data collection forms (e.g., Clinical Data Interchange Standards Consortium Operational Data Model, Patient Records, and Outcome Management Information System) can facilitate the data collection process [35]. Training in data collection procedures is key to reducing information bias and data misclassification and to achieve consistency amongst users and high quality data collection [25, 35]. Sustained investment in data collection and management is also critical as prospective data collection across the patient's lifespan can be expensive and onerous [78]. The capacity for a registry to embed clinical studies into its own database can also help sustain the registry and reduce costs associated with duplicated data collection efforts when conducting additional studies [31, 78].

Data collection tools such as computers, automation, smartphones, smartwatches, tablets, and medical devices (e.g., glucose monitors) can be valuable sources of electronic health data as well as increase registry participation, particularly from disparate geographic locations, which in turn can result in increased knowledge, improved patient outcomes, stronger patient advocacy, and enhanced equity through healthcare access [13, 37, 39, 40, 75, 79]. However, internet-based data collection may impact equity for data providers with limited access to the internet. Relationship building with physicians and patient groups who serve often excluded groups can facilitate greater equity and inclusion through referrals and knowledge translation efforts that promote and encourage registry participation [32, 40].

Sub-domain 5 – data quality and assurance

Data quality reflects various data attributes or dimensions that can be used to measure the calibre of the data [25] such as completeness (the extent that the stored data represents the potential data), uniqueness (no repeated or redundant data), timeliness (data is up to date at the time of release), validity (data conforms to the appropriate syntax [e.g., format, type, range]), accuracy (the data correctly reflects the object or event being described), consistency (there are no discrepancies when the data is compared across different databases or against its definition) [35, 64], and usefulness (the extent to which the outputs provide value) [25].

Data quality and assurance plans which include data validation (e.g., medical, clinical, and record audit) [61] and a review of RDR-generated studies ensure compliance with RDR-based studies' protocols and ethical and regulatory requirements [12, 25, 32, 69]. Data quality and assurance processes necessitate routine data quality checks and data cleaning to ensure the enrolment of eligible patients, data completeness, validity, and coherence while mitigating record duplication and errors [26, 31, 35, 45, 55]. Data audits can be performed by internal registry staff or an external service provider, or some combination [37]. Regular feedback to data providers about these data quality activities and findings encourages prompt remedial action and learning, thus improving the quality of the RDR data [25, 45, 50].

Sub-domain 6 – data analysis and reporting

In addition to study protocols, RDR-based studies benefit from the development of statistical analysis plans (SAPs), whether for internal registry objectives or external research with third-party partners [12, 13, 25, 45]. A SAP facilitates the production of trustworthy results that can be more easily interpreted and accepted by various stakeholders (e.g., registry participants, patient groups, researchers, decision-makers, or the general public) [25].

SAPs should provide a list of variables and confounders captured in the data and details on the statistical methods used to answer the study question(s) and to deal with missing or censored data [25, 45]. Adoption of guidelines such as the Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement, or the Patient-Centered Outcomes Research Institute (PCORI) Methodology Report improves transparency and accuracy when reporting RDR findings [13]. Details about dissemination activities such as study reports and communication strategies are usually included in the study protocols [25, 44, 45].

Domain – information technology infrastructure

Information technology (IT) infrastructure was discussed in 29 of the 64 documents in terms of physical and virtual infrastructure, and software infrastructure. IT infrastructure refers to the critical infrastructure that is required to collect, share, link, and use patient and clinical data [16, 32, 37], and importantly, securely store, transmit, and manage this private data [32, 33, 63]. Table 3 summarizes the literature which is described below in more detail below.

Sub-domain 1 – physical and virtual infrastructure

High quality RDRs are characterized by procedures and processes that ensure the digitally stored private data is secure [29, 32, 35, 62, 80] by housing the data on dedicated servers with intrusion detection systems [35]. Critical decisions include where the server(s) are held (e.g., centralized database versus distributed); how these location(s) are secured, and how and who has access to the RDR data. Registries can safeguard their systems through several processes (e.g., analysis of threats and countermeasures [63]) and tools (e.g. data software and

access policies [63, 81, 82]). An independent external security or threat risk assessment is recommended to document compliance with security and privacy standards [83].

Sub-domain 2 – software infrastructure

The adoption of FAIR principles at the data source can facilitate data connections and exchanges across multiple RDRs and bolster data quality that supports both clinical research and patient care [31, 56, 68, 73, 75, 81]. FAIR data principles stand for Findability (easy to find for both humans and computers), Accessibility (easily retrievable and accessible by users), Interoperability (easily integrates with other data), and Reusability (well-described so it can be replicated or applied in another setting). Since FAIR principles enable the extensive and efficient use of registry data while mitigating duplication, recollection, and errors [25, 56], it is recommended that the registry data infrastructure complies with FAIR principles [18, 25, 26, 37, 38, 59, 75, 83, 84].

The technology choices, software architecture design and software development practices have a dramatic impact on software sustainability, legacy software support, ease of software modification, enhancements and interoperability [81]. With this in mind, software solutions can either be "out-of-the-box" commercial software or "home-built", custom-designed in-house software, the latter often being more powerful but more resource and time-intensive [25]. Either way, drop-down menus, pop-up explanatory notes, and tab-to-jump options will aid in rapid and user-friendly data entry [32, 72]. A user-friendly web interface with the capacity to upload and download data can also facilitate data sharing [25, 38]. Since registry data is often collected from several sources, machine-readable files could facilitate the interoperability of pseudonymized data subsets, reduce duplication, and make the data more findable [26, 63, 67, 75]. However, data heterogeneity can prove a barrier to automation [79].

Regardless of the methods used to collect, store, and manage data sets, data encryption and firewalling of servers are standard [59, 63]. The encryption of data while in transit is an added layer of data security [27, 32, 40]. As it might become necessary to delete data occasionally (e.g., participants who revoke their consent), standardized procedures for data deletion help maintain database integrity and mitigate errors [13, 38, 41]. Enhanced technological literacy and adoption of technology tools (e.g., electronic health records, automated data capture, internet and mobile devices) [44, 54] and the integration of a broader group of stakeholders [23, 30, 54] can also facilitate quality improvements.

Table 3 Literature guidance on information technology infrastructure

Sub-domain
Physical and virtual infrastructure
• Secure storage of data [29, 32, 35, 62, 80].
• Adoption of independent external security or regular threat risk assessments [82, 83].
Software infrastructure
• Compliance with FAIR principles at the data source to improve data quality and enhance data connections and exchanges across multiple RDRs [18, 25, 26, 31, 37, 38, 56, 59, 68, 73, 75, 81, 83, 84].
• User-friendly web design and data entry portal, when applicable [32, 72] ideally allowing upload and download of data [25, 38].
• Adoption of encryption and firewalling of registry data [27, 32, 40, 59, 63].
• Provision and procedures to scale up the use of technological tools to enhance registry quality [44, 54, 67, 79].
• Promotion of technological literacy among its stakeholders to facilitate quality improvement [23, 30, 54].

Discussion

Because RDRs can be designed around different purposes (e.g., patient advocacy, enhanced clinical practice, epidemiological and research goals) [25, 41, 46], RDRs often vary in quality, and are structurally and operationally diverse [38]. As a result, their fitness for purpose as a source of data to support decision making around the approval or funding of treatments for rare diseases must be assessed on a case-by-case basis [17, 18]. Fortunately, the RDR literature offers a range of quality standards that define the essential characteristics leading to the development and maintenance of a quality RDR. The guidance from the 64 sources captured by the scoping review is synthesized within three dominant domains: (1) governance, which represents the many operational features of governance such as the governance structure, stakeholder engagement, sustainability, ethic and regulatory oversight, and training; (2) data, which represents standardized ontology and common data elements and standardized processes for data entry, verification, and auditing and reporting; and (3) information technology, which represents physical and software infrastructure and security, guided by FAIR principles.

While many guidelines focused on certain dimensions of RDRs' quality (e.g., governance, core data elements), only three papers provided overall guidance on an extensive set of elements required to set up and maintain high-quality RDRs. Among those, in 2018, Kodra et al. 2018 [25] reported on the set of 17 recommendations to improve the quality of RDRs developed by a select group of experts convened by the Italian National Center of Rare Diseases in collaboration with other European countries [25]. These recommendations touched on 11 topics (registry definition; registry classification; governance; data sources; data elements; case report form; IT infrastructure complying with FAIR principles; quality information; documentation; training; and data quality audit) [25, 61]. Building on these recommendations and expert meetings, in 2020, Ali et al. [38] surveyed the RDR community to determine the consensus level regarding 17 criteria that could be considered essential when assessing the quality of a RDR in terms of registry governance (9 items), data quality (5 items), and IT infrastructure (6 items). The responses of 35 respondents representing 40 RDRs across the United States, Canada, the United Kingdom, and Europe indicated a high level of consensus among the RDRs with more than 90% of respondents agreeing with most of the 17 criteria. Of note, 30% of respondents did not feel that patient involvement in the registry governance was necessary, conceding that although patient involvement in RDR governance may be best practice, there may be a limited role for patients in some scenarios such as physician-driven registries. The 2021 European Medicine Agency (EMA) guidance

[45] integrated guidelines from PARENT Joint Action Methodological Guidance [20], the EUnetHTA's Registry Evaluation and Quality Standards Tool (REQueST) [19], the US Agency for Healthcare Research and Quality (AHRQ)'s Users' Guide on registries [13], and the European Reference Network Patient Registry Platform [85]. The EMA guidance provides information regarding two main domains: Administrative Information (subdomains include governance, consent and data protection) and Methods (subdomains include objectives, data providers, population, data elements, infrastructure, quality requirements). Our review broadly aligns with all three of these sources in terms of content, but it is more closely aligned with Ali et al. 2021 in terms of organization and number of domains (registry governance, data quality, IT infrastructure). However, compared to these guidances based on consensus panels or surveys, the evidence leading to our framework was based on a scoping review of the literature and we synthesized a broader set of quality indicators deemed essential by the literature. In December 2023, the FDA released its finalized real-world data guidance regarding a registry's fitness to support regulatory decision making, which is consistent with our framework (e.g., governance, data, information technology infrastructure) [86]. However, this guidance was published after the completion of the scoping review and as such was not included in this review.

Despite the literature on quality standards for RDRs, a review of 37 publications reporting on RDRs between 2001 and 2021 found that while most of these publications reported on collecting informed consent (81%) and provided information on data access, data sharing or data protection strategies (75%), fewer publications reported on quality management (51%) or maintenance (46%). Furthermore, fewer RDRs reported using core data elements (22%) or ontological coding systems (24%), which is key for interoperability and for linking registries [21]. It is however possible that RDRs had such policies in place but did not report on them. Initiatives such as those undertaken in Europe to develop guidance to improve the quality, reporting, and assessment of patient registries from both regulatory and HTA perspectives facilitate the integration of registry data into decision-making processes. For example, the REQueST tool has been used in Europe by HTA agencies to evaluate the quality of registries being used as a source of data to support decision-making [17]. However, the REQueST assessment of 25 Canadian RDRs based on publicly available information on these RDRs highlighted the importance of developing standards for Canadian RDRs [21]. In this context, the results of this scoping review could be used to help develop a Canadian consensus on the core standards defining high-quality RDRs from regulatory and HTA perspectives. Compliance with RWE guidance [87, 88]

and acceptance of evidence from other jurisdictions are other important considerations when using RDR data for decision-making, especially in countries with relatively small populations such as Canada.

While adhering to existing and future RDR and HTA guidance will certainly improve the quality of RDRs and their use in decision-making, it should be recognized that this may require significant investment in terms of human and financial resources which may not be easily available to all RDRs. However, at least for Canada, the recent announcement in March 2023 by Health Canada to invest $1.5 billion over three years in support of a National Strategy for Drugs for Rare Diseases [89] represents a unique opportunity to develop a national infrastructure of sustainable, standardized, and quality RDRs while aligning with the pan Canadian Health Data Strategy [90].

As a next step, for RDRs interested in the harmonization of their data collection with other registries, the European Platform on Rare Disease Registration (EU RD Platform) [91]serves as an example of how this might be achieved. With over 600 diverse and fragmented rare disease registries in Europe, the European Commission's Joint Research Centre in collaboration with stakeholders took on the tremendous task of establishing standards for integration, training, and interoperability of RDR data across Europe. A core element of the EU RD Platform is the European Rare Disease Registry Infrastructure (ERDRI) [92], which consists of a directory of registries, a data repository, a pseudonymization tool and importantly the EU RD Platform comprising of a set of 16 common data elements [93] that capture the characteristics of rare disease patients such as demographic, and clinical, diagnostic, and genetic information [43, 56, 76, 91, 94].

Limitations
Before interpreting the results of this scoping review, several limitations should be considered. First, due to the unique characteristics of RDRs, we limited our scoping review to RDRs, and we did not search the literature to improve the quality of non-RD registries. However, we identified several non-RDR guidances [12, 13, 20, 45, 64, 68] when checking the references of the RDR papers included in the scoping review. Second, although we took a systematic approach when selecting the papers to be included in our scoping review, it is always possible that we missed one or several studies, even though all included publications' references were checked to identify relevant studies not included in our final list of documents. Third, while we summarized the guidance under three domains and 16 sub-domains, we did not develop recommendations. This is left for future research. Despite these limitations, the results of this scoping review of 64 documents published between 2010

and April 2023 add to the body of the literature offering suggestions to improve the quality of RDRs. The results of this scoping review provide the foundation to develop quality standards for RDRs in Canada or other countries lacking guidelines for quality RDRs. For example, these results could be used by a Delphi panel to develop standards and processes to enhance the quality of data in RDR registries.

This review has also identified a few areas which merit further consideration. First, from a Canadian standpoint, future work is needed to develop a database of Canadian RDRs along with information on their key characteristics (e.g., purpose, population, funding) and information regarding their governance, data and IT infrastructure. Second, although the literature agrees on the importance of being able to link with international registries, it is also important to be able to link RDRs with health administrative databases to provide HTA agencies and decision makers with information on short and long-term outcomes, healthcare resource utilization, and expenditures associated with RDs. Similarly, issues of equity and diversity were discussed by only a few papers in the context of data collection methods to encourage patient participation [27, 40, 80], and relationships with physicians and patient groups working with disadvantaged groups [32, 40]. A broader RDR equity lens could be achieved by using equity tools such as the PROGRESS (Place of resident, Race, Occupation, Gender, Religion, Education, Socioeconomic status, Social capital) framework [95], which would facilitate a greater understanding of how equity-deserving populations are affected by RDs or represented in RDR registries. Finally, the integration of patients' experiences and insights when designing and interpreting results is an important avenue of research to enhance the quality and acceptance of RDR studies by generating patient-centered RWE [96].

Conclusion
Although RDRs face numerous challenges due to their small and dispersed populations, RDRs can generate quality data to support healthcare decision-making through the use of standards and principles on strong governance, quality data practices, and IT infrastructure.

Supplementary Information
The online version contains supplementary material available at https://doi.org/10.1186/s13023-024-03193-y.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Supplementary Material 4

Data availability
The authors confirm that a complete list of the sources used for data analyzed during this study is available in Appendix 2 of this published article. Example: https://doi.org/10.3390/ijerph15081644.

Declarations

Ethics approval and consent to participate
not applicable.

Consent for publication
all authors give their consent to publish.

Competing interests
no conflict of interest.

References

1. Canada's Drug and Health Technology Agency (CADTH): Ahead of the Curve: Shaping Future-Ready Health System. (2022). https://strategicplan.cadth.ca/wp-content/uploads/2022/03/cadth_2022_2025_strategic_plan.pdf. Accessed March 15, 2023.
2. National Institute for Health and Care Excellence (NICE): NICE strategy 2021 to 2026 - Dynamic, Collaborative, Excellent. (2021). https://static.nice.org.uk/NICE%20strategy%202021%20to%202026%20-%20Dynamic,%20Collaborative,%20Excellent.pdf. Accessed March 15, 2023.
3. Wang SV, Pinheiro S, Hua W, Arlett P, Uyama Y, Berlin JA, et al. STaRT-RWE: structured template for planning and reporting on the implementation of real world evidence studies. BMJ. 2021;372:m4856. https://doi.org/10.1136/bmj.m4856.
4. ISPOR: New Real-World Evidence Registry Launches. (2021). https://www.ispor.org/heor-resources/news-top/news/view/2021/10/26/new-real-world-evidence-registry-launches. Accessed March 15, 2023.
5. Corrigan-Curay J, Sacks L, Woodcock J. JAMA. 2018;320(9):867–8. https://doi.org/10.1001/jama.2018.10136. Real-World Evidence and Real-World Data for Evaluating Drug Safety and Effectiveness.
6. Finger RP, Daien V, Talks JS, Mitchell P, Wong TY, Sakamoto T, et al. A novel tool to assess the quality of RWE to guide the management of retinal disease. Acta Ophthalmol. 2021;99(6):604–10. https://doi.org/10.1111/aos.14698.
7. Schneeweiss S, Eichler HG, Garcia-Altes A, Chinn C, Eggimann AV, Garner S, et al. Real World Data in Adaptive Biomedical Innovation: a Framework for Generating evidence fit for decision-making. Clin Pharmacol Ther. 2016;100(6):633–46. https://doi.org/10.1002/cpt.512.
8. Gliklich RE, Leavy MB. Ther Innov Regul Sci. 2020;54(2):303–7. https://doi.org/10.1007/s43441-019-00058-6. Assessing Real-World Data Quality: The Application of Patient Registry Quality Criteria to Real-World Data and Real-World Evidence.
9. Reynolds MW, Bourke A, Dreyer NA. Considerations when evaluating real-world data quality in the context of fitness for purpose. Pharmacoepidemiol Drug Saf. 2020;29(10):1316–8. https://doi.org/10.1002/pds.5010.
10. Government of Canada: Building a National Strategy for High-Cost Drugs for Rare Diseases: A Discussion Paper for Engaging Canadians: A discussion Paper for Engaging Canadians. (2021). https://www.canada.ca/content/dam/hc-sc/documents/services/health-related-consultation/National-Strategy-High-Cost-Drugs-eng.pdf. Accessed July 20, 2023.
11. The Canadian Forum for Rare Disease Innovators (RAREi). Unique approach needed: Addressing barriers to accessing rare disease treatments. Submission to House of Commons Standing Committee on Health (HESA). https://www.ourcommons.ca/Content/Committee/421/HESA/Brief/BR10189782/br-external/CanadianForumForRareDiseasesInnovators-e.pdf (2013). Accessed November 8, 2023.
12. European Medicines Agency: Discussion paper: Use of patient disease registries for regulatory purposes – methodological and operational considerations. (2018). https://view.officeapps.live.com/op/view.aspx?src=https%3A%2F%2Fwww.ema.europa.eu%2Fen%2Fdocuments%2Fother%2Fdiscussion-paper-use-patient-disease-registries-regulatory-purposes-methodological-operational_en.docx&wdOrigin=BROWSELINK. Accessed November 17, 2023.
13. Gliklich RE, Leavy MB, Dreyer NA. Registries for evaluating patient outcomes: a user's guide (4th Eds.) (Prepared by L&M Policy Research, LLC under Contract No. 290-2014-00004-C with partners OM1 and IQVIA) (2020). https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-evaluating-patient-outcomes-4th-edition.pdf. Accessed March 7, 2023.
14. Canada's Drug and Health Technology Agency (cadth). Optimizing the Integration of Real–World Evidence as Part of Decision-Making for Drugs for Rare Diseases. What We Learned https://www.cadth.ca/sites/default/files/RWE/pdf/optimizing_the_integration_of_real_world_evidence_as_part_of_decision-making_for_drugs_for_rare_diseases.pdf. Accessed November 8, 2023.
15. Canada's Drug and Health Technology Agency (cadth): Report on a Best Brains Exchange. Optimizing the Use of Real-World Evidence as Part of Decision-Making for Drugs for Rare Diseases. (2022). https://www.cadth.ca/sites/default/files/RWE/pdf/MG0022_best_brains_exchange_optimizing_the_use_of_real_world_evidence_as_part_of_decision_making_for_drugs_for_rare_diseases.pdf. Accessed November 8, 2023.
16. Ali SR, Bryce J, Tan LE, Hiort O, Pereira AM, van den Akker ELT, et al. The EuRRECa Project as a model for Data Access and Governance policies for Rare Disease registries that collect clinical outcomes. Int J Environ Res Public Health. 2020;17(23). https://doi.org/10.3390/ijerph17238743.
17. Allen A, Patrick H, Ruof J, Buchberger B, Varela-Lema L, Kirschner J, et al. Development and Pilot Test of the Registry evaluation and quality standards Tool: an Information Technology-based Tool to support and review registries. Value Health: J Int Soc Pharmacoeconomics Outcomes Res. 2022;25(8):1390–8. https://doi.org/10.1016/j.jval.2021.12.018.
18. Jonker CJ, de Vries ST, van den Berg HM, McGettigan P, Hoes AW, Mol PGM. Capturing data in Rare Disease registries to Support Regulatory decision making: a Survey Study among Industry and other stakeholders. Drug Saf. 2021;44(8):853–61. https://doi.org/10.1007/s40264-021-01081-z.
19. EUnetHTA. REQueST: Tool and its vision paper. https://eunethta.eu/request-tool-and-its-vision-paper/ Accessed February 14, 2024.
20. Zaletel M, Kralj M. Methodological guidelines and recommendations for efficient and rational governance of patient registries (2015). https://health.ec.europa.eu/system/files/2016-11/patient_registries_guidelines_en_0.pdf. Accessed September 14, 2023.
21. Boyle L, Gautam M, Gorospe M, Kleiman Y, Dan L, Lynn E et al. Assessing Canadian Rare Disease Patient Registries for Real-World Evidence Using REQueST (2022). https://www.cadth.ca/sites/default/files/pdf/es0369-cadth-poster-laurie-lambert-final.pdf. Accessed March 15, 2023.
22. Sucharew H, Macaluso M. Methods for Research evidence synthesis: the Scoping Review Approach. J Hosp Med. 2019;14:416–8. https://doi.org/10.12788/jhm.3248.
23. Rare Disease Task Force: Patient registries in the field of rare diseases: overview of the issues surrounding the establishment, management, governance and financing of academic registries. (2011). https://www.orpha.net/actor/EuropaNews/2011/doc/RDTFReportRegistries2009Rev2011.pdf. Accessed March 7, 2023.
24. Ouzzani M, Hammady H, Fedorowicz Z, Elmagarmid A. Rayyan-a web and mobile app for systematic reviews. Syst Rev. 2016;5(1):210. https://doi.org/10.1186/s13643-016-0384-4.
25. Kodra Y, Weinbach J, Posada-de-la-Paz M, Coi A, Lemonnier SL, van Enckevort D, et al. Recommendations for improving the quality of Rare Disease registries. Int J Environ Res Public Health. 2018;15(8). https://doi.org/10.3390/ijerph15081644.

26. Gainotti S, Torreri P, Wang CM, Reihs R, Mueller H, Heslop E, et al. Eur J Hum genetics: EJHG. 2018;26(5):631–43. https://doi.org/10.1038/s41431-017-0085-z. The RD-Connect Registry & Biobank Finder: a tool for sharing aggregated data and metadata among rare disease researchers

27. Bellgard MI, Napier KR, Bittles AH, Szer J, Fletcher S, Zeps N, et al. Design of a framework for the deployment of collaborative independent rare disease-centric registries: Gaucher disease registry model. Blood Cells Mol Dis. 2018;68:232–8. https://doi.org/10.1016/j.bcmd.2017.01.013.

28. Biedermann P, Ong R, Davydov A, Orlova A, Solovyev P, Sun H, et al. BMC Med Res Methodol. 2021;21(1):238. https://doi.org/10.1186/s12874-021-01434-3. Standardizing registry data to the OMOP Common Data Model: experience from three pulmonary hypertension databases.

29. Boulanger V, Schlemmer M, Rossov S, Seebald A, Gavin P. Establishing patient registries for Rare diseases: Rationale and challenges. Pharm Med. 2020;34(3):185–90. https://doi.org/10.1007/s40290-020-00332-1.

30. Busner J, Pandina G, Domingo SZ, Berger AK, Acosta MT, Fisseha N, et al. Clinician-and patient-reported endpoints in CNS orphan drug clinical trials: ISCTM position paper on best practices for Endpoint Selection, Validation, Training, and standardization. Innovations Clin Neurosci. 2021;18(10–12):15–22.

31. Derayeh S, Kazemi A, Rabiei R, Hosseini A, Moghaddasi H. National information system for rare diseases with an approach to data architecture: a systematic review. Intractable rare Dis Res. 2018;7(3):156–63. https://doi.org/10.5582/irdr.2018.01065.

32. Marques JP, Carvalho AL, Henriques J, Murta JN, Saraiva J, Silva R. Design, development and deployment of a web-based interoperable registry for inherited retinal dystrophies in Portugal: the IRD-PT. Orphanet J Rare Dis. 2020;15(1):304. https://doi.org/10.1186/s13023-020-01591-6.

33. Rubinstein YR, Groft SC, Bartek R, Brown K, Christensen RA, Collier E, et al. Creating a global rare disease patient registry linked to a rare diseases biorepository database: Rare Disease-HUB (RD-HUB). Contemp Clin Trials. 2010;31(5):394–404. https://doi.org/10.1016/j.cct.2010.06.007.

34. National Cancer Registry Ireland: Data confidentiality in the National Cancer Registry. (2007). https://www.ncri.ie/data.cgi/html/confidentialitypolicy.shtml. Accessed March 7, 2023.

35. Kourime M, Bryce J, Jiang J, Nixon R, Rodie M, Ahmed SF. An assessment of the quality of the I-DSD and the I-CAH registries - international registries for rare conditions affecting sex development. Orphanet J Rare Dis. 2017;12(1):56. https://doi.org/10.1186/s13023-017-0603-7.

36. Coi A, Santoro M, Villaverde-Hueso A, Lipucci Di Paola M, Gainotti S, Taruscio D, et al. The quality of Rare Disease registries: evaluation and characterization. Public Health Genomics. 2016;19(2):108–15. https://doi.org/10.1159/000444476.

37. Mordenti M, Boarini M, D'Alessandro F, Pedrini E, Locatelli M, Sangiorgi L. Remodeling an existing rare disease registry to be used in regulatory context: lessons learned and recommendations. Front Pharmacol. 2022;13(no pagination). https://doi.org/10.3389/fphar.2022.966081.

38. Ali SR, Bryce J, Kodra Y, Taruscio D, Persani L, Ahmed SF. The Quality evaluation of Rare Disease Registries-An Assessment of the essential features of a Disease Registry. Int J Environ Res Public Health. 2021;18(22). https://doi.org/10.3390/ijerph182211968.

39. Maruf N, Chanchu G. Planning a rare disease registry (2022). https://orphan-reach.com/planning-a-rare-disease-registry/. Accessed September 29, 2023.

40. Hessl D, Rosselot H, Miller R, Espinal G, Famula J, Sherman SL, et al. The International Fragile X Premutation Registry: building a resource for research and clinical trial readiness. J Med Genet. 2022;59(12):1165–70. https://doi.org/10.1136/jmedgenet-2022-108568.

41. Vitale A, Della Casa F, Lopalco G, Pereira RM, Ruscitti P, Giacomelli R, et al. Development and implementation of the AIDA International Registry for patients with still's disease. Front Med. 2022;9:878797. https://doi.org/10.3389/fmed.2022.878797.

42. Stanimirovic D, Murko E, Battelino T, Groselj U. Development of a pilot rare disease registry: a focus group study of initial steps towards the establishment of a rare disease ecosystem in Slovenia. Orphanet J Rare Dis. 2019;14(1):172. https://doi.org/10.1186/s13023-019-1146-x.

43. Kinsner-Ovaskainen A, Lanzoni M, Garne E, Loane M, Morris J, Neville A, et al. A sustainable solution for the activities of the European network for surveillance of congenital anomalies: EUROCAT as part of the EU platform on Rare diseases Registration. Eur J Med Genet. 2018;61(9):513–7. https://doi.org/10.1016/j.ejmg.2018.03.008.

44. Pericleous M, Kelly C, Schilsky M, Dhawan A, Ala A. Defining and characterising a toolkit for the development of a successful European registry for rare liver diseases: a model for building a rare disease registry. Clin

45. Med J Royal Coll Physicians Lond. 2022;22(4). https://doi.org/10.7861/CLINMED.2021-0725.

45. European Medicines Agency: Guideline on registry-based studies. (2021). https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en-0.pdf. Accessed March 7, 2023.

46. Bellgard MI, Snelling T, McGree JM. RD-RAP: beyond rare disease patient registries, devising a comprehensive data and analytic framework. Orphanet J Rare Dis. 2019;14(1):176. https://doi.org/10.1186/s13023-019-1139-9.

47. Isaacman D, Iliach O, Keefer J, Campion DM, Kelly B. Registries for rare diseases: a foundation for multi-arm, multi-company trials. (2019). https://www.iqvia.com/library/white-papers/registries-for-rare-diseases.

48. Chorostowska-Wynimko J, Wencker M, Horvath I. The importance of effective registries in pulmonary diseases and how to optimize their output. Chronic Resp Dis. 2019;16:1479973119881777. https://doi.org/10.1177/1479973119881777.

49. Liu P, Gong M, Li J, Baynam G, Zhu W, Zhu Y, et al. Innovation in Informatics to Improve Clinical Care and Drug Accessibility for Rare diseases in China. Front Pharmacol. 2021;12:719415. https://doi.org/10.3389/fphar.2021.719415.

50. EUCERD: EUCERD Core Recommendations on Rare Disease Patient Registration and Data Collection. (2013). http://www.rd-action.eu/eucerd/EUCERD_Recommendations/EUCERD_Recommendations_RDRegistryDataCollection_adopted.pdf. Accessed September 28, 2023.

51. Bellgard MI, Macgregor A, Janon F, Harvey A, O'Leary P, Hunter A, et al. A modular approach to disease registry design: successful adoption of an internet-based rare disease registry. Hum Mutat. 2012;33(10):E2356–66. https://doi.org/10.1002/humu.22154.

52. Garcia M, Downs J, Russell A, Wang W. Impact of biobanks on research outcomes in rare diseases: a systematic review. Orphanet J Rare Dis. 2018;13(1):202. https://doi.org/10.1186/s13023-018-0942-z.

53. EURORDIS-NORD-CORD:, EURORDIS-NORD-CORD Joint Declaration of 10 Key Principles for Rare Disease Patient Registries. (2012). https://download2.eurordis.org/documents/pdf/EURORDIS_NORD_CORD_JointDec_Registries_FINAL.pdf. Accessed September 28, 2023.

54. Marques JP, Vaz-Pereira S, Costa J, Marta A, Henriques J, Silva R. Challenges, facilitators and barriers to the adoption and use of a web-based national IRD registry: lessons learned from the IRD-PT registry. Orphanet J Rare Dis. 2022;17(1):323. https://doi.org/10.1186/s13023-022-02489-1.

55. Hageman IC, van der Steeg HJJ, Jenetzky E, Trajanovska M, King SK, de Blaauw I, et al. A Quality Assessment of the ARM-Net Registry Design and Data Collection. J Pediatr Surg. 2023;25:25. https://doi.org/10.1016/j.jpedsurg.2023.02.049.

56. Kaliyaperumal R, Wilkinson MD, Moreno PA, Benis N, Cornet R, Dos Santos Vieira B, et al. Semantic modelling of common data elements for rare disease registries, and a prototype workflow for their deployment over registry data. J Biomedical Semant. 2022;13(1):9. https://doi.org/10.1186/s13326-022-00264-6.

57. European Commission. Principles of the GDPR. https://commission.europa.eu/law/law-topic/data-protection/reform/rules-business-and-organisations/principles-gdpr_en.

58. Office of the Privacy Commissioner of Canada. PIPEDA in brief. https://www.priv.gc.ca/en/privacy-topics/privacy-laws-in-canada/the-personal-information-protection-and-electronic-documents-act-pipeda/pipeda_brief/ (2019).

59. Bettio C, Salsi V, Orsini M, Calanchi E, Magnotta L, Gagliardelli L, et al. The Italian National Registry for FSHD: an enhanced data integration and an analytics framework towards Smart Health Care and Precision Medicine for a rare disease. Orphanet J Rare Dis. 2021;16(1):470. https://doi.org/10.1186/s13023-021-02100-z.

60. U.S. Department of Health and Human Services. : Rare Disease Registry Program, Data Ownership https://registries.ncats.nih.gov/glossary/data-ownership/.

61. Deserno TM, Haak D, Brandenburg V, Deserno V, Classen C, Specht P. Integrated image data and medical record management for rare disease registries. A general framework and its instantiation to theGerman Calciphylaxis Registry. J Digit Imaging. 2014;27(6):702–13. https://doi.org/10.1007/s10278-014-9698-8.

62. Blumenthal S, The, NQRN Registry Maturational Framework: Evaluating the Capability and Use of Clinical Registries. EGEMS, Washington. DC). 2019;7(1):29. https://doi.org/10.5334/egems.278.

63. Lautenschlager R, Kohlmayer F, Prasser F, Kuhn KA. A generic solution for web-based management of pseudonymized data. BMC Med Inf Decis Mak. 2015;15:100. https://doi.org/10.1186/s12911-015-0222-y.

64. DAMA U.K. Working Group: The Six Primary Dimensions For Data Quality Assessment, Defining Data Quality Dimensions. (2013). https://www.sbctc.edu/resources/documents/colleges-staff/commissions-councils/dgc/data-quality-deminsions.pdf. Accessed September 28, 2023.
65. McGlinn K, Rutherford MA, Gisslander K, Hederman L, Little MA, O'Sullivan D. FAIRVASC: a semantic web approach to rare disease registry integration. Comput Biol Med. 2022;145:105313. https://doi.org/10.1016/j.compbiomed.2022.105313.
66. Song P, He J, Li F, Jin C. Innovative measures to combat rare diseases in China: the national rare diseases registry system, larger-scale clinical cohort studies, and studies in combination with precision medicine research. Intractable rare Dis Res. 2017;6(1):1–5. https://doi.org/10.5582/irdr.2017.01003.
67. Sernadela P, Gonzalez-Castro L, Carta C, van der Horst E, Lopes P, Kaliyaperumal R, et al. Linked registries: connecting Rare diseases patient registries through a semantic web layer. Biomed Res Int. 2017;2017:8327980. https://doi.org/10.1155/2017/8327980.
68. U.S. Food and Drug Administration. Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry: Draft Guidance (2021). https://www.regulations.gov/document/FDA-2021-D-1146-0041. Accessed August 15, 2023.
69. Santoro M, Coi A, Di Lipucci M, Bianucci AM, Gainotti S, Mollo E, et al. Rare disease registries classification and characterization: a data mining approach. Public Health Genomics. 2015;18(2):113–22. https://doi.org/10.1159/000369993.
70. Daneshvari S, Youssof S, Kroth PJ. The NIH Office of Rare Diseases Research patient registry Standard: a report from the University of New Mexico's Oculopharyngeal muscular dystrophy patient Registry. AMIA Annual Symp Proc AMIA Symp. 2013;2013:269–77.
71. Rubinstein YR, McInnes P, NIH/NCATS/GRDR. Contemp Clin Trials. 2015;42:78–80. https://doi.org/10.1016/j.cct.2015.03.003. R Common Data Elements: A leading force for standardized data collection.
72. Bellgard MI, Render L, Radochonski M, Hunter A. Second generation registry framework. Source Code Biol Med. 2014;9:14. https://doi.org/10.1186/1751-0473-9-14.
73. Choquet R, Maaroufi M, de Carrara A, Messiaen C, Luigi E, Landais P. J Am Med Inf Association: JAMIA. 2015;22(1):76–85. https://doi.org/10.1136/amia-jnl-2014-002794. A methodology for a minimum data set for rare diseases to support national centers of excellence for healthcare and research.
74. Mullin AP, Corey D, Turner EC, Liwski R, Olson D, Burton J, et al. Standardized data structures in Rare diseases: CDISC user Guides for Duchenne Muscular Dystrophy and Huntington's Disease. Clin Transl Sci. 2021;14(1):214–21. https://doi.org/10.1111/cts.12845.
75. Groenen KHJ, Jacobsen A, Kersloot MG, Dos Santos Vieira B, van Enckevort E, Kaliyaperumal R, et al. The de novo FAIRification process of a registry for vascular anomalies. Orphanet J Rare Dis. 2021;16(1):376. https://doi.org/10.1186/s13023-021-02004-y.
76. Taruscio D, Mollo E, Gainotti S, Posada de la Paz M, Bianchi F, Vittozzi L. Archives public health = Archives belges de sante publique. 2014;72(1):35. https://doi.org/10.1186/2049-3258-72-35. The EPIRARE proposal of a set of indicators and common data elements for the European platform for rare disease registration.
77. Roos M, Lopez Martin E, Wilkinson MD. Adv Exp Med Biol. 2017;1031:165–79. https://doi.org/10.1007/978-3-319-67144-4_9. Preparing Data at the Source to Foster Interoperability across Rare Disease Resources.
78. Ahern S, Sims G, Earnest A, Bell C. Optimism, opportunities, outcomes: the Australian cystic Fibrosis Data Registry. Intern Med J. 2018;48(6):721–3. https://doi.org/10.1111/imj.13807.
79. Maaroufi M, Choquet R, Landais P, Jaulent M-C. Towards data integration automation for the French rare disease registry. AMIA Annual Symposium proceedings AMIA Symposium. 2015;2015:880-5.
80. Bellgard MI, Sleeman MW, Guerrero FD, Fletcher S, Baynam G, Goldblatt J, et al. Rare Disease Research Roadmap: navigating the bioinformatics and translational challenges for improved patient health outcomes. Health Policy Technol. 2014;3(4):325–35. https://doi.org/10.1016/j.hlpt.2014.08.007.
81. Bellgard M, Beroud C, Parkinson K, Harris T, Ayme S, Baynam G, et al. Dispelling myths about rare disease registry system development. Source Code Biol Med. 2013;8(1):21. https://doi.org/10.1186/1751-0473-8-21.
82. Vasseur J, Zieschank A, Gobel J, Schaaf J, Dahmer-Heath M, Konig J, et al. Development of an interactive dashboard for OSSE Rare Disease registries. Stud Health Technol Inform. 2022;293:187–8. https://doi.org/10.3233/SHTI220367.
83. Amselem S, Gueguen S, Weinbach J, Clement A, Landais P. RaDiCo, the French national research program on rare disease cohorts. Orphanet J Rare Dis. 2021;16(1):454. https://doi.org/10.1186/s13023-021-02089-5.
84. Hooshafza S, Mc Quaid L, Stephens G, Flynn R, O'Connor L. Development of a framework to assess the quality of data sources in healthcare settings. J Am Med Inf Assoc. 2022;29(5):944–52. https://doi.org/10.1093/jamia/ocac017.
85. European Reference Network. ERN-RND Registry. https://www.ern-rnd.eu/ern-rnd-registry/#registry-objectives Accessed February 14, 2024.
86. U.S. Food and Drug Administration. Real-World Data: Assessing Registries To Support Regulatory Decision-Making for Drug and Biological Products. Final (2023). https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-registries-support-regulatory-decision-making-drug-and-biological-products. Accessed April 4, 2024.
87. Canada's Drug and Health Technology Agency (cadth): Guidance for Reporting Real-World Evidence. (2023). https://www.cadth.ca/sites/default/files/RWE/MG0020/MG0020-RWE-Guidance-Report-Secured.pdf. Accessed November 16, 2023.
88. National Institute for Health and Care Excellence (NICE): NICE Real-World Evidence Framework. (2022). https://www.nice.org.uk/corporate/ecd9/resources/nice-realworld-evidence-framework-pdf-1124020816837. Accessed November 16, 2023.
89. Government of Canada. Government of Canada improves access to affordable and effective drugs for rare diseases. https://www.canada.ca/en/health-canada/news/2023/03/government-of-canada-improves-access-to-affordable-and-effective-drugs-for-rare-diseases.html (2023).
90. Public Health Agency of Canada: Moving Forward on a Pan-Canadian Health Data Strategy. (2022). https://www.canada.ca/en/public-health/programs/pan-canadian-health-data-strategy.html. Accessed November 16, 2023.
91. European Commission: European Platform on Rare Disease Registration (EU RD Platform). https://eu-rd-platform.jrc.ec.europa.eu/_en Accessed February 9, 024.
92. European Commission. European Rare Disease Registry Infrastructure (ERDRI). https://eu-rd-platform.jrc.ec.europa.eu/erdri-description_en Accessed February 14, 2024.
93. European Commission. Set of Common Data Elements. https://eu-rd-platform.jrc.ec.europa.eu/set-of-common-data-elements_en Accessed February 14, 2024.
94. Abaza H, Kadioglu D, Martin S, Papadopoulou A, Dos Santos Vieira B, Schaefer F, et al. JMIR Med Inf. 2022;10(5):e32158. https://doi.org/10.2196/32158. Domain-Specific Common Data Elements for Rare Disease Registration: Conceptual Approach of a European Joint Initiative Toward Semantic Interoperability in Rare Disease Research.
95. O'Neill J, Tabish H, Welch V, Petticrew M, Pottie K, Clarke M, et al. Applying an equity lens to interventions: using PROGRESS ensures consideration of socially stratifying factors to illuminate inequities in health. J Clin Epidemiol. 2014;67(1):56–64. https://doi.org/10.1016/j.jclinepi.2013.08.005.
96. Oehrlein EM, Schoch S, Burcu M, McBeth JF, Bright J, Pashos CL, et al. Developing patient-centered real-world evidence: emerging methods recommendations from a Consensus process. Value Health. 2023;26(1):28–38. https://doi.org/10.1016/j.jval.2022.04.1738.

Publisher's Note