So You Want to Build Your Disease's First Online Patient Registry: An Educational Guide for Patient Organizations Based on US and European Experience

Paul Wicks[1] · Lindsey Wahlstrom-Edwards[2] · Sam Fillingham[3] · Andrea Downing[4] · Elin Haf Davies[5]

Abstract

Patient registries fulfill a number of key roles for clinicians, researchers, non-profit organizations, payers, and policy makers. They can help the field understand the natural history of a condition, determine the effectiveness of interventions, measure safety, and audit the quality of care provided. Successful registries in cystic fibrosis, Duchenne's muscular dystrophy, and other rare diseases have become a model for accelerating progress. However, the complex tasks required to develop a modern registry can seem overwhelming, particularly for those who are not from a technical background. In this Education article, a team of co-authors from across patient advocacy, technology, privacy, and commercial perspectives who have worked on a number of such projects offer a "Registry 101" primer to help get started. We will outline the promise and potential of patient registries with worked case examples, identify some of the key technical considerations you will need to consider, describe the type of data you might want to collect, consider privacy risks to protect your users, sketch out some of the paths towards long-term financial sustainability we have observed, and conclude with plans to mitigate some of the challenges that can occur and signpost interested readers to further resources. While rapid growth in the digital health market has presented numerous opportunities to those at the beginning of their journey, it is important to start with the long-term goals in mind and to benefit from the learnings of those who have walked this path before.

✉ Paul Wicks
paul@wicksdigitalhealth.com

1   Wicks Digital Health Ltd, Lichfield, UK

2   Sano Genetics, Cambridge, UK

3   PIP-UK, Stockport, UK

4   Light Collective, Eugene, OR, USA

5   Aparito Ltd, Wrexham, UK

Key Points for Decision Makers

Patient organizations are frequently encouraged by third parties to "build a registry" but are offered little guidance on whether that is the right decision, and if so, whether to build, buy, or borrow one from a platform provider.

Patient registries have the potential to serve a variety of stakeholders including patients, researchers, clinicians, pharmaceutical companies, and payers. By starting with the end in mind and identifying the aims and intentions of these stakeholders from the outset, the registry itself will be more useful for all concerned.

Setting up and maintaining a registry involves a range of costs including information technology staff, server costs, data management, and marketing. There are a range of approaches to obtaining initial funding (such as a grant or a consortium of industry sponsors) and maintaining ongoing support (such as cost recovery from academics or fee-for-service approaches).

With the advance of technology, the barriers to building a registry are becoming lower, but the expectations of patients and caregivers are growing higher as they have daily access to social networks, smartphones, and wearable devices. Privacy, interoperability, and the ability to move to another platform in the future are key technical considerations as you plan your activities.

1 Background

Patient registries are a relatively modern invention, dating back some 50 years and defined as "an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more predetermined scientific, clinical, or policy purposes" [1]. A registry differs from a simple "contact database" that might be used to store the personal details and basic demographics of people living with a medical condition for such purposes as mailing lists or fundraising. It is also different from a "study database" that might be developed while trying to answer a single scientific question for a single research study. Finally, it is different from the "forums" or "online support groups" found online that primarily exist to enable peer-peer communication, even though their proper function depends on a database too. Instead, a well-designed registry

has much broader functions, in that it can potentially serve as the foundation for multiple studies from different disciplines that serve many stakeholders [1].

Traditionally, registries have fulfilled a variety of functions for clinicians, researchers, non-profit organizations, payers, and policy makers such as helping to understand the natural history of a poorly understood condition, determining the effectiveness of interventions outside the confines of a randomized controlled trial, measuring safety, and/or measuring quality [1]. More recently though, there has been growing interest in evolving registries from siloed databases purely intended for scientists to study into more dynamic systems that allow for the development of "learning health systems" to benefit many stakeholders [2]. Such progress includes increasing patient and caregiver involvement in governing a registry, connecting registry data to clinical care, and supporting advocacy for non-profit organizations looking to generate evidence [2].

We are a group of authors specializing in digital health research (PW), registry partnerships with patient organizations (LWE), setting up a patient organization's first registry (SF), data governance and privacy issues for patients (AD), and regulations, registry development, and commercialization (EHD). Based on our experience in the USA and Europe, through the course of this education article we aim to provide an initial primer for leaders at patient organizations who might be considering setting up a patient registry for the first time. We will outline the potential benefits of developing a registry, cite some examples for review, outline technical considerations, describe data collection approaches, summarize ethical issues, suggest ways to make a registry commercially sustainable, and outline some of the key privacy issues involved.

2 The Promise and Potential of a Well-Designed Patient Registry

Given that there seem to be so many patient registries today, what are some of the potential benefits that others have experienced? There is opportunity to support four levels of beneficiaries: individuals, communities, organizations, and scientific fields.

At an individual level, a shared experience of all those affected by ill health is uncertainty. Common questions include: What is this thing I have? What will this do to me? What might help me get better [3]? Patients and family members seek a variety of sources to answer these questions including their healthcare providers, the medical/scientific community, and the experiences of other peers like them who have been down a similar path. However, most individuals quickly realize that there are no solid

answers and that whatever information they gather is likely to be biased. Because registries have a scientific orientation towards collecting uniform data, they are the best chance we have to elevate individuals from "I" to a community of "We" in improving our shared understanding of a condition (Fig. 1). Beyond fulfilling their own needs, most patients and caregivers altruistically want their experiences to count for something, to be measured, and to be put to good use so that others might learn from their victories and their mistakes [4]. Being seen, being counted, and being connected are key drivers of value for a community. Whereas once, being part of the "patient community" meant being a formal member of a charitable organization or non-profit organization and attending in-person meetings or paying annual dues to receive newsletters, the changes brought by the Internet have broadened who counts as a member of a community. Increasingly, patient advocates use a variety of social networking platforms and can self-identify with a hashtag in their messages (e.g., #BCCW for Breast Cancer Chat Worldwide), a note of diagnosis on their profile (e.g., "breast cancer survivor"), or by interacting with other stakeholders informally in public or in private. The COVID-19 pandemic accelerated such developments, as shown for example by the Patient-Led Research Collaborative of people living with Long Covid who themselves are also researchers from around the world operating outside a formal organizational structure [5].

At an organizational level, registries give authority and credibility for non-profit organizations to present data-driven insight and business cases. Governance and control over a registry act as powerful convening forces to attract external stakeholders who would like to learn more about a condition, such as pharmaceutical companies, funders, and policy makers. A registry is a tangible asset to attract investment, a rationale for professionalizing, and a mechanism for delivering impact. In collaborating with researchers, a registry lowers the barriers inherent in answering a range of questions. The provision of electronic surveys fielded easily through low-cost online tools expands the pool of research hypotheses that can be tested without needing to stand up their own data collection infrastructure. This can be important to answer questions relating to topics such as the health economic impact of disease [6], which might be important from a policy perspective but rarely attract as much research funding as interventional clinical trials, for instance. Where clinicians are tightly woven into the activity of a registry, it can become possible to conduct quality improvement work to better understand how care delivery and patient outcomes interact, and what can be done better [7].

Once a multi-national field of study and practice becomes large enough, it is not uncommon for there to be multiple (sometimes even competing) registries in a given condition. They might have originated in different geographies, fulfill different objectives, or even been developed as direct counter positioning (i.e., a non-profit version of a for-profit organization's registry, or two different medicines each with their own safety registry). For example, some industry-funded registries only collect data specific to a single product, such as the Hunter Outcome Survey, which
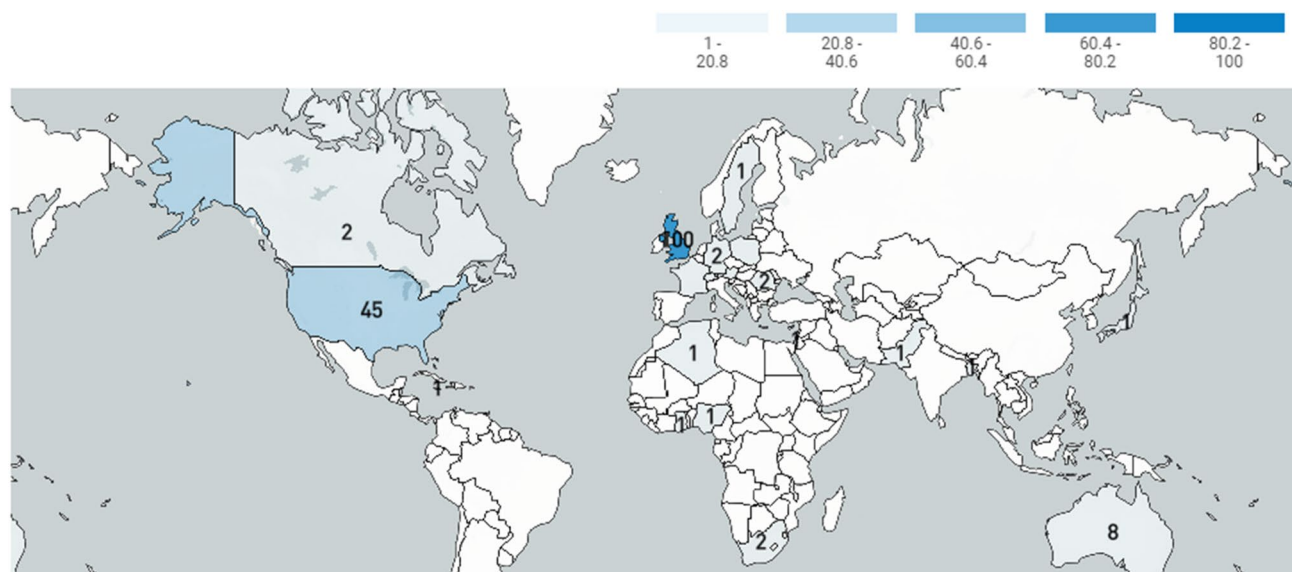
Fig. 1 Example dashboard view for a patient registry participant showing real-time location of registrations. Through an engaged community of younger users on TikTok and Instagram, the registry quickly surpassed its target recruitment of $N = 100$ for the year in just over a month. Courtesy of Poland Syndrome Community Register and Pulse Inforame

excludes potential participants who are not taking the funding manufacturer's product [8]. In some fields, a higher order harmonization group can come together to ensure that similar core data elements are collected by multiple registries and to offer an integrated view [7, 9, 10].

Regulators are a key stakeholder and harmonizer of standards in the health field, and groups such as the European Medicines Agency Cross-Committee Task Force on Patient Registries provide guidance and advice on how to best structure registries [11]. Between 2005 and 2013, the European Medicines Agency requested that over 30 drug manufacturers develop a registry to inform long-term safety and risk-management profiles, particularly in rare or "orphan" diseases [11]. In the USA, the Food and Drug Administration has draft guidance for industry on how to ensure the quality of data captured can support regulatory decisions [12]. Even if your registry is not *primarily* intended to assess safety or the real-world efficacy of products, these regulatory guidelines may still be the standards by which pharmaceutical sponsors (and the scientists that work there) will judge the robustness of your registry data. Table 1 lists some well-regarded registries in the space across a range of size, age, and therapeutic area, each of which provide additional context and examples of best practice in governance, member engagement, and scientific outputs. They all have websites and scientific publications that can serve as templates and inspiration for planning your registry.

3 Starting with the End in Mind

Throughout this Education article, we repeat the importance of starting with the end in mind. The point of a registry is not merely to *collect data* [13]. The point of a registry is to *answer questions*. The types of questions we can answer with a registry might include scientific, clinical, and policy concerns. Scientific questions might include: What sort of people have this disease? Where can we find people who might be eligible to enroll in clinical trials? Could we stratify patients into different forms of the disease, for example, moderate/severe? Clinical questions might include: How are the outcomes of people with this disease changing over time? What are the most important symptoms to manage? How well do drugs and other interventions work in this disease? Finally, policy questions might include: What sort of services and support are people seeking and getting? Is there enough funding being provided to support people with this condition nationally as well as locally? Are people with this disease still able to work, study, and be productive? If not, what is getting in their way? There are several different types of project associated with the word "registry" and Table 2 attempts to differentiate between the most common

examples of terms used in the field, but these are not always applied consistently.

Before you get started, it is worth searching for your condition and the word "registry" in scientific search engines such as PubMed or Google Scholar to see if there are already similar projects underway. It can also be worth searching for any recent "systematic reviews" of your disease field to identify what gaps remain to be filled in the literature. As you start thinking through your objectives for a patient registry, it can be useful to keep a list of these questions, because this will shape what data you collect, how large a sample you need, and how burdensome it might be to take on this endeavor. As you meet with other stakeholders, it is also important to interview them to understand what sort of questions they are hoping to answer with a registry. If you do not have all the context as to why they might want a particular question answered, dive a little deeper. The developers of the European Cystic Fibrosis Society Patient Registry caution against setting the community's expectations too high or trying to capture too much information from the outset that will never be used [13]. For further information, and a helpful checklist, The Genetic Alliance has produced a detailed "Registry Bootcamp" (https://geneticalliance.org/registries/bootcamp) to support your efforts.

For example, once you have set up your registry, a pharmaceutical company might ask where in your country most of the patients are. There might be several reasons they are asking this and knowing why will help shape your approach. If they are at an early stage of R&D and designing a clinical trial, they might be trying to figure out which hospitals they should invite to be trial sites, so it would be important to know how far your users are from major cities. If they have already completed their trials and have recently had their product approved, they might be in the commercial launch phase and trying to figure out where to inform more doctors about their product. Each of these use cases has different data needs, privacy implications, and nuances of interpretation, which we will explore in more detail later.

As you develop your list, remember *you* are a stakeholder too, and for each question that your registry might answer, try and give an honest assessment; *why* do you want to answer this question? *How* will you act differently once you know the answer? What *size* of difference between groups might cause you to act differently? *Who* else will have to be convinced before you can make a decision? What would happen if you *did not* know the answer to a high degree of confidence; would a decision or action still be taken anyway? Too often we see registries set up that will only *describe* the state of a given population, with the intent of generating hypotheses that can be tested later. However, it is worth going through this exercise up front because in many cases

Table 1 Examples of patient registries that range in size, age, and focus

Name (year founded), registered users	Condition focus	Ownership, scope, funding	Key features
Cystic Fibrosis Foundation Patient Registry [13] (1960s), 50k+	Cystic fibrosis and associated cystic fibrosis transmembrane conductance regulator-related disorders	Non-profit ownership, USA, charitable	Clinical integration covering > 95% of clinic visits, browsable quality measure data for care centers, >100 scientific projects
Autoimmune Research Collaborative [14] (ARC, 2016), 50k+	Rheumatoid arthritis, spondyloarthritis, Crohn's disease, ulcerative colitis, multiple sclerosis, vasculitis	Partnership, USA, primarily federal with some charitable and industry funding	Mobile app, electronic patient-reported outcome measures, patient newsletters, harmonization to PCORnet Common Data Model
North American Research Committee on Multiple Sclerosis [15] (NARCOMS, 1996), 41k+	Multiple sclerosis or clinically isolated syndrome	Clinical-Academic Consortium, USA but expanding globally, charitable and industry funding	Surveys, trial recruitment, patient newsletters, de-identified datasets for researchers, 152 peer-reviewed publications, harmonization to MS Metadata Collective
Hunter Outcome Survey [8] (2005), 1k+	Mucopolysaccharidosis type II (Hunter syndrome)	Industry funded, global, associated with use of a specific pharmaceutical product (idursulfase, Elaprase®)	Data from routine clinic visits, emphasis on product safety, natural history, PROs, exploratory data analyses, 12+ peer-reviewed publications
The Duchenne Registry [16] (formerly Duchenne Connect, 2007), 5k+	Duchenne or Becker muscular dystrophy	Non-profit ownership with commercial delivery partners, global (115+ countries), charitably funded	Mobile app, genetic testing, recruitment for 100+ studies, patient-entered data, EHR integration, clinical integration, industry partnership, platform trial, harmonization to TREAT-NMD
International Niemann-Pick Disease Registry [17] (INPDR, 2014), 300+	Acid sphingomyelinase deficiency (types A and B) and Niemann–Pick disease type C	Non-profit ownership with commercial delivery partners, available in 9 countries, EU grant funded with industry support	Web app platform, clinical records for diagnostics, genetics, and outcome data; patient-generated health data by patient and/or caregiver(s)
Gaucher Registry for Development, Innovation & Analysis of Neuronopathic disease [18] (GARDIAN, 2022) N = 30	Type 2 and 3 (neuronopathic) Gaucher disease	Non-profit ownership with commercial delivery partners, available globally in 8 languages, industry funded	Web app platform, surveys, patient-reported diagnosis (verified by professional) and patient/caregiver-reported clinical data, examining PROs, clinical phenotypes, drug use, and impact on caregiver
Poland Syndrome Community Register [19] (2022) N = 227	Poland syndrome	Non-profit ownership with commercial delivery partners, available in English, charitably funded	Web app, surveys, patient/caregiver-reported data, surveys PROs, SF-36, medical history, and demographics recorded

EHR electronic health record, *EU* European Union, *PROs* patient-reported outcomes, *SF-36* Short Form 36 questionnaire

Table 2 Traditional origins for different types of registry (or similar concepts) and common terms. Definitions are not always used consistently over time, geographies, and disease areas

Term	Definition	Notes
Disease/clinical registry (e.g., CDC ALS Registry [20])	Interactive database that collects, organizes, and displays healthcare information with an emphasis on the natural history of disease in a real-world setting. Not restricted to one drug or medical device, can capture "off-label" use	Typically originate at a hospital, group of hospitals, or from a multinational specialist consortium of researchers. More likely to use health record data or physician-reported data rather than allowing patient-generated health data or social networking. Outputs published in journals but slowly
Product registry (Hunter Outcome Survey [8])	Database about patients taking a particular drug product with the objective of gathering outcomes and potentially demonstrating "real-world efficacy" of an intervention	Typically originates from an industry sponsor, health system, or consortium. Data are typically only shared for that sponsor's purposes (e.g., regulatory filing, health technology assessment, partnership with allied doctors)
Safety registry (e.g., Tysabri pregnancy exposure registry [21])	Ongoing database developed to fulfill a pharmacovigilance regulatory obligation, such as tracking cases of an adverse event for the regulators (e.g., EMA or FDA)	The focus may often only be on drug side effects with relatively little data about the patients themselves. Tends to be specific to a region like the USA. Often shut down once the product goes generic or leaves the market, data typically not shared widely
Device registry (e.g., deep brain stimulation for Tourette's syndrome [22])	Focused on tracking the safety and effectiveness of medical devices (e.g., surgical implants) once they are on the market ("post-market surveillance")	Can help determine the need for product upgrades and recalls, may be important in litigation
Registry-based study (e.g., effect of rheumatoid arthritis drugs [23])	Investigation of a research question using the data collection infrastructure or patient population of one or more patient registries	This might be a clinical trial or an observational study run through the registry itself, and might either be a secondary analysis of existing registry data or new primary data
Biobank (e.g., Department of Veteran Affairs Biorepository Brain Bank [24])	The combination of a registry with biological samples such as saliva, blood, or biopsy samples or processed biological data such as DNA sequences	Particularly useful in conditions with causative genetic mutations or to identify potential biomarkers of disease. Because of the expense and complexity, sample sizes tend to be small and localized
Online community/forum (e.g., breast cancer Facebook group [25])	Digital tool that allows for synchronous or asynchronous interaction between users, primarily focused on social interaction and support with minimal objective data features	Typical examples include Facebook groups or threaded bulletin board messaging platforms. Such tools typically are unable to generate quantitative data that can be used for research but may be used for a qualitative text analysis or as a recruitment tool for surveys
Patient-powered research network (e.g., PatientsLikeMe neuromyelitis optica community [26])	Patient-focused registry with online components that may incorporate patient-reported outcome measures, electronic health records, social features, and mobile apps	PPRNs are often run by patients and caregivers themselves, or they have more decision-making influence than traditional registries. From 2009 a US funding body, PCORI, helped develop around 30 such networks

EMA European Medicines Agency, *FDA* Food and Drug Administration, *PCORI* Patient-Centered Outcomes Research Institute, *PPRNs* Patient-Powered Research Networks

the decisions you make at an early stage will become hard to change later on. If the data you gather have no path of becoming information needed to guide decisions, it is just being conducted for curiosity's sake.

4 Technical Considerations

Early registries in the 1990s may once have been simple databases on the hard drive of a spare computer in a clinic. Today, most registries use cloud-based systems that allow for a greater degree of robustness against accidental data loss but also enable a variety of programs and services. For example, an academic clinical researcher might need a data dashboard that displays anonymous aggregate-level data to perform a statistical analysis. A patient might need a mobile app on their smartphone that collects questionnaires on a regular basis and syncs up with their smart watch. The non-profit running the registry may need a series of administrative tools that allow maintenance, software upgrades, or to set the permissions for other categories of user. Each of these users has a slightly different set of requirements and permissions to access or modify the registry data, but they also share some common requirements. They will have certain expectations that the software they use will be responsive, will work on a variety of devices (such as a tablet, smartphone, or desktop web browser), that it will be secure, and that it will conform with their local laws and regulations. Certain types of data, such as brain imaging files or whole genome sequencing data, can be very large and can only be usefully accessed with specialist software. A large amount of data can also be confusing (for any user) and might need to be contextualized by visualizing it as a chart or timeline of some type.

The most important users are the registry participants themselves and they will have their own needs and expectations. Depending on the condition they have, they are also likely to have a range of accessibility requirements such as adjustable font sizes, high-contrast modes for visual contrast issues, and compatibility with screen readers or assistive and augmentative communication devices [27]. For many conditions, it might be important for data from or about a patient to come from one or more caregivers, for example, one or more parents or other caregivers. While there is an increasing expectation that "there's an app for that", smartphone users risk being overwhelmed by the range of notifications, permissions, and settings that need to be managed when controlling their health data. There are also additional privacy concerns when their personal devices host sensitive health data when that device might be shared with other family members. Finally, there is a risk of widening the digital divide when relying only on the latest most expensive versions of hardware that are not available to all participants equally. Throughout the development of any registry, it is important to use a "human centered design" approach and continually gain feedback from the different stakeholders that will power your registry [28].

While a full discussion of registry technology is outside the scope of this article, there are a few high-level considerations to bear in mind (see Table 3). Broadly speaking, most registries are either bespoke (i.e., built just for you by a development team) or on a platform (i.e., a common core of generic features with some optional customization for your purposes from a menu of choices). While bespoke registries can be cheaper to develop initially and give more perceived control, there is a risk of being reliant on a very small team of individuals that know how it works and can make changes. Adding features like a mobile app or clinical trial modules could be too significant an undertaking for the team that originally built a simple web-based data collection tool, and there is a risk of the code becoming out of date as browsers and mobile devices evolve. If key individuals leave, the company closes, or there is a change of control, it might become more challenging to maintain control. Conversely, a platform-based registry may be more expensive upfront but will already cover some of the basic technical considerations and have a more intuitive user experience. They may also offer useful features that allow stakeholders to gain value from the registry much faster. In either case, there is a risk of "vendor lock-in" where it becomes harder to transfer your data and your community from one data environment to another in the event your needs change. A common complaint for either bespoke or platform-based registries is the presence of "bugs" and a months-long waiting period before the implementation of what seem like relatively small upgrade requests.

5 Data Collection Considerations

You have probably noticed that various questionnaires you complete ask you about the same thing in different ways, whether it is the order in which you are asked for a date (e.g., DD/MM/YY vs MM/DD/YYYY) or whether you enter your height in inches or centimeters, everybody seems to do it a little differently. That is annoying in daily life, but it can be crucial when it comes to designing a patient registry. Unlike when you hand a piece of paper over to a human and they can check for errors and ask what exactly you meant (e.g., did you mean to write that you were "5 foot 9" where you have put "59 inches"?), online data entry is typically a one-shot process, so it has to be right the first time.

Even the way we ask users to enter things as simple as their age (or date of birth), sex (and/or gender identity),

Table 3 Summary of approaches to registry development (non-exhaustive)

Approach	Description	Advantages	Disadvantages
Build	Develop IT skills within an organization (e.g. front-end, web app development, community management, back-end database management, data science) and own ongoing responsibility	Total control of development, data, and commercial options. Customization to fit needs, staff development, incorporation of registry at core of operations, can develop at own pace	Time, skills gap, management of multiple IT professionals, lack of skilled oversight, competitive job market for skilled professionals, regulatory/compliance overhead
Buy	Conduct a tendering process for a third party to be paid to develop and maintain a registry under contract	Faster, control, partial customization, working with skilled professionals, up to date with industry standards, fewer staff commitments	Expensive, few independent developer studios have relevant experience, can be sensitive to scope creep, procurement is complex, vendor lock-in
Borrow	Agree to manage registry on an existing platform at low or no cost, typically in exchange for provision of software (and/or a share of revenue)	Fastest option, many decisions already set as default or admin-able options at set-up, leverage experience with similar platforms, benefit from marketing footprint, can be maintained by skeleton crew, not reinventing the wheel	Lack of control and leverage over platform. Difficult to "lift and shift" community if functionality degrades or relationship turns out poorly. Slow to respond to bugs and new feature requests

IT information technology

and location can become complicated very quickly. When it comes to entering dates about things that happened a long time ago many respondents will have to guess (hence a high proportion of dates entered as "January 1"). In general, the harder you make it to enter the right answer the more likely you will encounter errors in the data, or your participants will simply give up. While it can be appealing to consider the participants' doctors entering data on their behalf, in practice this is almost impossible because of a lack of time in the brief clinical encounter or information technology security policy restrictions on hospital computers.

The method of data structuring becomes more critical as you consider with whom the data might be shared once they are collected. If you have a national registry then at some point you might want to compare your data with that from another country. If you think a researcher, a regulator, or a pharma company might want to look at your data, then its quality will be much higher if you enforce some data standards, i.e., guidelines by which data are described and recorded. You do not have to invent these yourself, there are existing standards like Logical Observational Identifier Names and Codes, which describe laboratory tests and their results, or International Classification of Diseases codes; however, these standards do not apply across all types of data that might be collected as part of a registry, and many rare diseases do not yet have an International Classification of Diseases code.

At a minimum, organizations should keep a record as to how data were collected and structured to allow for technical integration and/or further configuration down the road. If your registry has plenty of explanatory text on web pages or uses branching logic, it is best practice to build a "codebook" showing how the data are entered, validated, and what the user sees, preferably with screenshots, and to keep it updated as you make changes to data entry screens. That way if you change something on the website in the future, you can understand why you might be seeing changes or errors in the data. In our earlier example of height, perhaps we started out with an "open text box," but then in a later version we made people choose to enter units as either centimeters or feet and inches. In a further iteration, we might decide to reduce the likelihood of out-of-range data by giving users a dropdown menu that only lists "feasible" heights (bearing a few unusual "edge cases" in mind, such as children, outliers at both ends, or even amputation).

If trying to merge or compare two or more different data sets, then considerable effort might go into "data harmonization," i.e., deciding on which approach to prefer when comparing two datasets. A data scientist can help automate these rules so that you can compare those datasets more rapidly in the future, but it would be better to start with the same standards. In some conditions, an

organization like the International Consortium of Health Outcomes Measurement or a consortium specific to your disease might have already done the work to define a "Core measure set" or "common data elements." This can be a slow and deliberative process that takes place over several years though, so do not be surprised if nobody's done that work yet in your field, but it is worth asking around because there may well be such a project underway.

One group created a comprehensive patient registry software systems checklist called CIPROS [29]. It may not be necessary to go into this level of detail when just starting out, but it could help structure the questions you want to ask of potential registry vendors as you work through your options. The European Cystic Fibrosis Society Patient Registry also published an extensive "lessons learnt" focusing on the collection, use, and improvement of data in their registry in an open-access publication [13].

6 Data Governance, Privacy, and Security

Data governance remains the cornerstone that enables patient registries to thrive. Typically, data are subject to the regulations of the country in which they are collected, with some countries requiring data collected about their citizens to be housed within that country. For those wishing to host a multi-national registry, there are some cloud-based solutions that allow you to specify a "hosting country", or you might narrow eligibility to individuals in a specific country. Similarly, data collected about individuals who live in countries in which the European Union's General Data Protection Regulation is in effect are beholden to the General Data Protection Regulation (in addition to local laws) regardless of the location of the organization collecting the data [30].

One regulation to be aware of when operating within the USA is the Health Insurance Portability and Accountability Act. The Health Insurance Portability and Accountability Act applies to the transfer of protected health information from one covered entity to another, i.e., health plans, healthcare clearing houses, and certain healthcare providers [31]. While the Health Insurance Portability and Accountability Act does not typically apply to a registry or a platform for which an individual shares their data, users may still have specific expectations with regard to the privacy and sharing rights applicable to their data.

Consent governs your permission to collect and store data, and to get in touch with your participants for the purposes of marketing, recontact, or invitation to a specific study. It may be efficient to try and secure "broad consent" for all of these potential use cases when they first register to avoid needing to go back and "re-consent". However, there is growing push back against consenting individuals in this manner [32], with a move towards "dynamic consent", for which individuals are given more control over the use of their data on a case-by-case basis. The consequences of this more specific consent, however, will be more complex administration, reduced interoperability, and potentially lower sample sizes for specific studies.

Truly *informed* consent should ensure that each individual patient understands who the guardian of the data is and how decisions are made about access and use. Some patients' groups may involve a third-party vendor (e.g., contract research organizations) that are for-profit entities that manage the risk of holding and analyzing or reporting on the data. This has value to ensure the smooth operations, processing, and ownership of data within a legal framework, but ultimately patient groups should remain the "data owner". They should have transparent rules and processes in place specifying under which predetermined criteria that anonymized and aggregated data might be shared with third-party organizations such as clinicians and academics for research purposes, regulators and reimbursement agencies for drug evaluation purposes, and pharmaceutical or medical devices companies for drug/device development.

Anticipating these data scenarios will allow for careful consideration of the informed consent and commercial arrangements to be put in place. All such financial arrangements should be made publicly available in a "declaration of interest statement" that is kept up to date. One common approach to ensure good data governance is to create a "data access committee" that convenes several times a year and is responsible for evaluating each data request received. Such a committee should include key opinion leaders in the field to ensure the scientific value of each request, patient advocates to represent the needs of the patients, and independent methodologists. While US and European Union regulators are considering which regulations might enforce such best practice consistently, for the time being, a well-governed data control process remains paramount for maintaining the trust of the community.

Beyond focus on informed consent, running a patient registry will also require you follow laws and standard practices to ensure your data are stored securely. For example, the National Institute of Standards and Technology offers standards on implementing what is called *Zero Trust* as an architecture that considers potential risks of your registry data. These standards also support implementing workflows to ensure your organization is prepared and trained to protect sensitive health data of your registrants [33]. Furthermore, you will need to learn the basics of how to properly encrypt the database of your registry, and ensure there are clearly defined internal control policies to access the data in ways that honor the consent model entrusted by your registrants [34].

7 Building a Research Agenda

Before you develop a registry, it is useful to develop a research agenda. A simple heuristic is something we call "The Rumsfeld Research Agenda". Based on a famous quote [35], the questions that are "knowns knowns" describe your sample and assess how it compares to the scientific literature. If you have a large and relatively unbiased method of recruiting participants, then we might find that the participants in your registry are very similar to those reported in the published scientific literature. For example, in an analysis comparing 10,255 members of the PatientsLikeMe multiple sclerosis registry with 4039 members of a specialist academic center database, the two samples were similar on age, age at onset, disease duration, gender ratio, family history, race, MS subtype, and even education level. However, owing to the large sample sizes, these were statistically significant differences and reflected the fact that PatientsLikeMe members were recruited via social media sites like Facebook, which tend to skew younger and more female [36]. While somewhat *un*interesting that the data were fairly similar to another data source, this is an important cornerstone to understand any bias in the data you are collecting. If you were to skip this step and just start discovering "unknown unknowns" from the outset, then you would face inevitable questions of bias that might result from online methods [37].

Once you have established the representativeness of your sample, the next step is to study "known unknowns" [35]. A patient registry lowers the barriers to answer research questions that are obscure, under-researched, would be challenging to fund, or represent a long-standing gap in the literature. Within the PatientsLikeMe MS community, a survey on the impact of menopause in women with MS on their symptoms was able to quickly recruit $N = 513$ respondents and established that postmenopausal status, surgical menopause, and earlier age at menopause were associated with more severe symptoms [38]. The clinical-scientific collaborators on this study had been interested in the topic for many years but because it crossed multiple disciplines, was in the historically under-funded domain of women's health, and had no direct impact on treatment, the study was otherwise challenging to conduct. This work has now been cited by other peer-reviewed articles over 40 times and informed several follow-up studies and systematic reviews. In this case, the "female bias" in the sample revealed by the earlier study was an advantage, not a limitation.

Finally, your registry can be a jumping-off point for hypothesis generation, innovation, and a rich foundry for "unknown unknowns" [35]. Registries have unexpectedly revealed that cancer drugs worked faster than the pharma

company that made them noticed [39], been the launchpad for patient-led drug trials, [40] spawned dozens of offshoots in other countries [41], connected families to genetic counselors (e.g., https://www.duchenneregistry.org/), and become the basis for auditing the quality of care [42].

8 Commercialization and Sustainability

A registry can only be sustained if it generates revenue or secures larger donations. While grants have historically been a major funding source for starting registries, these are time limited. When the grant comes to an end, then either the data capture or maintenance of the registry comes to an end, or the work becomes reliant on the unpaid (and finite) good will of the host organization. Because this is a risky position, most funders now ask for sustainability plans through strategic partnerships such as with pharmaceutical and biotech partners (see Table 4).

Registry costs to consider include staff time (i.e., management, product management, front end user interface, back end database development, quality assurance testing, infrastructure operations), platform maintenance and upkeep, recruitment support and initiatives, community moderation, marketing, and web hosting. This is without analysis and reporting costs, or efforts to publish and disseminate results. Therefore, there is no truly "free" service. There are, however, a variety of registry providers with various business models tailored to the organization. Some provide a free platform for advocacy organizations and use shared data ownership to help them recover the costs of maintaining the platform. Other platforms, like REDCap, charge organizations a small fee for long-term programs and for support services [43]. Still others charge organizations a licensing fee for use of the platform, but do not require shared data ownership. Depending on the organization's needs, it may be possible to offset the licensing fees through data sharing or a similar agreement. At the time of writing, commercial providers of patient registries include (in alphabetical order and with no implied endorsement) Aparito, ArborMetrix, Clinical Pursuit, CorEvitas, IQVIA, Invitae, Luna, OM1, PatientsLikeMe, Pulse Infoframe, Sano Genetics, Syneos, and Thread Research, amongst others. There are also non-profit organizations such as the National Organization for Rare Disorders IAMRARE® registry program or the Rare-X data platform.

Another method organizations use to secure support for their work is to have a "corporate circle" or other membership scheme through which organizations solicit sponsorship from several relevant for-profit partners through which they maintain the registry. Examples

Table 4 Summary of potential commercial approaches to sustainability (non-exhaustive)

Approach	Description	Advantages	Disadvantages
Initial set-up grant	Multimillion USD institutional funding from e.g., NIH, NIHR, large charities, EU Horizon funding	Significant capital bolus and initial burst of energy, interest, resources, grant writing consultants available for support	Grant writing is arduous, complex coordination, reporting requirements, limited flexibility to change direction, usually goes to academies
Use a third-party platform to host registry	Avoid costs by running registry through a commercial provider who manage the platform	Free at point of development, access to advanced features, continuous upgrades	Limited ability to share in revenue upside, vendor lock-in, lack of control, data ownership
Cost recovery on academic grants	Charge a fee to academic / clinical partners as pass-through costs on their grants when they use your registry data for research	Respectable funding source, proportionate to output, associated with scientific outputs	Unpredictable and low overall chance of success for any single grant, may require upfront work that is not funded, only covering costs or low margin
Single-sponsor pharmaceutical funder	A single sponsor steps in to take on all set-up costs and a fixed window of maintenance in exchange for access/control rights	Access to depth of expertise and network, rapid funding decision, significant capital	May insist on excluding competitors, loss of control, perceptions of bias, lose interest at end of product patent life
Multiple-sponsor "corporate circle"	Two or more sponsors collaborate in a managed consortium for initial funding or access to data and services	Diversified income, multiple perspectives, less perception of bias to a single organization, more funds available	Complex political management, corporate desires can dominate priorities, potential regulatory/compliance issues
Fee for service	Charging market or commercial rates for services such as clinical trial referral, market research surveys, data extracts, and reports	Proportionate to output, opportunity for significant revenue from multiple sources, can fund growth	Risk of survey fatigue, need trained staff to deliver high-quality services, competing with commercial CROs, claiming attribution

EU European Union, *NIH* National Institutes of Health, *NIHR* National Institute for Health and Care Excellence, *CROs* Contract Research Organisations

△ Adis

include the American Association of Kidney Patients and NephCure Kidney. This method of sponsorship may be most relevant for organizations working in one or multiple conditions in which several pharmaceutical or biotech companies have an ongoing interest. Where there is only a single pharmaceutical sponsor responsible for funding, developing, and maintaining a registry, there is a risk that because of their interests and regulatory constraints that data will be restricted only to their products. In the longer term as more therapeutic options emerge, this may lead to a fragmentation of data. In addition, pharmaceutical companies can undergo many changes such as changes in therapeutic focus, the exhaustion of a patent, or corporate changes of control. Therefore, reliance on a single pharmaceutical sponsor presents an additional risk.

Aside from unrestricted financial support, many registries operate on a fee-for-service basis for more transactional services such as data access, advertising, academic research partnerships, and consulting services. It can be helpful to look at other resources in your space such as biobanks or other repositories and ask about their costing structure, which might include set-up fees, data licenses, and a variable fee depending on the number of participants, depth of data, and any additional support services needed billed out at an hourly rate.

Use caution and carefully evaluate the tools that you choose to run your registry. While some platforms use advertising revenue to support their platform, this can be problematic. While a simple business model, platforms may have little control over which ads they serve through an ad network, and the most common ads in the health space are either direct-to-consumer pharmaceutical ads (which may be restricted by global compliance regulations to certain territories) or may be for complementary and alternative medicine approaches with limited evidence of utility. The technology involved in targeting ads may also lead to greater privacy concerns for your users. As you evaluate which tools to use, consider how you negotiate with platforms to ensure the digital rights to data in your community are preserved. One such framework for evaluating and negotiating with platforms is provided by the Light Collective (https://light collective.org/trust/).

Many registries field surveys on behalf of academics, agencies, or pharmaceutical companies, and again it is important to consider the potential burden on your population, the potential for survey fatigue, and whether you will need to assist in the design, improvement, or implementation of a survey. Even a questionnaire assembled by clinical or research experts might benefit from patient expertise, which is a valuable service in itself. Where academics are writing grants, they may not yet have the funding in hand, but you should provide a quote to allocate to their budget. In general, a given grant has only a 5–10%

chance of being accepted, so if the volume of such requests is becoming unmanageable it would be reasonable to either offer only a standard menu of service offerings or to charge for the time involved in making estimates. If pressed, researchers often do have access to patient and public involvement grants or may have discretionary research funds. After all, if 90% of the time your work supporting their grant writing will be in vain, then this may not be the best use of your time in the long term. Finally, there is the potential for generating revenue specifically by offering "consulting services" around engaging your community on behalf of your organization. Many nonprofit organizations give this away freely because they want sponsors to engage in their space, but in many cases, there are layers and layers of agencies, consultancies, data brokers, and advisors, all of whom are being paid. It would be illogical that those working (or volunteering) for a non-profit organization are the only agents in such activity not being paid for their contribution.

Some organizations have successfully built registries that facilitate paid participant engagement in clinical trials and/or market research, such as the COPD Foundation's Patient-Powered Research Network [44]. Quality of engagement is valuable and should be reflected in the fees charged. Sponsors may wish to pay only for those potential participants that are randomized into a trial; however, many well-qualified leads you refer to them will either be rejected for inclusion/exclusion criteria beyond your control or may be lost to follow-up because of issues at the site like responding to enquiries in a timely manner [45]. Therefore, you might wish to structure access on the basis of a flat fee for a messaging campaign, or based on the number of message opens or "click throughs" rather than enrollments, which are beyond your control. This is also important because typically a sponsor will go to many recruitment sources all at once, and there may be duplicates in registrations across the vendors. For example, a potential trial enrollee "A" may be contacted through "ad agency X's" e-mail campaign and then again through "hospital trial site Y's" Facebook ad campaign, but then finally enrolls via your "non-profit registry Z." So, who gets the credit, who can prove it, and who gets paid for participant A's enrollment? Sometimes it is best to win this game by not playing it.

Finally, it is worth bearing in mind that commercial providers are charging above the level of mere cost recovery and that this is widely accepted practice. Ultimately, you cannot run your organization without resources, and you do not need to subsidize organizations who have endowments, revenues, and investors. If for some reason they lose interest in a few years' time, you will still be here, so will your registry, and so will your (rising) costs. If you are good at something, never do it for free.

Table 5 Risks and mitigation strategies

Risk	Description	Possible mitigation strategy
Community engagement	While an organization may be ready to build a registry, it does not necessarily mean the community is ready to engage. Onboarding and maintaining engagement in the community can be a full-time job that organizations do not anticipate. This role is critical, and turnover can impact community engagement. Similarly, communities need to be ready for a registry offering	Conduct market research prior to launching a registry to understand community readiness, key messages that may resonate with the audience to encourage participation. Assess staffing availability and resources available to promote community engagement. As much as possible, share the role of community engagement across multiple team members and/or volunteers to prevent attrition if/when changes in staff occur
Platform delays	Any technical product that is not off-the-shelf can face delays in development, from the build itself through to quality testing and data migration. These delays can be significant and can impact your business plan if not properly accounted for in your timelines	Do not try and launch near holidays. Consider licensing a platform from a third-party vendor, rather than attempting to build and maintain a platform in-house. While on the surface it can seem like a more expensive solution, it often ends up saving money because of the ongoing support and platform maintenance a vendor will provide
Changes in key staff	When building, there are several key staff members who can and should be involved: technical, scientific, community engagement, and fundraising. Changes in these key staff positions can threaten the continuity of the program	Clear and comprehensive documentation of processes, goals, changes in the platform, and changes relevant to data security (i.e., new privacy policies or regulations) is the best way to mitigate the loss of important institutional knowledge due to changes in staff
Shifting internal priorities	Internal targets and priorities can shift with new scientific developments, changes in staff, and external influences. A registry is a long-term commitment requiring leadership buy-in and staff commitment to its maintenance and growth, and therefore can fall victim to shifts in internal short-term priorities	Ensuring the buy-in of all key decision makers prior to a registry launch is critical. A registry platform should also be sufficiently flexible to easily adapt to new areas of focus. For example, once a new treatment is discovered, a registry can be used to match potentially eligible individuals to said treatment. Establishing comprehensive requirements for multiple use cases before a build commences is the best mitigation against shifting priorities
Shifting external priorities	Creating a registry is committing to a long-term commitment that requires funding for platform maintenance, growth, and community engagement. If funding is secured through a single avenue (i.e., industry grants), it can put your registry at risk should the funding source change their funding priorities, e.g., due to a failed drug trial	The primary goal for mitigating shifts in external priorities is to diversify the funding sources. In addition to securing industry grants, individual donations, family foundations, and government grants may be available to help ensure the long-term sustainability of the program. Philanthropic donors tend to prefer starting initiatives over supporting existing initiatives
Researcher engagement	While many researchers are more likely to engage with organizations that have an established registry or community, having a registry does not guarantee researchers will engage with your organization or condition—they may even view you as competition	As they undertake a registry build, some organizations have also begun funding one or more researchers who are interested in their health condition. This ensures the data collected through the registry are used for their intended purpose and that research moves forward in the condition
Changes to data privacy standards	The rollout of the GDPR in 2018 demonstrated the critical importance of understanding the data privacy regulations within each geographic area of operation. Typically, data are beholden to the regulations in which they are collected, and the requirements of each country (or state) vary widely. In addition, privacy regulations can influence the ability of registry owners to collaborate with research partners	Using a third-party registry vendor can help organizations navigate various regulatory environments as qualified vendors will have staff and/or external counsel assisting them to ensure adherence to all standards. Organizations that build their own platform may choose to only allow individuals from specific countries to join a registry and share their data, which helps to manage regulatory requirements, but can hinder research progress. Users can also enter false data about where they are located
Data quality/data analysis	Every data set will have its biases, but when your registry is just starting it is likely to be at its most unrepresentative. Early adopters may not be typical patients, and recruiting through social media may bias younger and more female; "Word of mouth" tends to attract similar users. The biases of your dataset might call conclusions drawn from it into question	Following the "Rumsfeld Research Agenda" described earlier should help you understand, and potentially minimize sources of bias. Try and do research iteratively in small projects first, and reincorporate those learnings into a revised data capture plan. Invite statisticians, epidemiologists, or population health researchers to your advisory board. Thoughtfully plan outreach to under-represented groups

GDPR General Data Protection Regulation

Table 6 Further resources

Name	Description	URL
Genetic Alliance Registry Bootcamp	A free resource from an experienced rare disease non-profit to help organizations and communities set up and sustain their own registry	https://registrybootcamp.org/
Registries for Evaluating Patient Outcomes: A User's Guide: 4th Edition	"The registry bible", 426-page comprehensive US AHRQ Report detailing registry planning, design, operations, and analysis as well as sections on common data elements, standardized outcome measurements, governance, stakeholder engagement, ethics, privacy, informed consent, and real-world evidence	https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/users-guide
European Medicines Agency—Patient Registries	Directory of European Registry resources including searchable inventory of existing registries, stakeholder collaboration activities, reports from registry workshops, and guidelines on registry studies	https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries
Development of Harmonized Outcome Measures for Use in Patient Registries and Clinical Practice: Methods and Lessons Learned	US AHRQ Report on how to develop an Outcome Measures Framework for registries	https://effectivehealthcare.ahrq.gov/sites/default/files/wysiwyg/white-paper-standardized-library.pdf
European Network of Centres for Pharmacoepidemiology and Pharmacovigilance	Searchable inventory of disease registries in EU	https://www.encepp.eu/encepp/search.htm

AHRQ Agency for Healthcare Research and Quality, *EU* European Union

9 Privacy Risks

It is important to consider the very real risks of holding highly sensitive data. The trust of a community takes years to earn, minutes to lose, and can take years to rebuild. In this section, we discuss how to mitigate the risk of community data misuse.

We start with an example of how things can go wrong quickly. Since 2013, the non-profit organization Crisis Text Line has hosted a SMS text and social-media based suicide hotline for people in a mental health crisis to get help. Although the service provided support to millions of users, since 2017, a commercial partnership with the artificial intelligence spin-off company Loris.ai used anonymized datasets of over 200 million messages extracted from the service to optimize the performance of customer service chatbots. Critics questioned whether a 50-paragraph "terms of use" agreement merited proper consent from people in need of urgent help, and the organization's own volunteers were unaware of the data repackaging [46]. To avoid similar issues yourself, it is worth using internal marketing campaigns, surveys, and user interviews to ensure there are "no surprises" with how data are being used.

Decisions beyond your control can incur privacy risks too. Technology platforms you adopt will inevitably want to test new business models, particularly in response to changing privacy policies or leadership, but these may not always benefit the community. For example, many advocacy organizations created Facebook Groups to provide social support, reach an audience, and to grow their communities [47]. However, because such tools prioritize "engagement" they may inadvertently reveal personal information by disclosing membership of a sensitive group such as being a carrier for a disease-associated genetic variant. Their engagement algorithms may also inadvertently promote misinformation [47]. While mainstream social media platforms can support fundraising, connection, and advocacy, they were not built with the same intent or constraints as true registries.

Consider the privacy issues that you create for your community over the long term. Many new or emerging registry tools may appear to be an "easy fix," to reach your community. However the lack of rights to your collective data may quickly sow mistrust and even cause real harm, as shown by patient groups who tried to emerge from using Facebook as a lightweight registry [48]. Consider how a third-party platform may target ads to your community or resell your community's data [49], and consider how sensitive data about your community can be leaked to third parties or data brokers [50].

Regularly put yourself in the shoes of someone joining your registry today. Data have the potential to heal, but it also holds the power to cause harm when in the wrong hands. With great power comes great responsibility. You will need to understand how to make sure the technology and tools you use to implement your strategy are worthy of your trust, and the trust of your community. Consider your rights to the data as you adopt any new technology, and if you are uncertain, it is usually worth investing in legal counsel to help you understand your country's health data privacy and compliance laws. There are additional potential risks to consider presented in Table 5 along with ways to mitigate them.

10 Conclusions

For someone just starting their journey on developing a patient registry, this brief (but broad) education article may appear daunting. There is certainly the potential for greater depth and complexity behind any of the topics we have covered including data, research agendas, governance, privacy, sustainability, and mitigating the types of challenges that may appear, and we would encourage interested readers to explore the references provided. Fortunately, this is now a well-worn path and you do not have to create every aspect of your registry from a blank slate. Increasingly, there are now a range of developers and software providers who can get you started much quicker than in the past, though continuity and interoperability must always be primary considerations when partnering with another organization.

Table 6 details additional resources you can explore in greater depth and use as a starting point for developing your own registry, but the most practical advice can always be found by connecting to other people who have built their own registry in adjacent areas. It might be another pediatric indication like yours, or a different form of cancer, or someone who has developed a registry around a similar drug or device that might apply in your condition but for a different indication. Because technologies, regulation, and funding opportunities are so dynamic, it is important to connect with those who have launched their own registries recently, ideally in the same territory.

We leave the closing words to someone sharing their lived experience, first as someone who has been through this journey and emerged successfully on the other side (author SF, Electronic Supplementary Material), and finally from a participant themselves: "The register means a huge deal to me, as a parent … it would be so incredible to have this data coming from *our community* that *we* can then take to the (health service), to the professions, to the organizations and say "this is what *we know* to be true about this condition".

Declarations

References

1. Gliklich RE, Leavy MB, Dreyer NA. Registries for evaluating patient outcomes: a user's guide. Fourth edition. Agency for Healthcare Research and Quality (AHRQ); 2020. https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/users-guide. Accessed 9 Mar 2023.

2. Nelson EC, Dixon-Woods M, Batalden PB, Homa K, Van Citters AD, Morgan TS, et al. Patient focused registries can improve health, care, and science. BMJ. 2016. https://doi.org/10.1136/bmj.i3319.

3. Okun S, Goodwin K. Building a learning health community: by the people, for the people. Learn Health Syst. 2017;1: e10028.

4. Bradley M, Braverman J, Harrington M, Wicks P. Patients' motivations and interest in research: characteristics of volunteers for patient-led projects on PatientsLikeMe. Res Involv Engagem. 2016;2:33.

5. Davis HE, Assaf GS, McCorkell L, Wei H, Low RJ, Reem Y, et al. Characterizing long COVID in an international cohort: 7 months of symptoms and their impact. eClinicalMedicine. 2021;38: 101019.

6. Landfeldt E, Lindgren P, Bell CF, Schmitt C, Guglieri M, Straub V, et al. The burden of Duchenne muscular dystrophy: an international, cross-sectional study. Neurology. 2014;83:529–36.

7. European Medicines Agency. Patient Registries Workshop, 28 October 2016. 2017. https://www.ema.europa.eu/en/documents/report/report-patient-registries-workshop_en.pdf. Accessed 9 Mar 2023.

8. Muenzer J, Jones SA, Tylki-Szymańska A, Harmatz P, Mendelsohn NJ, Guffon N, et al. Ten years of the Hunter Outcome Survey (HOS): insights, achievements, and lessons learned from a global patient registry. Orphanet J Rare Dis. 2017;12:82.

9. TREAT-NMD Neuromuscular Network. TREAT-NMD diseases: data and Information. 2022. https://treat-nmd.org/what-we-do/data-and-information/treat-nmd-diseases/. Accessed 9 Mar 2023.

10. Maelstrom. Multiple Sclerosis Metadata Collective: a collaborative effort of North American observational studies in multiple sclerosis. 2022. https://www.maelstrom-research.org/network/msmdc. Accessed 24 Oct 2022.

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

12. US Food and Drug Administration. Assessing registries to support regulatory decision-making for drug and biological products guidance for industry: draft guidance. 2021. https://www.fda.gov/media/154449/download. Accessed 9 Mar 2023.

13. Viviani L, Zolin A, Mehta A, Olesen HV. The European Cystic Fibrosis Society Patient Registry: valuable lessons learned on how to sustain a disease registry. Orphanet J Rare Dis. 2014;9:81.

14. Nowell WB, Merkel PA, McBurney RN, Young K, Venkatachalam S, Shaw DG, et al. Patient-powered research networks of the Autoimmune Research Collaborative: rationale, capacity, and future directions. Patient. 2021;14:699–710.

15. Marrie RA, Cutter GR, Fox RJ, Vollmer T, Tyry T, Salter A. NARCOMS and other registries in multiple sclerosis: issues and insights. The Consortium of Multiple Sclerosis Centers. Int J MS Care. 2021;23:276–84.

16. Rangel V, Martin AS, Peay HL. DuchenneConnect registry report. PLoS Curr. 2012;4:1309.

17. Bolton SC, Soran V, Marfa MP, Imrie J, Gissen P, Jahnova H, et al. Clinical disease characteristics of patients with Niemann-Pick disease type C: findings from the International Niemann-Pick Disease Registry (INPDR). Orphanet J Rare Dis. 2022;17:51.

18. Collin-Histed T, Milce J, Reed S, Braham-Chaouche L, Jaffe D, Revel-Vilk S, et al. SA73 addressing unmet needs of patients with neuronopathic Gaucher disease type 2 and type 3: creation of the GARDIAN Patient Registry. Value Health. 2022;25:S497.

19. PIP UK. Poland Syndrome Community Register. https://pip-uk.org/poland-syndrome-community-register. Accessed 25 Jan 2023.

20. Rechtman L, Brenner S, Wright M, Ritsick M, Rahman F, Han M, et al. Impact of the National Amyotrophic Lateral Sclerosis Registry: analysis of registry-funded research. Ann Clin Transl Neurol. 2022;9:1692–701.

21. Friend S, Richman S, Bloomgren G, Cristiano LM, Wenten M. Evaluation of pregnancy outcomes from the Tysabri® (natalizumab) pregnancy exposure registry: a global, observational, follow-up study. BMC Neurol. 2016;16:150.

22. Deeb W, Rossi PJ, Porta M, Visser-Vandewalle V, Servello D, Silburn P, et al. The international deep brain stimulation registry and database for Gilles de la Tourette syndrome: how does it work? Front Neurosci. 2016. https://doi.org/10.3389/fnins.2016.00170/abstract.

23. Padula AS, Pappas DA, Fiore S, Blachley TS, Ford K, Emeanuru K, et al. The effect of targeted rheumatoid arthritis therapeutics on systemic inflammation and anemia: analysis of data from the CorEvitas RA registry. Arthritis Res Ther. 2022;24:276.

24. Brady CB, Trevor KT, Stein TD, Deykin EY, Perkins SD, Averill JG, et al. The Department of Veterans Affairs Biorepository Brain Bank: a national resource for amyotrophic lateral sclerosis research. Amyotrophic Lateral Sclerosis and Frontotemporal Degeneration. 2013;14:591–7.

25. Kashian N, Jacobson S. Factors of engagement and patient-reported outcomes in a stage IV breast cancer Facebook group. Health Commun. 2020;35:75–82.

26. Eaneff S, Wang V, Hanger M, Levy M, Mealy MA, Brandt AU, et al. Patient perspectives on neuromyelitis optica spectrum disorders: data from the PatientsLikeMe online community. Multiple Scler Relat Disord. 2017;17:116–22.

27. Lazar J, Goldstein DF, Taylor A. Ensuring digital accessibility through process and policy. Amsterdam: Morgan Kaufmann/Elsevier; 2015.

28. Goodwin K. Designing for the digital age: how to create human-centered products and services. Indianapolis: Wiley Publications; 2009.

29. Lindoerfer D, Mansmann U. Data for the elaboration of the CIPROS checklist with items for a patient registry software system: examples and explanations. Data Brief. 2017;14:494–7.

30. Voigt P, Von dem Bussche A. The eu general data protection regulation (GDPR): a practical guide, vol 10. 1st ed. Cham: Springer International Publishing; 2017. p. 10–5555.

31. US Department of Health and Human Services National Institute of Health. To whom does the privacy rule apply and whom will it affect? 2022. https://privacyruleandresearch.nih.gov/pr_06.asp#:~:text=Covered%20entities%20are%20defined%20in,which%20HHS%20has%20adopted%20standards. Accessed 24 Oct 2022.

32. Steinsbekk KS, KåreMyskja B, Solberg B. Broad consent versus dynamic consent in biobank research: is passive participation an ethical problem? Eur J Hum Genet. 2013;21:897–902.

33. Rose S. Planning for a zero trust architecture: a planning guide for Federal administrators. Gaithersburg (MD): National Institute of Standards and Technology; 2022: Report No.: NIST CSWP 20. https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.20.pdf. Accessed 9 Mar 2023.

34. Stine K, Dang Q. Encryption basics. J Am Health Inform Manage Assoc. 2011;82:44–6.

35. Wikipedia. There are known knowns. 2022. https://en.wikipedia.org/wiki/There_are_known_knowns. Accessed 9 Mar 2023.

36. Bove R, Secor E, Healy BC, Musallam A, Vaughan T, Glanz BI, et al. Evaluation of an online platform for multiple sclerosis research: patient description, validation of severity scale, and exploration of BMI effects on disease course. PLoS ONE. 2013;8: e59707.

37. Goodson N, Wicks P, Morgan J, Hashem L, Callinan S, Reites J. Opportunities and counterintuitive challenges for decentralized clinical trials to broaden participant inclusion. NPJ Digit Med. 2022;5:58.

38. Bove R, Vaughan T, Chitnis T, Wicks P, De Jager PL. Women's experiences of menopause in an online MS cohort: a case series. Mult Scler Relat Disord. 2016;9:56–9.

39. Michelucci P, editor. Handbook of human computation. New York: Springer; 2013.

40. Wicks P, Vaughan T, Heywood J. Subjects no more: what happens when trial participants realize they hold the power? BMJ. 2014;348: g368.

41. Sheppard DN. The European cystic fibrosis patient registry: the power of sharing data. J Cystic Fibrosis. 2010;9:S1-2.

42. Schechter MS, Fink AK, Homa K, Goss CH. The Cystic Fibrosis Foundation Patient Registry as a tool for use in quality improvement. BMJ Qual Saf. 2014;23:i9-14.

43. University of Florda Clinical and Translational Science Institute. REDCap pricing. 2022. https://www.ctsi.ufl.edu/research/study-design-and-analysis/redcap/redcap-pricing/. Accessed 24 Oct 2022.

44. Thomashow B, Walsh J, Malanga E. The COPD Foundation: celebrating a decade of progress and looking ahead to a cure. J COPD F. 2014;1:4–16.

45. Bedlack RS, Wicks P, Heywood J, Kasarskis E. Modifiable barriers to enrollment in American ALS research studies. Amyotrophic Lateral Scler. 2010;11:502–7.

46. Politico. Suicide hotline shares data with for-profit spinoff, raising ethical questions. 2022. https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617. Accessed 9 Mar 2023.

47. Zhang S. Facebook groups as therapy. 2018. https://www.theatlantic.com/technology/archive/2018/10/facebook-emotional-support-groups/572941/. Accessed 24 Oct 2022.

48. Prior R. This breast cancer advocate says she discovered a Facebook flaw that put the health data of millions at risk. 2020. https://edition.cnn.com/2020/02/29/health/andrea-downing-facebook-data-breach-wellness-trnd/index.html. Accessed 24 Oct 2022.

49. Ohlheiser A. Facebook is bombarding cancer patients with ads for unproven treatments. 2022. https://www.technologyreview.com/2022/06/27/1054784/facebook-meta-cancer-treatment-ads-misinformation/. Accessed 9 Mar 2023.

50. Feathers T, Fondrie-Teitler S, Waller A, Mattu S. Facebook is receiving sensitive medical information from hospital websites. 2022. https://themarkup.org/pixel-hunt/2022/06/16/facebook-is-receiving-sensitive-medical-information-from-hospital-websites. Accessed 24 Oct 2022.