Chapter

International Data Sharing and Rare Disease: The Importance of Ethics and Patient Involvement

Adrian Thorogood

Abstract

Improving our understanding of rare disease and developing new therapies can only succeed through global collaboration. Whole genome sequencing is increasingly being deployed to diagnose rare disease, and can be combined with machine-learning tools that analyze patient photos to identify phenotypes. Clinical interpretation of genomes and phenotypic data in rare disease depends on sharing individual patient data internationally. Data sharing is essential in rare disease contexts, to support the diagnosis of patients, recruitment into trials, the development of precision diagnostics and therapies, and clinical trial transparency. The sharing of rich molecular and phenotypic data presents privacy risks for rare disease patients, though many want to see their data made available to improve their care and advance research. Informed consent, access governance, and access technologies are important to realize the benefits of data sharing while mitigating risks. Rare disease patients should be involved in the design of data sharing governance to ensure it responds to their particular needs and preferences.

Keywords: rare disease, data sharing, law, ethics, consent, privacy, patient involvement

1. Introduction

There is great interest in adopting data-intensive approaches as part of both care pathways and research in rare disease contexts. Indeed, many rare diseases "have no treatments, are incurable, and have a devastating impact on patients and their families" [1]. One of the first areas where whole-genome sequencing is already demonstrating clinical utility is in helping to provide a genetic diagnosis of individuals with rare disease [2]. Whole-genome sequencing can be a powerful tool to resolve diagnoses for patients with rare disease. Receiving a timely and accurate diagnosis can have a number of direct benefits for patients, "enabling a better understanding of their prognosis, more personalized treatment and tailored management and surveillance" [3]. An ethics report from Canada's health technology assessment body, CADTH, recently concluded that genome sequencing could be effective for patients with unexplained developmental disabilities and multiple congenital abnormalities, if responsibly administered [4].

Data-intensive medicine is powered by data sharing. Data sharing practice and policy has long been a hallmark of genomic research. Many health research funders

and journals now require researchers to deposit sequence data in repositories or otherwise make data available to the broader research community. "Data sharing enables researchers to rigorously test the validity of research findings, strengthen analyses through combined datasets, reuse hard-to-generate data, and explore new frontiers of discovery" [5]. Data sharing in health care contexts is also growing in importance. The American College of Medical Genetics and Genomics (ACMG), for example, "advocates for extensive sharing of laboratory and clinical data from individuals who have undergone genomic testing" [6]. Data sharing is expected to have a range of benefits, including improving the diagnoses of other patients, informing the development of diagnostic approaches and tools, and powering research. Some have even argued that sharing minimal information about variant interpretations should be the standard of care in genetics [7].

Data sharing is of particular importance in rare disease contexts. Making a diagnosis is often dependent on many forms of data sharing. Databases of population genetic variation are needed as a reference to help filter out benign variants from test results. Comparison of family trios can help filter candidate disease-causing variants even further. Making a genetic diagnosis available through publications or public genetic variant databases can offer confirmation and can inform and accelerate the diagnosis of future patients. Data-intensive approaches are not limited to genomic data. Facial recognition technologies such as Face2Gene can inform diagnosis based on images of facial morphology [8]. Because the meaning of all this data is not fully understood, data-intensive medicine for rare disease depends on adoption of a learning health system approach. In learning health systems, rich data are generated as part of routine clinical care and are subsequently made available for quality improvement and research. Data sharing between rare disease clinicians, laboratories and scientists can help to refine interpretive techniques and analysis pipelines. Images and videos of facial morphology can be used to train machine-learning algorithms and improve diagnostic tools [9].

The impetus to make genomic and health-related data collected as part of routine clinical care available for research is stronger in the rare disease context, where there are numerous barriers and limited incentives. These data could also serve as a rich resource for natural history studies to better understand the progression of rare diseases, for biomarker discovery, as registries to recruit patients into precision clinical trials, and as a resource for ongoing surveillance of the clinical and cost effectiveness of rare disease therapies.

Biobanking and biobank networks are essential infrastructure for genomic research, which require collection and analysis of biospecimens [10]. Biobanks are organized collections of samples and associated data. Samples have to be collected, stored, and shared following scientific and technical standards in order to be comparable. Biobanks must meet standards of quality and size to be scientifically valuable. Biobank networks are established to enable aggregation of samples and data from geographically disperse patients. This is essential in the rare disease context. Fostering standardization is a complicated challenge for biobank networks. In the era of Big Data, the value of biobanks increasingly lies in their datafication. Datafication includes: (1) the collection of rich associated demographic, health and clinical information about patients, and (2) the analysis of samples to generate molecular, imaging or other forms of biological data. Of course, the quality and compatibility of data collected or generated is also essential to aggregation of data across biobank networks. In order to attract researchers and additional resources to understudied areas like rare disease, biobank networks and datafication are key. This article will focus primarily on data sharing, but the reader should keep in mind that the generation of data from biological samples is an essential step, one that is organizationally and scientifically non-trivial.

Little will be achieved for rare disease patients without collaboration and international data sharing. No single institution, laboratory, or even country is likely to encounter a sufficient number and diversity of patients with a given rare disease to be able to advance research alone. In the next section, I review some major data-driven initiatives to improve rare disease care and research. These learning health system approaches, powered by international data sharing, are essential to deliver data-intensive medicine for rare disease. International data sharing does raise concerns about the privacy of patients with rare disease. I discuss issues of privacy and consent in data-intensive rare disease medicine and research. However, it is important to note that many rare disease patients want to make their data available to improve their care and to support research. It is therefore important that patients are involved in the development and implementation of data sharing governance to ensure the benefits of data sharing are achieved while managing risks to patient privacy.

2. Rare disease data sharing initiatives

This section discusses the importance of data sharing for rare disease, across the research lifecycle, from diagnosis and basic research, to clinical trial transparency, to health technology assessment.

2.1 Diagnosis and drug discovery

A number of national and international initiatives have emerged to demonstrate the potential of data to improve rare disease patient care and to accelerate research. All of these initiatives seek to adopt data-intensive approaches to accelerate rare disease diagnosis. They adopt learning health system strategies, which involve collecting rich data as part of routine clinical care and making these data available for research to improve diagnostics and therapies. Finally, the initiatives all recognize the importance of international data sharing in the rare disease context.

The Genomics England 100,000 Genomes Project has "committed to sequencing 100,000 whole human genomes, from 70,000 patients, by the end of 2018" [11], with a focus on rare and infectious diseases [12]. This project will facilitate the introduction of genomic medicine in NHS care while contributing to the personalization of its medicine [11]. Clinicians are hoping to achieve earlier diagnoses and develop more effective treatments with this data [13]. Researchers also hope to gain a better understanding of cancer.

Genome Canada has proposed a national, clinical genomics project, which aims to advance precision medicine for all Canadians, with an initial pilot focused on rare disease [14]. The proposal is to introduce genomic testing as part of clinical care. The data will then be made available as a research platform. The vision is to establish a national cohort, perhaps through a federation of provincial datasets.

The European Joint Program on Rare Diseases (EJP RD) brings 130 institutions together across 27 EU Member States as well as Canada, Armenia, Georgia, Israel, Norway, Serbia, Switzerland and Turkey, to accomplish its two main goals: [15].

1. "To ***improve*** the integration, the efficacy, the production and the social impact of research on [rare disease] through the development, demonstration and promotion of Europe/world-wide sharing of research and clinical data, materials, processes, knowledge and know-how

2. To ***implement*** and further *develop* an efficient model of financial support for all types of research on [rare disease] (fundamental, clinical, epidemiological,

social, economic, health service) coupled with accelerated exploitation of research results for benefit of patients" [15].

To achieve its objectives, the EJP-RD has developed a five pillars structure subdivided in various themes and activities such as Joint Transnational Calls for collaborative research projects; a common virtual platform for discoverable data and resources for rare disease research; capacity building and training of patients and researchers in rare disease research and processes, all to accelerate the validation, use and development of innovative methodologies tailored for clinical trials in rare diseases [15].

Through the Breaking Barriers to Health Data Project, the World Economic Forum is "partnering with genomics institutes in the United Kingdom, the United States, Canada and Australia" [16] to pilot a governance framework "to support the effective and responsible use of federated data systems to advance rare disease diagnostic and treatment-related research" [16]. Federated data systems enable researchers to query a distributed network of secure databases. The individual-patient data remains hidden in each of the secure nodes. This pilot project aims to demonstrate a proof-of-concept for federated data systems, accompanied with an economic analysis and a scalable governance framework [16].

An example of a commercial initiative to overcome the geographic barriers to rare disease research is the start-up RDMD [17]. This company aims to generate a rich, regulatory-grade biobank, database, and registry of patients with rare disease from across the United States (US) and internationally. The start-up leverages the rights of patients in the US and in other countries to request access to their health records and biospecimens for onward transfer to RDMD. RDMD then looks to enter into partnerships with pharmaceutical companies to accelerate their research into rare disease therapies. Patients are provided with access to their aggregated and structured medical record through an app.

2.2 Clinical trial transparency

Improving the transparency of clinical trials has been an important public health priority for regulators and policy-makers in recent years. Clinical trial transparency encompasses the registration of clinical trials before recruitment, the timely dissemination of results—whether positive or negative, and the sharing of individual patient data supporting those results [18]. Sharing of individual patient data enables reproducibility studies to confirm the validity of results, and facilitates meta-analyses. Transparency can also accelerate research and reduce duplicative trials that waste resources and expose participants to unnecessary risks. Regulators increasingly publish the clinical data submitted by pharmaceutical companies seeking market approval [19]. Some sponsors also proactively make individual patient data available. There are now several data sharing platforms that facilitate clinical trial data sharing, including Yale Open Data Access project (YODA) [20], ClinicalStudyDataRequest [21], and Vivli [22].

Ensuring that the results of clinical trials as well as the underlying data are made available is perhaps more important for rare disease clinical trials. Regulators sometimes allow more flexibility and accept greater clinical uncertainty to accelerate approvals of drugs for rare diseases with high unmet need. This is because of the "unique challenges that hinder efficient and effective traditional clinical trials, including low patient numbers, limited understanding of disease pathology and progression, variability in disease presentation, and a lack of established endpoints" [1]. Where there is greater uncertainty over the meaning of research data, there is a greater need for transparency to support regulators, prescribing physicians, and

patients. Sharing individual patient data does raise concerns about patient privacy, discussed below. The tension between transparency and privacy, however, tends to be overstated, as benefits can be promoted and risks can be reduced through governance mechanisms. Moreover, rare disease patients are generally supportive of greater transparency, as long as their privacy is protected, appropriate steps are taken to seek their consent, and patient groups are involved in the design of data sharing governance.

2.3 Access to medicines

Even where approved medicines are available for rare disease, an additional hurdle is convincing health technology assessment bodies that these medicines—which are often very expensive per patient—are cost-effective [23]. There is often significant uncertainty over the clinical and economic value offered by rare disease therapies, in part because of the limits to generating clinical evidence in small patient populations. One potential solution to accelerate patient access is through managed access programs. Where countries offer these programs, drugs may be given early approval despite some uncertainty over value, under the condition of ongoing collection of data to fill in evidentiary gaps. Real world evidence is collected through post-market surveillance to confirm the drug delivers value. Post-market surveillance, however, is challenging and requires effective data sharing strategies and infrastructure. Moreover, it is important to involve patients in decisions to approve drugs where there is greater uncertainty over benefits and risks. Patient can also be engaged in establishing the conditions under which a drug would meet or fail to meet the conditions of a managed access agreement. Indeed, patients are increasingly involved in health technology assessment to ensure that the drugs are delivering the clinical, economic, and personal value that matters to them [24]. Given the diverse burdens of disease on rare disease patients and their caregivers, they have important perspectives on the true value that can be delivered by new therapies.

3. Privacy

Data-intensive medicine, and the research, biobanking, and data sharing that necessarily accompany it, all raise privacy concerns for patients. In the Big Data era, increasingly rich data are being generated as part of clinical care and research protocols. Traditionally, privacy in research was primarily protected by removing or separating identifiers from research data. Rich, multi-dimensional health data can no longer be definitively de-identified. Genomic data for example is rich, unique to the individual, stable over time, and shared across families. It also contains potentially sensitive information about the health predispositions of individuals and their families. Genomic data therefore raise particular concerns about the limits of de-identification [25]. But the problem is broader than just genomic data. A recent study also showed that 99.98% of American can be reidentified from a database with less than 15 demographic attributes [26]. Re-identification is increasingly seen as an inherent risk in research. This risk increases as the dimensionality of data increases, as more publicly available data becomes available, and as new statistical re-identification tools emerge. If patients are re-identified, sensitive information about their health may be disclosed to unauthorized third parties, including employers, insurers, and family members, and may be used to discriminate against or stigmatize the individual or their family. Sharing patient data with clinicians and researchers around the world may heighten concerns over privacy. Where data are

copied and distributed to many different parties, there is a greater potential for a breach of confidentiality or security, and lower confidence that the breach will be identified and rectified.

Rare disease patients may face a greater risk of re-identification or subsequent harm. Rare disease patients may be easier to single-out in a dataset, given their unique genotypes and phenotypes, and the small number of participants in a study. Rich data is often collected or generated about rare disease patients, such as whole genome sequences and pictures and videos of their phenotypes. In order to match similar patients to inform a diagnosis, or to conduct a study with an acceptable sample size, information about rare disease patients must often necessarily be shared beyond institutions and national borders. Moreover, in part because of institutional and geographical barriers to care and participation in research, many rare disease patients share rich health information about themselves online with patient support groups or researchers. In fact, there are numerous academic and commercial research efforts that enable remote participation of rare disease patients to overcome geographic barriers [27]. The public availability of patient information could potentially increase the risk of re-identification in research datasets.

At the same time, many patients with rare disease see the important clinical and scientific value of data sharing and are willing to participate if research involves appropriate consent processes, safeguards, and patient involvement. A number of solutions have evolved to reduce the tension between privacy and openness. The first solution is to develop more transparent consents about how data are shared. This is recommended by the Global Alliance for Genomics and Health (GA4GH) Consent Policy [28]. Consent is discussed in greater detail in the next section. The second solution is through safeguards and governance, including robust de-identification, security protections, and access controls. Responsible data governance aims to maximize uses of data that benefit science and society, minimize risks to data subjects, and strike a proportionate balance where these interests come into conflict [29]. Risks of data breaches or misuse when sharing data can be significantly reduced through governance mechanisms including due diligence review of access requests by an expert committee, data access agreements that protect participant privacy, and ongoing monitoring of data use. Sharing data within secure cloud environments can enhance security and accountability by limiting the distribution of copies of datasets. Federated network technologies now allow researchers to submit search queries or run research analyses across multiple secure patient databases, without ever having to access the patient results. The World Economic Forum is exploring such an approach specifically for international rare disease research (see above). A third solution, also discussed below, is greater patient involvement in the design or research and data sharing governance, to ensure their input on priorities and the balancing of risks and benefits under uncertainty.

There is also a risk of too much privacy protection in the rare disease context. Data privacy laws are tightening globally in response to concerns over commercial and law enforcement surveillance practices. Europe's *General Data Protection Regulation* (GDPR) is now in force, and California will soon be introducing its own comprehensive consumer data privacy regime [30]. The GDPR imposes stricter, more formal procedural and security safeguards for the protection of personal data, particularly for special categories of data, for example, health and genetic. It also imposes higher consent standards with regards to the purposes of processing, and transfers between organizations and across borders. Different national and institutional interpretations of the GDPR have hampered international health research collaborations [31]. Formal legal safeguards and strict transparency requirements leave organizations with less flexibility to share samples and data about rare disease

patients, especially internationally, even where researchers seek explicit patient consent and/or patient involvement in data sharing governance.

4. Consent

This section considers three issues concerning consent in data-intensive medicine for rare disease. The first concern is that it is impossible to fully specify all the potential users and uses of patient data at the time of collection. A common solution when seeking consent to research and sharing of samples and data is seeking broad consent to future not-fully-specified uses accompanied with ongoing governance. A second issue is re-use or sharing of legacy collections of samples and data that have significant scientific and societal value but where the original consent is absent or is silent about key matters. A third issue is consent in the pediatric context, which raises special concerns about capacity, protection from harm and exploitation, inclusion, and shared decision-making. Arguably, the tension between promoting science and respecting individual autonomy is greater in the rare disease context.

4.1 Broad consent

The International Rare Diseases Research Consortium (IRDiRC) in collaboration with the GA4GH has shared template consent clauses for rare disease research [32]. These clauses emphasize some of the special characteristics of rare disease research, such as the collection of photos and videos of patient phenotypes, the participation of and feedback of health findings to family members, as well as the imperative of international data sharing to support both research as well as to match patients to inform diagnoses. This initiative also demonstrates the importance of engaging patients in the development of research governance and consents.

Biobanking and data sharing aim to make samples and data available for research that cannot be fully specified at the time of recruitment and collection. This present risks to patient autonomy: how much information can and should be provided at the time of consent? What kinds of meaningful choices can and should be offered to patients about who can access their samples and data? As samples and data are typically stored for long periods of time, can and should patients be able to withdraw consent or change their preferences over time? Broad consent—consent to not-fully-specified research uses coupled with ongoing governance—has been adopted in many research contexts internationally, is now expressly permitted under the US Common Rule and recognized under the EU GDPR (rec 33) [33].

Especially where samples and data are collected in clinical care contexts, there is concern that sharing those samples and data for research may be done coercively, or that patients may have limited knowledge or comprehension. Moreover, rare disease patients may see data sharing as a necessity for receiving a diagnosis, or to advance research on a cure, and thus may feel compelled to forgo their privacy. Where consents cover a broad set of purposes, this can be seen as coercively tying purposes together, unless patients are given granular choices. But these kinds of arguments can result in inefficient sharing and use of data that precludes effective care and research, and that contradicts the wishes of many rare disease patients. Broad consent may be especially important in rare disease, given the scarcity of data and the risk of losing that data if every subsequent use is subject to re-consent. There is also an argument that rare disease data, again considering its scarcity, should be made available for a wide range of purposes, including diagnostic matchmaking, research to discover new biomarkers, natural history studies to better understand the nature

of a rare disease, and the recruitment of individuals into precision medicine clinical trials. Generating multiple siloed resources for multiple different purposes is simply not feasible.

4.2 Legacy collections

Legacy collections of samples and data are those collected without consent or without consent covering core consent elements required to conduct research or data sharing. These collections present an ethical dilemma: they often continue to have great scientific and societal value if shared and used for research, but the consent to do so is missing or insufficient. This is particularly a problem in biobanking, where samples are often collected many years before they are able to be distributed, aggregated, or analyzed. As the years pass, scientific and data sharing practices can change, and regulatory and ethical frameworks can evolve. As a result, the existing consents may become insufficient. One could argue the ethical dilemma is even more pressing in rare disease contexts, because of the high unmet need for research into novel diagnostics and therapies, as well as the associated practical difficulties of recruiting geographically disperse patients and collecting samples.

Solutions have been developed for legacy collections that aim to strike an appropriate balance between making them available for research, while also making best efforts to communicate with and respect the expectations of patients. When seeking to study or share legacy collections, an important starting point is to assess the existing consent materials (if applicable) [34]. If core elements of consent are met, then the research may be able to proceed. If the consent is silent on the desired research or data sharing, then patients should be recontacted to renew their consent. If permitted by applicable norms, it may also be sufficient to notify the patients and provide them with an opportunity to opt-out. In many cases, however, re-consent or re-contact will be impossible if patients can no longer be found. In such cases, some jurisdictions allow research ethics committees (RECs) to alter or waive consent requirements, as long as certain conditions are met. In Canada for example, a consent alteration/waiver is available where research is minimal risk, consent is impracticable, the alteration/waiver will not adversely harm individuals, there are appropriate safeguards in place, and there has been no clear refusal by the individual (art 3.7B (samples); art 5.5A (identifiable information)) [35].

Where consent is silent, an ethics waiver may be more easily justified. Where consent makes a specific commitment (e.g., guarantees data will be kept confidential, or will only be used for a specific research project), it may be harder to justify a waiver. Because consent is often take-it-or-leave-it, however, it is not necessarily clear if the commitment was determinant to the patient's decision, or what the patient would have preferred. An active refusal by the patient to participate in research or data sharing is a more clear-cut case [36]. For example, the patient may have been offered the option to participate in research or to share their data and may have refused. Indeed, the revised US Common Rule for research on human subjects in the US prohibits use of an ethics waiver when a patient has rejected a broad consent (§46.116) [37]. Practically, however, given the limits of tracking systems, it is unclear how these refusals can be tracked and respected over time.

Typically, ethics waivers are used to approve a specific research project using legacy collections, but in some cases, they have also been used to approve the deposit of data into international databases for onward sharing. The GA4GH recommends ethics waivers under certain conditions for international sharing of genomic and health-related data [34].

A final consideration offering some additional flexibility in rare disease contexts is patient involvement. Patient groups or representatives may lend moral support

to a particular interpretation of an existing consent, or to a particular decision to re-use or share legacy collections with an ethics waiver. This can help to alleviate uncertainty over what patients would have wanted in cases of uncertainty. If core consent elements are legally required, however, such modifications will not be possible.

4.3 Consent and capacity: pediatrics

An additional challenge that tends to be overlooked in data sharing discussions is that many patients with rare disease are minors, who are generally legally presumed to lack the capacity to consent to genetic testing, research participation, or release of personal data on their own [38]. Moreover, a number of rare diseases involve intellectual disability, which may diminish the decision making capacity of rare disease patients. Regulatory frameworks developed in the context of experimental research are traditionally protectionist, aiming to ensure vulnerable individuals are protected from harm and exploitation. These frameworks place several limitations on research involving minors. Where research is allowed, additional safeguards must be in place for minors. The guiding legal and ethical principle is the best interests of the child, though this principle is somewhat modified in the research context. A modified best interests limitation is only allowing research if it offers direct benefit to the individual or to individuals with a similar age or condition, and that the benefit is favorable vis-a-vis the individual risks, as determined by a REC. Where research is permitted, the minor is protected by parental (or legal guardian) representation, who must provide informed consent. The parent would be provided with detailed information about the research and would be asked to consent on behalf of the child.

This protectionist approach has been somewhat modified both by public health and human rights concerns. From a public health perspective, the inclusion of children in research is imperative to ensuring that the standard of care improves for conditions that predominantly affect children. Given the physiological differences between children and adults, drugs demonstrated to be safe and effective for adults may not be so for children. Human rights instruments and discussions have also highlighted the principle of non-discrimination, which argues that some attempts to protect children ultimately result in their exclusion from participation in society.

Another important human rights and ethical principle is respect for the developing autonomy of the child. Mature-minor exceptions address the developing capacity of minors to make decisions. These regimes allow exceptions for children below the age of majority to make certain their own decisions (e.g., for health care) if they demonstrate their capacity to understand information and appreciate the consequences of decisions. Regardless of who has ultimate legal capacity to consent, children should generally be given appropriate opportunities to be involved in decisions concerning them. Many health research ethics guidelines recommend that children should be involved in decisions through assent, where the child is provided with age appropriate information and asked if they would like to participate, and dissent, where a clear objection to participation must also generally be respected by researchers. Dissent is more clearly applicable in experimental research, such as distress caused by a needle, than in data-intensive research.

Based on the principle of inclusion, as well as practical implications, the GA4GH Pediatric Task Team has argued that those who generate pediatric data as part of research or clinical care have an obligation to offer minors and their parents an opportunity to share their data, so as to benefit the care of children in the future [39]. The best interests of the child are ensured in this context through the benefit–risk assessment, ongoing data governance (to maximize scientific and societal

benefits while minimizing risk), parental representation and informed consent, and the child's involvement through assent processes.

5. Patient involvement

This section discusses patient involvement (also referred to as patient engagement) in the governance of research, challenges to involving patients effectively and responsibly, and how rare disease patients may be involved specifically in the governance of biobanking and data sharing.

Respect for communities is an important ethical principle in health research. According to Charles Weijer *et al*. researchers have "an obligation to respect the values and interests of the community in research and, wherever possible, to protect the community from harm" [40]. This principle is also prominently featured in the international health research ethics guidelines of the Council for International Organizations of Medical Sciences and World Health Organization: "[r]esearchers, sponsors, health authorities and relevant institutions should engage potential participants and communities in a meaningful participatory process that involves them in an early and sustained manner in the design, development, implementation, design of the informed consent process and monitoring of research, and in the dissemination of its results" (Guideline 7) [41]. Communities are not only defined geographically, but include subpopulations affected by or able to influence research. Patients affected by a particular condition are clearly a key stakeholder in related research.

The involvement of patients and patient advocacy organizations across the research, drug-development, and delivery of care life cycle is increasingly practiced and is the subject of numerous national and international research ethics guidelines. Patient involvement is when patients "meaningfully and actively collaborate in the governance, priority setting, or design and conduct of research" [42]. The term patient is understood broadly to include "those having or a risk of having a medical condition, their families, and their caregivers" [43]. Patients have intimate, lived experience with and understanding of their medical condition, and how symptoms affect their everyday lives. These perspectives can inform the priorities, goals, and conduct of research and the ultimate value new diagnostic tools and therapies can deliver to patients.

The CIOMS/WHO guidelines also reflect the importance of engaging patient communities in the governance of research. Patients have valuable perspectives on both the potential value of research and the acceptability of associated physical and privacy risks. Patients also have perspectives on how consent documentation can meaningfully communicate the nature of research, its benefits and risks, and the safeguards in place to limit risks. Patient involvement in the governance of research can supplement the efforts of RECs to ensure the ethical conduct of research. There is a spectrum of patient involvement approaches, from feedback through surveys or workshops, to advisory boards, to formal leadership roles within research, to patient-led initiatives where patient groups decide when and how to engage experts [44].

Patient involvement can involve important costs and potential delays for research. There is a risk that involvement initiatives successful in specific research and community contexts are extrapolated by policy-makers or oversight bodies into generic ethical requirements [45]. Researchers alone cannot be expected to bear the financial and administrative burden of patient involvement without appropriate support from funding agencies and institutions. Imposing specific forms of involvement as an ethical requirement may also encourage a compliance mentality where

researchers want to get it out of the way rather than developing meaningful community and partnership with patients. Systemic barriers and negative perceptions can discourage patients from meaningfully engaging with researchers. Patients are already dealing with the burden of living with a disease and potentially also the burden of participating in research. It may therefore be difficult for them to visit research sites in order to participate in unpaid involvement activities. Patients may also have negative perceptions that they will not actually be listened to [42]. Some critics have highlighted that patient involvement is usually designed to advance an institutional agenda rather than truly give a voice to all patients [45]. Involvement activities do not necessarily mean that patients have significant decision-making power. Researchers may preferentially seek to engage with patients who have considerable experience as research participants, as well as experience with research involvement activities, which may reduce opportunities to hear other voices.

The organizational governance of patient advocacy groups is often informal, which raises concerns about democratic representation and managing conflicts of interest. There can often be major differences of opinion within a patient community. Some patient advocacy groups receive significant financial support from pharmaceutical companies and may not have formal processes in place for declaring and managing these conflicts [46]. Patient involvement challenges may be exacerbated in rare disease contexts. Patient involvement for rare disease may be difficult for the same reasons that doing research on rare disease is difficult. Patients may be small in number, geographically dispersed, and may have very heterogeneous experiences with the disease. This makes it hard to survey patients about their views on research. Many patients with rare disease struggle to even receive an accurate diagnosis, which may affect their ability to identify with and organize a specific community in the first place.

Patients are also increasingly engaged in the governance of biobanking and data sharing. YOURDNAYOURSAY is an interactive, international, online survey exploring public perspectives about the international sharing of genomic and health-related data. The results of the survey address public fears over potential harms, public willingness to release their data, and how trust differs between organizations [47, 48]. A European survey specific to rare disease patients found they were supportive of data sharing to improve research and health care, as long as steps were taken to provide individual patients with meaningful choices, to protect patient privacy, and to provide patients with transparent information about how their data are shared and used [49]. Patients can also be engaged in the design of governance documents for biobanks and databases, such as access policies, privacy safeguards, and consent forms (see previous sections). This involvement can provide assurances that governance strikes an appropriate balance between openness and promotion of science with protection of participant privacy. In many cases, patients may be highly supportive of greater openness in research and their perspectives may serve as a counterweight to overly protective stances by oversight bodies like RECs. Patients may also participate directly on biobank access committees, influencing decisions about which researchers receive samples and data, for which research projects.

Patient involvement in the governance of biobanking and data sharing can raise tensions between community control and scientific openness. In particular, the value of biobanks and data sharing is often dependent on their integration into networks allowing integration of multiple resources to increase statistical power. This is particularly true for rare disease. There is a recognized need for harmonized ethical and legal governance of biobanks and databases to enable such integration. One potential solution is for involvement activities to address the importance of harmonization with patient groups, to make sure this value is taken into account in

the co-development of governance [50]. Another potential solution is to develop more concerted public involvement efforts in the development of international standards for biobank and data sharing governance.

6. Conclusion

Rare disease research continues to be hampered by lack of academic and commercial incentives and practical barriers to conducting research involving small, geographically dispersed populations. This results in limited understanding of rare diseases, delayed diagnoses and a lack of therapeutic options for patients. There is hope that international biobank networks and data sharing can improve care and advance research into rare disease. Research ethics concerns about protecting patient privacy, enabling individuals to make informed decisions, and involving patients in governance deserve concerted and nuanced attention in these contexts. Standard governance approaches may need to be re-calibrated for rare disease contexts, given the necessity of openness, high unmet need, and the willingness of many rare disease patients to contribute to biobanks and databases, despite minor privacy risks. This is not to say that rare patients do not care about privacy or about being offered meaningful choices. Involving patients in the governance of biobanks and data sharing, while appropriately highlighting the importance of international collaboration, can help to ensure these activities ultimately improve the prospects of those with rare disease.

Acknowledgements

The author would like to thank Genome Canada, Genome Quebec, and the Canadian Institutes of Health Research for funding support, the hosts of the International Symposium on Rare Diseases held in Santiago, Chile in September 2019, and Kristina Kékesi-Lafrance for helping with the preparation of the chapter.

Author details

Adrian Thorogood
Centre of Genomics and Policy, McGill University, Canada

*Address all correspondence to: Adrian.thorogood@mcgill.ca

References

[1] Mulberg AE, Bucci-Rechtweg C, Giuliano J, Jacoby D, Johnson FK, Liu Q, et al. Regulatory strategies for rare diseases under current global regulatory statutes: A discussion with stakeholders. Orphanet Journal of Rare Diseases. 2019;**14**(1):36

[2] Might M, Wilsey M. The shifting model in clinical diagnostics: How next-generation sequencing and families are altering the way rare diseases are discovered, studied, and treated. Genetics in Medicine. 2014;**16**(10):736-737

[3] Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: Diagnosing rare disease in children. Nature Reviews Genetics. 2018;**19**(5):253-268

[4] CADTH. Genome-wide sequencing: ethical considerations [Internet]. 12 November 2019. 37 p. Available from: https://www.cadth.ca/sites/default/files/hta-he/he0020-genome-wide-sequencing-ethical-considerations.pdf [cited: 17 December 2019]

[5] National Institutes of Health. DRAFT NIH Policy for data management and sharing [Internet]. November 2019. 4 p. Available from: https://osp.od.nih.gov/wp-content/uploads/Draft_NIH_Policy_Data_Management_and_Sharing.pdf [cited: 17 December 2019]

[6] ACMG Board of Directors. Laboratory and clinical genomic data sharing is crucial to improving genetic health care: A position statement of the American College of Medical Genetics and Genomics. Genetics in Medicine. 2017;**19**(7):721-722

[7] Wright CF, Ware JS, Lucassen AM, Hall A, Middleton A, Rahman N, et al. Genomic variant sharing: A position statement [version 2; peer review: 2 approved]. Wellcome Open Research. 2019;**4**(1):22

[8] Gurovich Y, Hanani Y, Bar O, Nadav G, Fleischer N, Gelbman D, et al. Identifying facial phenotypes of genetic disorders using deep learning. Nature Medicine. 2019;**25**(1):60

[9] Hallowell N, Parker M, Nellåker C. Big data phenotyping in rare diseases: Some ethical issues. Genetics in Medicine. 2019;**21**(2):272-274

[10] Dove ES. Biobanks, data sharing, and the drive for a global privacy governance framework. The Journal of Law, Medicine & Ethics. 2015;**43**(4):675-689

[11] NHS England [Internet]. 100,000 genomes project; [about 2 screens]. Available from: https://www.england.nhs.uk/genomics/100000-genomes-project/ [cited: 10 December 2019]

[12] Department of Health and Social Care [Internet]. DNA mapping to better understand cancer, rare diseases and infectious diseases. 5 July 2013 [about 3 screens]. Available from: https://www.gov.uk/government/news/dna-mapping-to-better-understand-cancer-rare-diseases-and-infectious-diseases [cited: 10 December 2019]

[13] NHS England [Internet]. 100,000 genomes project; [about 2 screens]. Available from: https://www.gov.uk/government/news/dna-mapping-to-better-understand-cancer-rare-diseases-and-infectious-diseases [cited: 10 December 2019]

[14] Genome Canada, Canadian genomics partnership for rare diseases (CGP4-RD): Mission statement. Available from: https://www.genomecanada.ca/en/programs/precision-health-strategy/canadian-genomics-partnership-rare-diseases-cgp4-rd-mission

[15] EJP RD [Internet]. About EJP RD [about 2 screens]. Available from:

http://www.ejprarediseases.org/index.php/about/ [cited: 10 December 2019]

[16] World Economic Forum [Internet]. Breaking barriers to health data project; [about 3 screens]. Available from: https://www.weforum.org/projects/breaking-barriers-to-health-data-project/ [cited: 10 December 2019]

[17] Constine J. RDMD attacks rare diseases with data mined from health records [online]. TechCrunch; 2018. Available from: http://social.techcrunch.com/2018/08/20/rdmd/

[18] Bruckner T. (TranspariMED) [written in close consultation with Cochrane, CRIT and Transparency International's Pharmaceuticals and Healthcare Programme]. Clinical trial transparency: Aa guide for policy makers [Internet]. UK. 2017. p. 40. Available from: https://docs.wixstatic.com/ugd/01f35d_def0082121a648529220e1d56df4b50a.pdf [Transparency International Report] [cited: 17 December 2019]

[19] European Medicines Agency [Internet]. Transparency; [about 9 screens]. Available from: https://www.ema.europa.eu/en/about-us/how-we-work/transparency [cited: 17 December 2019]

[20] Krumholz HM, Waldstreicher J. The Yale open data access (YODA) project—A mechanism for data sharing. New England Journal of Medicine. 2016;**375**(5):403-405

[21] ClinicalStudyDataRequest [Internet]. Our mission; [about 1 screen]. Available from: https://clinicalstudydatarequest.com/Default.aspx [cited: 17 December 2019]

[22] Vivli: Center for Global Clinical Research Data [Internet]. About Vivli: Overview; [about 2 screens]. Available from: http://vivli.org/about/overview/ [cited: 17 December 2019]

[23] Menon D, Clark D, Stafinski T. Reimbursement of drugs for rare diseases through the public healthcare system in Canada: Where are we now? Healthc Policy. 2015;**11**(1):15-32

[24] Perfetto EM, Oehrlein EM, Boutin M, Reid S, Gascho E. Value to whom? The patient voice in the value discussion. Value in Health. 2017;**20**(2):286-291

[25] Gymrek M, McGuire AL, Golan D, Halperin E, Erlich Y. Identifying personal genomes by surname inference. Science. 2013;**339**(6117):321-324

[26] Rocher L, Hendrickx JM, De Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. Nature Communications. 2019;**10**(1):1-9

[27] Rothstein MA, Zawati MN, Knoppers BM. Regulatory landscape of international irect-to-participant (DTP) genomic research: Time to untie the Gordian knot? The Journal of Law, Medicine & Ethics. 2019;**47**(2):1-14

[28] Global Alliance for Genomics and Health (GA4GH) in their Consent Policy. 2019. Available from: https://www.ga4gh.org/work_stream/regulatory-ethics/

[29] Council of Canadian Academies. Accessing health and health-related data in Canada [Internet]. 2015. 260 p. Available from: https://cca-reports.ca/wp-content/uploads/2018/10/healthdatafullreporten.pdf [cited: 17 December 2019]

[30] General Data Protection Regulation (GDPR), Regulation (EU) 2016/679 [Internet]. 2016. Available from: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679 [cited: 17 December 2019]; California Consumer Privacy Act of 2018, A.B. 375 [cited: 28 June 2018]

[31] Rabesandratana T. Researchers sound alarm on European data law. Science. 2019;**366**(6468):936

[32] Nguyen MT, Goldblatt J, Isasi R, Jagut M, Jonker AH, Kaufmann P, et al. Model consent clauses for rare disease research. BMC Medical Ethics. 2019;**20**(1):55

[33] Contreras JL, Knoppers BM. The genomic commons. Annual Review of Genomics and Human Genetics. 2018;**19**:429-453

[34] Global Alliance for Genomics and Health. Consent policy [Internet]. 2019. 6 p. Available from: https://www.ga4gh.org/wp-content/uploads/GA4GH-Final-Revised-Consent-Policy_16Sept2019.pdf [cited: 17 December 2019]

[35] Canadian Institutes of Health Research, Natural Sciences and Engineering Research Council of Canada, Social Sciences and Humanities Research Council. Tri-council policy statement: Ethical conduct for research involving humans—TCPS 2018 [Internet]. 2018. 211 p. Available from: https://ethics.gc.ca/eng/documents/tcps2-2018-en-interactive-final.pdf [cited: 17 December 2019]

[36] Gainotti S, Turner C, Woods S, Kole A, McCormack P, Lochmüller H, et al. Improving the informed consent process in international collaborative rare disease research: Effective consent for effective research. European Journal of Human Genetics. 2016;**24**(9):1248-1254

[37] Revised common rule, Electronic Code of Federal Regulations, 45 CFR 46 [Internet]. 2018. Available from: https://www.ecfr.gov/cgi-bin/retrieveECFR?gp=&SID=83cd09e1c0f5c6937cd9d7513160fc3f&pitd=20180719&n=pt45.1.46&r=PART&ty=HTML [cited: 17 December 2019]

[38] Dalpé G, Thorogood A, Knoppers BM. A tale of two capacities: Including children and decisionally vulnerable adults in biomedical research. Frontiers in Genetics [Internet]. 2019;**10**:289. Available from: https://www.frontiersin.org/articles/10.3389/fgene.2019.00289/full [cited: 2 May 2019]

[39] Rahimzadeh V, Schickhardt C, Knoppers BM, Sénécal K, Vears DF, Fernandez CV, et al. Key implications of data sharing in pediatric genomics. JAMA Pediatrics. 2018;**172**(5):476-481

[40] Weijer C, Goldsand G, Emanuel EJ. Protecting communities in research: Current guidelines and limits of extrapolation. Nature Genetics. 1999;**23**(3):27-80

[41] Council for International Organizations of Medical Sciences (CIOMS) [in collaboration with the World Health Organization (WHO)]. International ethical guidelines for health-related research involving humans [Internet]. 2016. 119 p. Available from: https://cioms.ch/wp-content/uploads/2017/01/WEB-CIOMS-EthicalGuidelines.pdf [cited: 17 December 2019]

[42] CIHR. Draft CIHR ethics guidance for developing research partnerships with patients [Internet]. 2019. 42 p. Available from: https://ktcanada.org/wp-content/uploads/2019/04/ethics_guidance_developing_research-en.pdf [cited: 17 December 2019]

[43] Hoos A, Anderson J, Boutin M, Dewulf L, Geissler J, Johnston G, et al. Partnering with patients in the development and lifecycle of medicines: A call for action. Therapeutic Innovation & Regulatory Science. 2015;**49**(6):929-939

[44] Deverka PA, Gilmore D, Richmond J, Smith Z, Mangrum R, Koenig BA, et al. Hopeful and concerned: Public input on building a trustworthy medical information

commons. The Journal of Law, Medicine & Ethics. 2019;**47**(1):70-87

[45] Johannesen J. The trouble with Patient and Public Involvement (PPI) [Internet]. Session presented at: 2018 Cochrane Colloquium. Edinburg, UK; 2018. Available from: https://johannesen.ca/2018/09/the-trouble-with-patient-and-public-involvement-ppi-keynote-at-cochrane-colloquium-2018/ [cited: 10 December 2019]

[46] von Tigerstrom B. The patient's voice: Patient involvement in medical product regulation. Medical Law International. 2016;**16**(1-2):27-57

[47] Milne R, Morley KI, Howard H, Niemiec E, Nicol D, Critchley C, et al. Trust in genomic data sharing among members of the general public in the UK, USA, Canada and Australia. Human Genetics. 2019;**138**:1237-1246

[48] Middleton A, Milne R, Thorogood A, Kleiderman E, Niemiec E, Prainsack B, et al. Attitudes of publics who are unwilling to donate DNA data for research. European Journal of Medical Genetics. 2019;**62**(5):316-323

[49] Courbier S, Dimond R, Bros-Facer V. Share and protect our health data: An evidence based approach to rare disease patients' perspectives on data sharing and data protection-quantitative survey and recommendations. Orphanet Journal of Rare Diseases. 2019;**14**(1):175

[50] Thorogood A. Towards Legal Interoperability in International Health Research. University of Toronto (School of Graduate Studies—Theses). 2019. 203 p. Available from: http://hdl.handle.net/1807/98411 [cited: 17 December 2019]