

POSITION STATEMENT

Open Access



# Sharing is caring: a call for a new era of rare disease research and development

Nathan Denton<sup>1,7\*</sup> , Andrew E. Mulberg<sup>2</sup>, Monique Molloy<sup>3</sup>, Samantha Charleston<sup>3</sup>, David C. Fajgenbaum<sup>3,4</sup>, Eric D. Marsh<sup>3,5,6,7\*</sup> and Paul Howard<sup>7\*</sup>

## Abstract

Scientific advances in the understanding of the genetics and mechanisms of many rare diseases with previously unknown etiologies are inspiring optimism in the patient, clinical, and research communities and there is hope that disease-specific treatments are on the way. However, the rare disease community has reached a critical point in which its increasingly fragmented structure and operating models are threatening its ability to harness the full potential of advancing genomic and computational technologies. Changes are therefore needed to overcome these issues plaguing many rare diseases while also supporting economically viable therapy development. In “Data silos are undermining drug development and failing rare disease patients (*Orphanet Journal of Rare Disease*, Apr 2021),” we outlined many of the broad issues underpinning the increasingly fragmented and siloed nature of the rare disease space, as well as how the issues encountered by this community are representative of biomedical research more generally. Here, we propose several initiatives for key stakeholders - including regulators, private and public foundations, and research institutions - to reorient the rare disease ecosystem and its incentives in a way that we believe would cultivate and accelerate innovation. Specifically, we propose supporting non-proprietary patient registries, greater data standardization, global regulatory harmonization, and new business models that encourage data sharing and research collaboration as the default mode. Leadership needs to be integrated across sectors to drive meaningful change between patients, industry, sponsors, and academic medical centers. To transform the research and development landscape and unlock its vast healthcare, economic, and scientific potential for rare disease patients, a new model is ultimately the goal for all.

**Keywords** Rare diseases, Drug development, Knowledge bases, Registries, Research design, Data management, Clinical trials

\*Correspondence:

Nathan Denton  
ndenton@upenn.edu

Eric D. Marsh  
marshe@chop.edu

Paul Howard  
phoward@amicusrx.com

<sup>1</sup>Gene Therapy Program, Department of Medicine, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>2</sup>Neurogene Inc, New York, NY, USA

<sup>3</sup>Department of Medicine, Orphan Disease Center, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>4</sup>Translational Medicine & Human Genetics, Perelman School of Medicine, University of Pennsylvania, Pennsylvania, PA 19104, USA

<sup>5</sup>Departments of Neurology and Pediatrics, Perelman School of Medicine, Children's Hospital of Philadelphia, University of Pennsylvania, Philadelphia, PA 19104, USA

<sup>6</sup>Division of Neurology, Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA

<sup>7</sup>Amicus Therapeutics, Philadelphia, PA 19104, USA



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Background

This position statement aims to encourage meaningful dialogue on the issues of data standardization and sharing between key stakeholders in the rare disease space, principally patients, industry, sponsors, and academic medical centers. We previously outlined many of the broad issues underpinning the fragmentation and siloed nature of the rare disease space, as well as how these issues are not unique to this community but biomedical research more generally [1]. We and many others [2–6] firmly believe that unlocking the full therapeutic/curative and economic potential of the rare disease space requires supporting non-proprietary patient registries [7, 8], greater data standardization [9, 10], global regulatory harmonization [5, 11], and new business models [12, 13] that encourage data sharing and research collaboration as the default mode [14, 15]. Here, we propose several initiatives for key stakeholders to reorient the rare disease space and its incentives in a way that we believe would cultivate and accelerate innovation.

## Main text

The rare disease research community is experiencing an explosion in activity as our understanding has grown for many clinically described disorders with previously unknown etiologies, genetics, and mechanisms [16]. The emergence of improved mechanistic understanding, combined with new tools like gene therapy [17], gene editing [18], and next-generation sequencing [19], has inspired optimism in the patient, clinical, and research communities that disease-specific treatments are on the way. Growing research efforts accompany the notable success of several gene therapy programs to characterize the basic biology and clinical manifestations of rare diseases, with translational studies bridging this crucial gap [20].

Unfortunately, the successes are considerably outnumbered by the failed attempts to develop treatments for rare diseases. Indeed, more than 90% of all rare diseases do not have an approved treatment [21]. Drug development for this community faces multiple challenges, with most issues stemming from the fact that rare diseases affect small patient populations in a non-uniform manner [22]. These small, complex patient populations are often incompatible with many key design and statistical power principles required for traditional randomized controlled trials (RCTs) [23]. This general lack of information (in which available data are often heterogeneous and complex), combined with significant inter- and intra-patient variability (in terms of disease onset, presentation, progression, response to treatment, etc.), results in a high degree of unexplained variation that makes it difficult for researchers and regulators to confidently assess potential efficacy signals.

Registries and natural history studies are potentially well positioned to characterize inter- and intra-patient variability and inform precision-guided trials to test targeted therapeutics in stratified populations. However, many impediments prevent the optimization of these studies in the rare disease context such as poor data collection and data management that is perpetuated by longstanding “gaps in standardization, disparate privacy laws and international regulations, and the shortage of shared international regulatory endorsements of best practices for data collection.” [6] Better tools for accelerated development of rare disease therapeutics will continue to elude us so long as we remain dependent on building blocks (i.e., patient registry and natural history data sources) that are incomplete, flawed, or difficult to use.

That the field is not better organized should be unsurprising to many. Data collected by academic, pharmaceutical, and patient groups on rare disease patients are primarily locked away within proprietary databases that are often selectively leveraged to protect funding streams, as well as support publications and nascent intellectual property. Each of these goals is valid and may not impede rapid advances in more common disease states, like oncology and cardiovascular disease. In the rare disease context, however, this situation typically means that no individual stakeholder can accumulate a critical mass of knowledge to de-risk their rare disease drug development programs in favor of success in the vast number of cases. This thereby delays or denies highly vulnerable patients potential treatments while also promoting the waste of precious time, energy, and resources. Extrapolating current drug development timelines, failure rates, and limited funding streams to the >7,000 known rare diseases suggests that it would take many decades to develop treatments/cures for even a fraction of them [24]. Given the substantial yet oft-underestimated scale of the rare disease burden [22] and immense, growing healthcare costs [25], creating scalable platforms and common methods to leverage our knowledge of shared etiology and pathology across rare diseases would provide the greatest positive impact on this patient population [26].

We have arrived at this unfavorable situation by the key stakeholders merely operating within the current system of economic and academic incentives regarding our most scarce resource: rare disease patient data. The reluctance to develop new standards and incentives to actively encourage data sharing makes the already daunting task of developing rare disease therapies even more difficult, if not self-defeating. There is clear evidence that data sharing and collaboration are highly effective and efficient in improving health [27, 28] and creating value [14, 29] while also enhancing affordability [30]. Moreover, the current situation concerning the proprietary nature of

patient registry data goes against the consensus of the community whom the various stakeholders are supposed to serve – the rare disease community – who overwhelmingly wish for their data to be shared (and protected) [7, 31]. This is particularly pertinent given that patients and their families can find it increasingly difficult to participate in rare disease research due to misaligned priorities between stakeholders and a general lack of transparency and collaboration, despite being an extremely motivated and altruistic group who wish to advance science to benefit the community [32, 33]. We are unwitting agents in a system that is a means to an end, but it is unclear for whom and for what.

There is an immense moral imperative to rectify this situation rapidly, with the urgency being two-fold. In addition to rare disease patients' and their family's race against time, we are on the cusp of massive data proliferation. Technology and open data rules are democratizing data aggregation and registry formation such that any patient community can easily become data enabled. This looming explosion of registries in the current system is poised to exacerbate data loss, replication, and/or data gaps unless meaningful steps are taken [10]. In a field where high-quality data are already limited, not addressing this avoidable systemic flaw will ensure a rare disease diagnosis will continue to be a tragic and hopeless moment for many patients and their families. Failing to harness the explosion in data collection and analysis techniques will allow this unique, potentially revolutionary, window of opportunity to close and instead leave the field continuing along its path to become simultaneously fragmented yet swamped by a cacophony of largely unusable registries [9].

We believe that the convergence of timing and technology provides a unique and important opportunity for sponsors, patients, and providers to focus on core data use issues - integrity, curation, and interoperability – to help bridge clinical care and clinical research. While we can and should be agnostic about technology standards and platforms, we as a community must adopt foundational principles that enable and encourage the creation of informative patient registries based on FAIR data principles – Findable, Accessible, Interoperable, Reusable [34].

For example, interoperable patient registries or natural history studies can inform strategies for stratification in a clinical trial (e.g., genotype) by helping to identify and test innovative clinical endpoints and outcomes, as well as determine branch points for adaptive trial design. One of the article's authors can also attest to patient-submitted electronic medical record data and biospecimens providing rich additional detail in real-time to bridge gaps in clinical trial data or guide novel biomarker hypothesis. The community has similarly called for greater guidance

from regulators to ensure (adaptive) trials advance as swiftly as possible by utilizing patient- and caregiver-reported outcomes such as seizure frequency, experience of side effects, etc. to assess priors or provide external control arms [26].

This situation - and many others besides - also provides a prime opportunity to devise and deploy better methods to enhance our understanding of the rare disease patient journey and the basic etiology of rare diseases [35] as well as ensure alignment between stakeholders [36] on assessing meaningful outcomes to guide clinical care pathways and product reimbursement. Many of these innovative trial designs and regulatory frameworks based on the agile use of patient data have proven to be viable after their rapid deployment and assessment in response to the COVID-19 pandemic. Despite setbacks in some fields (e.g., oncology [37]), greater use of platform trials, remote and decentralized trials to protect medically fragile patients, and regulators' growing acceptance of real-world data (RWD) that may provide real-world evidence (RWE) to support both safety and efficacy assessments have enabled biomedical research to continue during the COVID-19 pandemic while being responsive to emerging medical challenges [38]. Generating and integrating FAIR data in the rare disease space will enable more innovative clinical trial design and execution to occur, for this space to evolve more quickly, and for the value of different data types to be determined at different stages to optimize the use of limited resources, time, and energy.

Before the COVID-19 pandemic, a handful of cases operating within the constraints of the current research and development ecosystem successfully implemented a 'next-generation' approach in their use of patient data. They were able to develop comprehensive datasets, accelerate understanding and innovation, and ultimately devise and evaluate effective interventions in a relatively short amount of time. For example, the Castleman Disease Collaborative Network combined pre-existing clinical data with proteomic data, machine learning, and statistics to compile and analyze a dataset containing fewer than 100 patients; from this, it was possible to create a molecular subtyping method that resulted in the delineation of distinct pathogenic mechanisms [39] that enabled the development of an effective treatment regimen [40]. A foundational aspect of this effort was the ACCELERATE natural history registry, which combined patient medical record data with patient-reported outcomes data and biospecimens from patients. This example also highlights the potential for drug repurposing as the new treatment regimen discovered through proteomics was a pre-existing cancer treatment that was repurposed for Castleman disease [41]. Significant improvements to patients' lives are evidently attainable by collating current resources - however modest - and

engaging in collaborative efforts to bridge gaps. As rare disease research advances, the benefits of combining omics with clinical phenotype data/biological samples [42] and conducting multi-omic research grow increasingly clear in various contexts [43], from obtaining diagnoses to the molecular characterization of disease and its subtypes to identifying disease biomarkers and novel therapeutic targets [44, 45]. Realizing these benefits, however, is contingent on a collaborative culture and standardization measures that are yet to fully manifest.

The Critical Path Institute's Rare Disease Cures Accelerator Data Analytics Platform (RDCA, supported by the U.S. FDA and NORD) and RARE-X also represent forward-thinking platforms focused on aggregating, curating, and integrating datasets to improve the rare disease research and development ecosystem. These ventures provide a window into the advantages and possibilities enabled by standards-based data sharing, aggregation, and analysis to inform study design (e.g., to validate novel endpoints) and prospective data collection (including patient-level data) in the rare disease space to support both regulatory and post-market use cases (including clinical care and outcomes-based contracting); they also provide a model of transparent, inclusive, and patient-empowering data governance practices through industry and regulatory engagement. Although these initiatives provide grounds for hope, it must be emphasized that they are still the exception, rather than the norm.

Groups at the NIH and FDA have started programs to address the issues affecting the rare disease field, but we and many others feel that there is a general lack of clarity and direction from regulators on data sharing and data standards to support regulatory decision making. With this in mind, we organized a webinar in September 2020 entitled "Let's Get Real: Harnessing Non-Proprietary Patient Registries and RWE to Accelerate Rare Disease Drug Development." [46] This series began a much-needed discussion of issues regarding regulatory guidance/best practices, data sharing, and novel approaches to natural history studies and the use of real-world data. Panelists from industry, the FDA, the NIH, academia, and patient foundations all contributed to the webinar series, which involved public sessions open to the community and closed whiteboarding sessions that explored the ideas and questions raised in each session in greater depth.

The last webinar presented a dialogue with Dr. Amy Abernethy, former Principal Deputy Commissioner of the FDA, on the agency's approach to real-world data, and the lessons learned from the COVID-19 pandemic. After these discussions, we wholeheartedly believe that it is possible to challenge current assumptions and to positively disrupt the research ecosystem by carefully deploying policies and incentives to effect mutually beneficial

change. It is abundantly clear that any stakeholder acting alone cannot tackle the challenges facing the rare disease space. Moreover, there is no single solution to this complex and evolving situation as the underlying technologies and analytical tools change over time. However, common guiding principles, agreement on best data practices/standards, and reusable infrastructure employed by the various national and international stakeholders could bring about significant positive change rapidly within the rare disease ecosystem. We overview the key problems and propose solutions in the following sections.

**Problem 1** Small, highly geographically dispersed patient populations across multiple regulatory jurisdictions make it extremely difficult to aggregate enough data to advance innovative tools for rare disease drug development.

**Proposed solution** Build national and international consensus among stakeholders - guided by regulatory leadership - to determine high-quality disease definitions and standardize clinical outcome assessments for individual diseases using Clinical Data Interchange Standards Consortium (CDISC) terminology and standards in anticipation of mapping, sharing, and harmonizing pre-existing datasets.

- Regulatory agencies - such as the FDA and EMA - should embrace the unique role of regulator, technical expert, and convenor/facilitator. They should provide regular forums (both in person and virtual) for discussion and engagement between the different stakeholders (i.e., patient groups, clinical, academic, industry, and regulators). The FDA's current Patient Focused Drug Development meetings [47] and/or Patient Listening Session [48] format could be used to reach consensus between stakeholders on priority research questions, terminology, data types and their uses (e.g., what constitutes 'regulatory-grade' data [49], dataset linkage mechanisms, and analytical methods, including endpoints and adaptive clinical trial design) that ultimately satisfy regulatory standards and span multiple disease groups. Best practices and guidelines should be determined according to successful, transparent collaborations between patient organizations and clinical research sponsors [50] (e.g., community advisory boards [51] and patient advocacy groups [52, 53]) and the academic communities at large (see the FAIR principles - Findable, Accessible, Interoperable, Reusable [34] - to guide the flexible (re)design of current and future registries). High-level research networks, such as the federation of children's hospitals, could represent key

sites to run pilot schemes to hone and standardize processes before broader roll out.

- Align international stakeholders to generate a ‘target data manifest’ for a given rare disease to support therapy development, in which the necessary data types and source(s) are determined for each stage and for what purpose. Data collection methods should be devised around the limitations of data sources to obtain a comprehensive longitudinal data landscape that can inform clinical decision-making and generate a knowledge base that is “fit for purpose” [54].
- Engage with the technology industry to develop systems that utilize common data ontologies to capture and share regulatory-grade data with minimal burden from any/all contributors (and confer the appropriate privacy protections). Regulators should act as a convener for developing common outcomes measures that include algorithms, wearable devices, medical imaging, etc., to create standardized, freely accessible tools which enable better quantitative assessment of core clinical concepts of interest (particularly for common outcomes/symptoms such as seizures/tremors). Utilize technology and flexible/adaptive strategies to optimize research design with a specific rare disease case to reduce wastage, improve efficiency, and maximize effectiveness [55]. Build connections among previously fragmented datasets and support the development of artificial intelligence (AI) and machine learning methods to enhance data analysis [56]. Enable patients to enroll into prospective natural history studies virtually without having to travel to a study site, and task central institutions with obtaining and entering data for each patient (see ACCELERATE for Castleman Disease [41]).

**Problem 2** The general absence of a cultural expectation of pre-competitive data sharing by key stakeholders has led the field to a situation in which data hoarding is the norm. This stifles innovation by making it difficult to develop novel drug development tools, as has been observed in more common therapeutic areas like oncology.

**Potential solution** Stakeholders in industry, academia, and patient groups should be educated about the benefits of data sharing and resource pooling, with a particular focus on key gatekeepers such as children’s hospitals, research director networks, RD CEOs, and patient group leaders. Academics and the NIH should reward responsible data sharing through prioritization of grants, tenure, and publications.

- Engage industry, academia (including journals [57]), patients, and regulators in dialogue so

all stakeholders become more comfortable in collecting and sharing medical data in new ways to encourage the voluntary adoption of common data models and best practices [58, 59]. Umbrella rare disease associations should take the lead by providing educational resources and a framework for grading how well organizations achieve FAIR data practices in patient registries and NIH-funded research. These organizations should also lead discussion of how strategies such as federated data exchanges [60] protect privacy and ownership of data. Pharmacosafety and/or pharmacovigilance may represent topic areas which could unify stakeholders and act as a starting point to develop data-sharing practices.

- Increase public funding by national (e.g., countries) and transnational (e.g., European Union, World Health Organization) entities for rare disease research to favor the generation of fewer centralized registries by covering the costs of data curation and data sharing assessments. Such a publicly funded central data infrastructure should support the deposition and dissemination of well-annotated data in formats that enable use by multiple groups. This would help ensure registries are adequately equipped to meet regulator, sponsor, and payer needs while disassembling the proprietary data siloes that have developed due to few major public (e.g., NIH) initiatives for rare diseases [61] coupled with a reliance on public fundraising from families. Targeting funding towards data-enabled registries, through platforms like the RDCA, for example, could enhance the value of future and existing registries by encouraging and supporting the development of biobanks linked to registry entries (in which sample collection, bioassay usage, etc. are standardized and shared). High-quality registry data linked to biospecimens could incentivize additional rare disease research from companies and venture capital by de-risking early-stage proof-of-concept trials. Encourage and support international collaboration (possibly even the merging) between rare disease registries.
- Encourage universities around the world to uncouple the academic promotion/tenure process from data ownership and outcomes that flow almost exclusively from it in favor of data-sharing practices and collective achievements (e.g., formal recognition of research productivity not just via manuscript authorship, but as a data contributor or analyst, in which frequency of citations, data reuse, and/or impact of data analysis, for example, could be included in assessments).

**Problem 3** Existing incentives discourage companies from engaging in more pre-competitive data-sharing platforms because of a lack of clear business models to reward sharing and insufficient incentives and/or pressure to use data responsibly.

**Proposed solution** Policymakers and regulators should define clear use cases for how aggregated, high-quality data sets can be used to satisfy pre- and post-market regulatory requirements, for instance, for Phase IV trials, externally controlled trials, and label expansion.

- Regulatory agencies (e.g., the EMA and FDA) should lead the way in defining use cases for developing stage-appropriate and innovative endpoints [62] that encourage data standardization and data sharing. This could involve providing guidance on how high-quality data from registries and real-world evidence can satisfy post-market requirements for confirmatory trials and create externally controlled trials. Regulators should support the development of a collaborative, non-proprietary/pre-competitive ‘data space’ that encourages data sharing, collaboration, and data curation to support endpoint development (see the FDA’s National Evaluation System for Health Technology [63] and Federal Health IT Strategic Health Plan 2020–2025 [64]).
- Use regulator-hosted forums to build familiarity between stakeholders to conduct innovative regulatory work in real-time to facilitate the meeting of objectives in rare disease research and development. For instance, such forums might be used to identify novel data modalities that enhance understanding and are acceptable for regulatory decision making; this could include patient-entered digital signals, FDA-approved device data, or other datasets (e.g., electronic health records/EHR and insurance claims). These forums should also provide greater representation of patient groups’ perspectives to ensure realistic, feasible, and appropriate standards are being applied consistently to the drug development/approval process in a disease-specific manner. They could also reduce industry’s perception of regulatory risk and create clear value for participants to share “lessons learned” (see FDA’s patient listening sessions [48]) to accelerate the development process.
- Draw upon Congress or equivalent national governments to empower regulators with expert and patient input that encourages and/or enforces data sharing. In a transitory period, Congress could incentivize data-sharing standards by making the training of AI algorithms to support data integration a national strategic priority (e.g., for electronic health records). The FDA could also be empowered

to fast-track RWE applications that incorporate data-sharing practices before mandating future applications adhere to data sharing criteria ([59], see Sect. 309).

- Empower patients to be involved in data governance [10] and promote data literacy among patient groups. Develop a policy framework that aspires towards maximal transparency in data usage during sharing (e.g., provide public information on data accession and utilization, akin to the SWIFT system for transactions in banking). Develop and implement best practices for obtaining patient consent in research settings that support patient recruitment and facilitate their role as data generators while protecting their privacy, rights, and well-being without constraining research [65, 66]. Ensure there is a framework in place to ensure that when people “leave their legacy in data,” the patient’s intent is honored and their data are used as effectively as possible and shared as widely as necessary (i.e., in accordance to the FAIR principles) instead of being hoarded.

### Conclusion

The rare disease world is entering a potentially irrevocable state that will exacerbate already-significant delays and obstacles to making advances for rare disease patients in whom permanent loss of function or mortality can be measured in mere months. Given that many rare diseases affect children, meaningful change in this domain is desperately required and massively overdue to support this vulnerable population better. The advent of potentially curative genomic technologies, along with advances in computing power and analytics, provides an opportune time to start a broad, international dialogue and reach consensus between stakeholders on key issues relating to registry design, data sharing, and data governance. At best, the rare disease space will continue to grow and innovate, but this growth will be limited in trajectory, scope, and success. At worst, current structural flaws and practices left unaddressed have the potential to fragment the rare disease research and development landscape permanently, locking the stakeholders in a self-motivated yet counterproductive struggle that stifles innovation and ultimately seals the fate of the vulnerable patients they are supposed to serve.

The necessary transformation of the research and development ecosystem, with its economic and academic incentives that precipitated the current predicament, requires leadership to be shown and dialogue to be started between patients, industry, sponsors, and academia. Regulators need to provide leadership and directionality to the various stakeholders to cultivate an international environment in which top-down guidance

(e.g., for regulatory-grade data collection and clear use cases) synergizes with bottom-up innovation (e.g., new business models founded on agile data curation, aggregation, and analytics according to patient-centered data control) to accelerate drug development. This guidance needs to be aligned globally and encouraged through a mixture of novel economic incentives and policies to shift the research ecosystem towards non-proprietary, shared patient registries as the default standard. This scenario would help usher in a new data ecosystem that rewards collaborators for generating regulatory-quality data and analytics that capitalize on the potential of new approaches like AI and machine learning, platform trials, adaptive and remotely conducted trial designs, and real-world data (e.g., from wearable devices) to accelerate rare disease drug development. The solutions are within our grasp. We hope that the various stakeholders will step up to the challenge to maximize the limited resources, time, and energy available in the rare disease world for the sake of the patients and their families.

#### List of abbreviations

AI	Artificial intelligence.
CDISC	Clinical Data Interchange Standards Consortium.
COVID-19	Coronavirus disease of 2019.
EHR	Electronic health record.
EMA	European Medicines Agency.
FAIR	Findable, Accessible, Interoperable, Reusable.
FDA	Food & Drug Administration.
FOCR	Friends of Cancer Research.
IT	Information technology.
NIH	National Institutes of Health.
NORD	National Organization for Rare Disorders.
ODC	University of Pennsylvania Orphan Disease Center.
RCT	Randomized controlled trial.
RDCA	Rare Disease Cures Accelerator.
RD CEO	Research Director-Chief Executive Officer.
RDCA-DAP	Rare Disease Cures Accelerator-Data and Analytics Platform.
RWD	Real-world data.
RWE	Real-world evidence.
SWIFT	Society for Worldwide Interbank Financial Telecommunication.

#### Acknowledgements

We extend our gratitude to all participants of the seminar series held during September 2020 hosted by Penn Medicine's Orphan Disease Center and Amicus Therapeutics entitled 'LET'S GET REAL: Harnessing Non-Proprietary Patient Registries and RWE to Accelerate Rare Disease Drug Development' seminar series: Amy Abernethy, Brian Alexander, Jeff Allen, Gideon Blumenthal, Betsy Bogard, Alicyn Campbell, Elizabeth Hart, Jonathan Hirsch, Jane Larkindale, Craig Lipset, Sean Khozin, Christopher Kim, Anne Pariser, Anthony Philippakis, Elizabeth Powers, Tracy Salazar-Dixon, Laura Schanberg, Eric Sid, Suzanne Thornton-Jones, Steve Usdin, Jill Weimer, and Eric Zuckerman.

#### Authors' contributions

N.D. – conceptualization; writing – original draft preparation; writing – review and editing.

A.E.M. – writing – review and editing.

M.M. – conceptualization; writing-review and editing.

S.C. – conceptualization; writing – review and editing.

D.C.F. – writing – review and editing.

E.D.M. – conceptualization; writing – review and editing.

P.H. – conceptualization; writing – original draft preparation; writing – review and editing.

All authors read and approved the final manuscript.

#### Funding

Not Applicable.

#### Data availability

Not applicable.

#### Declarations

##### Ethics approval and consent to participate

Not applicable.

##### Consent for publication

Not applicable.

##### Competing interests

The Authors declare no Competing Financial or Non-Financial Interests.

Received: 16 February 2022 / Accepted: 2 October 2022

Published online: 27 October 2022

#### References

- Denton N, Molloy M, Charleston S, Lipset C, Hirsch J, Mulberg AE, Howard P, Marsh ED. Data silos are undermining drug development and failing rare disease patients. *Orphanet J Rare Dis.* 2021;16(1):161.
- Boycott KM, Lau LP, Cuttillo CM, Austin CP. International collaborative actions and transparency to understand, diagnose, and develop therapies for rare diseases. *EMBO Mol Med* 2019, 11(5).
- Ambrosini A, Quinlivan R, Sansone VA, Meijer I, Schrijvers G, Tibben A, Padberg G, de Wit M, Sterrenburg E, Mejat A, et al. "Be an ambassador for change that you would like to see": a call to action to all stakeholders for co-creation in healthcare and medical research to improve quality of life of people with a neuromuscular disease. *Orphanet J Rare Dis.* 2019;14(1):126.
- Mascalzoni D, Petrini C, Taruscio D, Gainotti S. The Role of Solidarity(-ies) in Rare Diseases Research. *Adv Exp Med Biol.* 2017;1031:589–604.
- Mulberg AE, Bucci-Rechtweg C, Giuliano J, Jacoby D, Johnson FK, Liu Q, Marsden D, McGoohan S, Nelson R, Patel N, et al. Regulatory strategies for rare diseases under current global regulatory statutes: a discussion with stakeholders. *Orphanet J Rare Dis.* 2019;14(1):36.
- Bétourné A, Walls RL, Bateman-House A, Ollivier C, Huynh H, Olson D, Borens A, Barrett JS. Rare Diseases Cures Accelerator Data and Analytics Platform (RDCA-DAP) best practices and recommendations for FAIR data, toward alignment with International Regulatory agencies. *Critical Path Institute*; 2022.
- 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 J Rare Dis.* 2019;14(1):175.
- Boulanger V, Schlemmer M, Rossov S, Seebald A, Gavin P. Establishing Patient Registries for Rare Diseases: Rationale and Challenges. *Pharmaceut Med.* 2020;34(3):185–90.
- Chin L, Khozin S. A digital highway for data fluidity and data equity in precision medicine. *Biochim Biophys Acta Rev Cancer.* 2021;1876(1):188575.
- Kodra Y, Posada de la Paz M, Coi A, Santoro M, Bianchi F, Ahmed F, Rubinstein YR, Weinbach J, Taruscio D. Data Quality in Rare Diseases Registries. *Adv Exp Med Biol.* 2017;1031:149–64.
- Horgan D, Moss B, Boccia S, Genuardi M, Gajewski M, Capurso G, Fenaux P, Gulbis B, Pellegrini M, Manu Pereira MDM, et al. Time for Change? The Why, What and How of Promoting Innovation to Tackle Rare Diseases - Is It Time to Update the EU's Orphan Regulation? And if so, What Should be Changed? *Biomed Hub.* 2020;5(2):1–11.
- Boran T, Menezes-Ferreira M, Reischl I, Celis P, Ferry N, Gansbacher B, Krafft H, Lipucci di Paola M, Sladowski D, Salmikangas P. Clinical Development and Commercialization of Advanced Therapy Medicinal Products in the European Union: How Are the Product Pipeline and Regulatory Framework Evolving? *Hum Gene Ther Clin Dev.* 2017;28(3):126–35.
- Shahyari A, Saghaeian Jazi M, Mohammadi S, Razavi Nikoo H, Nazari Z, Hosseini ES, Burtscher I, Mowla SJ, Lickert H. Development and Clinical Translation of Approved Gene Therapy Products for Genetic Disorders. *Front Genet.* 2019;10:868.

14. Pappas G, Berlin J, Avila-Tang E, Carroll J, Drozda J, Dumont D, Gross T, Hewitt K, Kirtane A, Kong D et al: Determining value of Coordinated Registry Networks (CRNs): a case of transcatheter valve therapies. *BMJ Surgery, Interventions, & Health Technologies* 2019, 1.
15. Yla-Herttuala S. Bumps in the Road for Commercial Gene Therapy for Rare Diseases. *Mol Ther*. 2017;25(10):2225.
16. Austin CP, Cuttillo CM, Lau LPL, Jonker AH, Rath A, Julkowska D, Thomson D, Terry SF, de Montleau B, Ardigo D, et al. Future of Rare Diseases Research 2017–2027: An IRDiRC Perspective. *Clin Transl Sci*. 2018;11(1):21–7.
17. Wang D, Tai PWL, Gao G. Adeno-associated virus vector as a platform for gene therapy delivery. *Nat Rev Drug Discov*. 2019;18(5):358–78.
18. Wang D, Zhang F, Gao G. CRISPR-Based Therapeutic Genome Editing: Strategies and In Vivo Delivery by AAV Vectors. *Cell*. 2020;181(1):136–50.
19. Fernandez-Marmiesse A, Gouveia S, Couce ML. NGS Technologies as a Turning Point in Rare Disease Research, Diagnosis and Treatment. *Curr Med Chem*. 2018;25(3):404–32.
20. Hmeljak J, Justice MJ. From gene to treatment: supporting rare disease translational research through model systems. *Dis Model Mech* 2019, 12(2).
21. New Report Finds Medical Treatments for Rare Diseases Account for Only 11% of US Drug Spending; Nearly 80% of Orphan Products Treat Rare Diseases Exclusively [<https://rarediseases.org/new-report-finds-medical-treatments-for-rare-diseases-account-for-only-11-of-us-drug-spending-nearly-80-of-orphan-products-treat-rare-diseases-exclusively/>].
22. Ferreira CR. The burden of rare diseases. *Am J Med Genet A*. 2019;179(6):885–92.
23. Korn EL, McShane LM, Freidlin B. Statistical challenges in the evaluation of treatments for small patient populations. *Sci Transl Med*. 2013;5(178):178sr173.
24. 'Major' challenges require new approaches to rare disease research. Tufts. <https://www.outsourcing-pharma.com/Article/2019/07/17/Rare-disease-challenges-require-new-approaches-Tufts>.
25. The National Economic Burden of Rare Disease Study. In. EveryLife Foundation for Rare Diseases; 2021.
26. Gopal-Srivastava R, Kaufmann P. Facilitating Clinical Studies in Rare Diseases. *Adv Exp Med Biol*. 2017;1031:125–40.
27. Fink AK, Loeffler DR, Marshall BC, Goss CH, Morgan WJ. Data that empower: The success and promise of CF patient registries. *Pediatr Pulmonol*. 2017;52(S48):44-s51.
28. Julkowska D, Austin CP, Cuttillo CM, Gancberg D, Hager C, Halftermeyer J, Jonker AH, Lau LPL, Norstedt I, Rath A, et al. The importance of international collaboration for rare diseases research: a European perspective. *Gene Ther*. 2017;24(9):562–71.
29. Blay JY, Coindre JM, Ducimetiere F, Ray-Coquard I. The value of research collaborations and consortia in rare cancers. *Lancet Oncol*. 2016;17(2):e62–9.
30. Godman B, Bucsis A, Vella Bonanno P, Oortwijn W, Rothe CC, Ferrario A, Bosselli S, Hill A, Martin AP, Simoens S, et al. Barriers for Access to New Medicines: Searching for the Balance Between Rising Costs and Limited Budgets. *Front Public Health*. 2018;6:328.
31. McCormack P, Kole A, Gainotti S, Mascalzoni D, Molster C, Lochmüller H, Woods S. 'You should at least ask': The expectations, hopes and fears of rare disease patients on large-scale data and biomaterial sharing for genomics research. *Eur J Hum Genet*. 2016;24(10):1403–8.
32. Tingley K, Coyle D, Graham ID, Chakraborty P, Wilson K, Potter BK. In collaboration with the Canadian Inherited Metabolic Diseases Research N: Stakeholder perspectives on clinical research related to therapies for rare diseases: therapeutic misconception and the value of research. *Orphanet J Rare Dis*. 2021;16(1):26.
33. Europe ERD. Rare disease patients 'participation in research: A Rare Barometer survey. In.; 2018.
34. FAIR Principles. <http://www.go-fair.org/fair-principles/>.
35. Charon R, Wyer P. Narrative evidence based medicine. *Lancet*. 2008;371(9609):296–7.
36. Pulciani S, Taruscio D. Patient-physician alliance: from Hippocrates to Post-Genomic Era. *Commentary. Ann Ist Super Sanita*. 2017;53(2):93–5.
37. Upadhaya S, Yu JX, Oliva C, Hooton M, Hodge J, Hubbard-Lucey VM. Impact of COVID-19 on oncology clinical trials. *Nat Rev Drug Discov*. 2020;19(6):376–7.
38. Schneeman EKK. Lessons Learned From COVID-19: Are there Silver Linings For Biomedical Innovation? In.; 2021.
39. Pierson SK, Stonestrom AJ, Shilling D, Ruth J, Nabel CS, Singh A, Ren Y, Stone K, Li H, van Rhee F, et al. Plasma proteomics identifies a 'chemokine storm' in idiopathic multicentric Castleman disease. *Am J Hematol*. 2018;93(7):902–12.
40. Fajgenbaum DC, Langan RA, Japp AS, Partridge HL, Pierson SK, Singh A, Arenas DJ, Ruth JR, Nabel CS, Stone K, et al. Identifying and targeting pathogenic PI3K/AKT/mTOR signaling in IL-6-blockade-refractory idiopathic multicentric Castleman disease. *J Clin Invest*. 2019;129(10):4451–63.
41. Pierson SK, Khor JS, Ziglar J, Liu A, Floess K, NaPier E, Gorzewski AM, Tamakloe MA, Powers V, Akhter F, et al. ACCELERATE: A Patient-Powered Natural History Study Design Enabling Clinical and Therapeutic Discoveries in a Rare Disorder. *Cell Rep Med*. 2020;1(9):100158.
42. Garcia M, Downs J, Russell A, Wang W. Impact of biobanks on research outcomes in rare diseases: a systematic review. *Orphanet J Rare Dis*. 2018;13(1):202.
43. Delude CM. Deep phenotyping: The details of disease. *Nature*. 2015;527(7576):14–5.
44. Kerr K, McAneney H, Smyth LJ, Bailie C, McKee S, McKnight AJ. A scoping review and proposed workflow for multi-omic rare disease research. *Orphanet J Rare Dis*. 2020;15(1):107.
45. TMB Results. The Future Use of Complex Biomarkers <https://friendsofcancerresearch.org/blog/engaging-innovation/tmb-results-future-use-complex-biomarkers>.
46. Let's Get Real. Harnessing Non-Proprietary Patient Registries and RWE to Accelerate Rare Disease Drug Development [<https://www.orphaniseasecenter.med.upenn.edu/september-recorded-sessions>].
47. FDA Patient-Focused Drug Development Guidance Series for Enhancing the Incorporation of the Patient's Voice in Medical Product Development and Regulatory Decision Making [<https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical>].
48. FDA Patient Listening Sessions. <https://www.fda.gov/patients/learn-about-fda-patient-engagement/fda-patient-listening-sessions>.
49. European Medicines Agency - Human Regulatory - Post-authorisation - Patient registries [<https://www.ema.europa.eu/en/human-regulatory/post-authorisation/patient-registries>].
50. Forsythe LP, Szydlowski V, Murad MH, Ip S, Wang Z, Elraiyah TA, Fleurence R, Hickam DH. A systematic review of approaches for engaging patients for research on rare diseases. *J Gen Intern Med*. 2014;29(Suppl 3):788–800.
51. Roennow A, Sauve M, Welling J, Riggs RJ, Kennedy AT, Galetti I, Brown E, Leite C, Gonzalez A, Portales Guiraud AP, et al. Collaboration between patient organisations and a clinical research sponsor in a rare disease condition: learnings from a community advisory board and best practice for future collaborations. *BMJ Open*. 2020;10(12):e039473.
52. Merkel PA, Manion M, Gopal-Srivastava R, Groft S, Jinnah HA, Robertson D, Krischer JP, Rare Diseases Clinical Research N. The partnership of patient advocacy groups and clinical investigators in the rare diseases clinical research network. *Orphanet J Rare Dis*. 2016;11(1):66.
53. Radu R, Hernandez-Ortega S, Borrega O, Palmeri A, Athanasiou D, Brooke N, Chapi I, Le Corvec A, Guglieri M, Perera-Lluna A, et al. Global Collaborative Social Network (Share4Rare) to Promote Citizen Science in Rare Disease Research: Platform Development Study. *JMIR Form Res*. 2021;5(3):e22695.
54. Administration USFD. Genetic Database Recognition Decision Summary for ClinGen Expert Curated Human Variant Data (Q181150). In.; 2019.
55. Whicher D, Philbin S, Aronson N. An overview of the impact of rare disease characteristics on research methodology. *Orphanet J Rare Dis*. 2018;13(1):14.
56. Sharpless NE, Kerlavage AR. The potential of AI in cancer care and research. *Biochim Biophys Acta Rev Cancer*. 2021;1876(1):188573.
57. Naudet F, Siebert M, Pellen C, Gaba J, Axfors C, Cristea I, Danchev V, Mansmann U, Ohmann C, Wallach JD, et al. Medical journal requirements for clinical trial data sharing: Ripe for improvement. *PLoS Med*. 2021;18(10):e1003844.
58. Leading Real-World Data and Analytics Organizations Form Industry. Coalition to Advance Policies to Support Regulatory Use of Real-World Evidence <https://www.businesswire.com/news/home/20210519005219/en/Leading-Real-World-Data-and-Analytics-Organizations-Form-Industry-Coalition-to-Advance-Policies-to-Support-Regulatory-Use-of-Real-World-Evidence>.
59. Government US: 21st Century Cures 2.0 Discussion Draft. In.; 2021.
60. Rieke N, Hancox J, Li W, Milletari F, Roth HR, Albarqouni S, Bakas S, Galtier MN, Landman BA, Maier-Hein K, et al. The future of digital health with federated learning. *npj Digit Med*. 2020;3(1):119.
61. Moses H 3rd, Matheson DH, Cairns-Smith S, George BP, Palisch C, Dorsey ER. The anatomy of medical research: US and international comparisons. *JAMA*. 2015;313(2):174–89.
62. Cox GF. The art and science of choosing efficacy endpoints for rare disease clinical trials. *Am J Med Genet A*. 2018;176(4):759–72.



63. National Evaluation System for Health Technology (NEST). <https://www.fda.gov/about-fda/cdrh-reports/national-evaluation-system-health-technology-nest>.
64. Federal Health IT. Strategic Plan [[https://www.healthit.gov/sites/default/files/page/2020-01/2020-2025FederalHealthIT%20StrategicPlan\\_0.pdf](https://www.healthit.gov/sites/default/files/page/2020-01/2020-2025FederalHealthIT%20StrategicPlan_0.pdf)].
65. Nguyen MT, Goldblatt J, Isasi R, Jagut M, Jonker AH, Kaufmann P, Ouillade L, Molnar-Gabor F, Shabani M, Sid E, et al. Model consent clauses for rare disease research. *BMC Med Ethics*. 2019;20(1):55.
66. Darquy S, Moutel G, Lapointe AS, D'Audiffret D, Champagnat J, Guerroui S, Vendeville ML, Boespflug-Tanguy O, Duchange N. Patient/family views on

data sharing in rare diseases: study in the European LeukoTreat project. *Eur J Hum Genet*. 2016;24(3):338–43.

### **Publisher's note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.