Critical Bottlenecks in Rare Disease Research and Care: A Community Perspective

doi:10.5281/zenodo.14906643

Julie McMurry (Monarch, UNC, Alamya Health), Alison Sizer (Growth Impact), Robert Allaway (Sage Bionetworks), Cornelius Boerkoel (BC Children's, AlamyaHealth), AJ Chen (ELHS Institute), Jason Colquitt (Across), James Cummings (Consumer), Autri Dutta (Genomenon), Nasha Fitter (Citizen Health), Adam Hansen (Geneial), Eric Harker (Ada), Collin Hovinga (C-Path), Ari Kahn (UTexas), Hans Keil (Nurture Genomics), Rodger Kessler (DARTNet Institute), Lukas Lange (Probably Genetic), Arturo Loaiza-Bonilla (Massive Bio), Natalia Luchkina (EverythingALS), Eric Luellen (Turing Biosciences), Arezoo Movaghar (Wake Forest University), Tomi Pastinen (Children's Mercy), Elizabeth Rountree (COMBINEDBrain), Vivek Rudrapatna (UCSF), Adam Sand (Selfii), Kat Schmolly (zebraMD, UCSF), Patrick Short (Sano Genetics), Geoffrey Siwo (UMichigan), Marina Sirota (UCSF), Nicholas Tattonetti (Cedars Sinai), Farid Vij (Citizen Health), Samuel Volchenboum (UChicago), Anita Walden (UNC), Ramona Walls (C-Path), Peter Washington (UCSF), Dylan Wenzlau (Guava), Margaret Wenzlau (Truveta), Matthew T Wheeler (Stanford), Natan Vidra (Anote), Charlene Son Rigby (Global Genes), Melissa Haendel (Monarch, UNC, Alamya Health)

Introduction

The rare disease community faces significant challenges in advancing research, improving diagnosis, and delivering effective healthcare. According to the National Economic Burden of Rare Disease Study, more than >10,000 rare diseases affect 30 million Americans, half of whom are children. With a total economic burden of nearly $1 trillion, rare diseases present a massive opportunity for innovation in diagnostic efficiency and treatment development.

On January 23, 2025, an extraordinary demonstration of the rare disease community's collaborative spirit emerged as a result of unexpected circumstances. When a US government communication ban led to the last-minute cancellation of a planned ARPA-H RAPID meeting, the community didn't disperse - we came together. RAPID is the first significant opportunity of its kind being funded in the US to reduce the diagnostic odyssey for rare disease patients. We, the potential applicants, self-organized to hold our own meeting, setting aside competitive interests to focus on the fundamental challenges facing rare disease research and care.

This impromptu gathering was more than just a discussion of funding opportunities; it became a powerful example of how shared commitment to improving patient outcomes can transcend institutional boundaries and administrative hurdles. Our discussions revealed that while the community and the current funding opportunity shared noble goals, there is potential to better align priorities with the field's most pressing needs well beyond the funding opportunity. This white paper synthesizes perspectives from healthcare providers, academic researchers, industry experts, registry providers, and people with lived experience to identify critical bottlenecks that must be addressed to accelerate progress in the rare disease field. Here, we share the collective insights from the meeting - a testament to the rare disease community's resilience, collaboration, and unwavering focus on patient impact.

The Data Integration Challenge: From Fragmentation to Unity and Utility

Fragmentation. On average, rare disease patients see between 4-6 different providers, limiting the utility of focusing on any one center for data collection. This dispersed care creates significant challenges in data collection and coordination. Medical information becomes fragmented across multiple sources, from electronic health records and clinical trials to research databases and patient registries. The healthcare system struggles to connect these scattered pieces, failing to integrate various types of patient data - clinical observations, genetic information, medical imaging, patient feedback, and lifestyle factors - that together could provide a complete picture of a patient's health. Health records of rare disease patients often fall short in several critical ways. Clinicians may interpret patient narratives differently, longitudinal data tracking is incomplete, and symptom documentation lacks consistency. Quality of life metrics are rarely captured, family history information is often missing, and standardized coding for rare conditions is lacking. The traditional scientific publication process presents additional challenges. Publications are slow to share crucial findings, often providing only a snapshot of the typical patient experience rather than the detailed information needed to develop accurate diagnostic tools. These limitations extend beyond rare diseases, affecting our understanding of complex relationships between genetics and symptoms, as well as broader cause-and-effect patterns across all diseases. The fragmentation becomes even more apparent when considering the vast network required for rare disease care - multiple providers, healthcare facilities, and specialized entities must coordinate to diagnose, treat, and monitor these conditions effectively.

Critical role of patients and patient partnership. Our limited understanding of how rare diseases progress over time further compounds challenges and disease prognosis is often extremely limited. Patients are an under-utilized source of information both about their particular disease journey and knowledge about the disease more broadly; patients are the key to improving diagnosis and care for everyone. However, when patient-reported data is collected at all, the questions are often off-target and too limited both in scope and in temporality. Without tracking patients' experiences throughout their disease journey, we are missing vital clues about potential biological markers, making it harder to identify reliable indicators of disease severity, progression, quality of life, or treatment effectiveness. As a result, both diagnosis and treatment development move more slowly than they could, leaving patients waiting longer for effective care. It is also harder to properly design clinical trials when we don't clearly understand how diseases unfold and their variability in doing so.

Health information registries will be crucial in this process through several key mechanisms. Rare communities display the most advanced behavior of participatory responsibility, willingness to share, and urgency, providing a unique window into what universal consumer behavior could be. The data must flow back to the patient since patients and patient groups - not researchers or pharma - are often the ones driving their care and creating treatments. For people with rare diseases, research IS care.

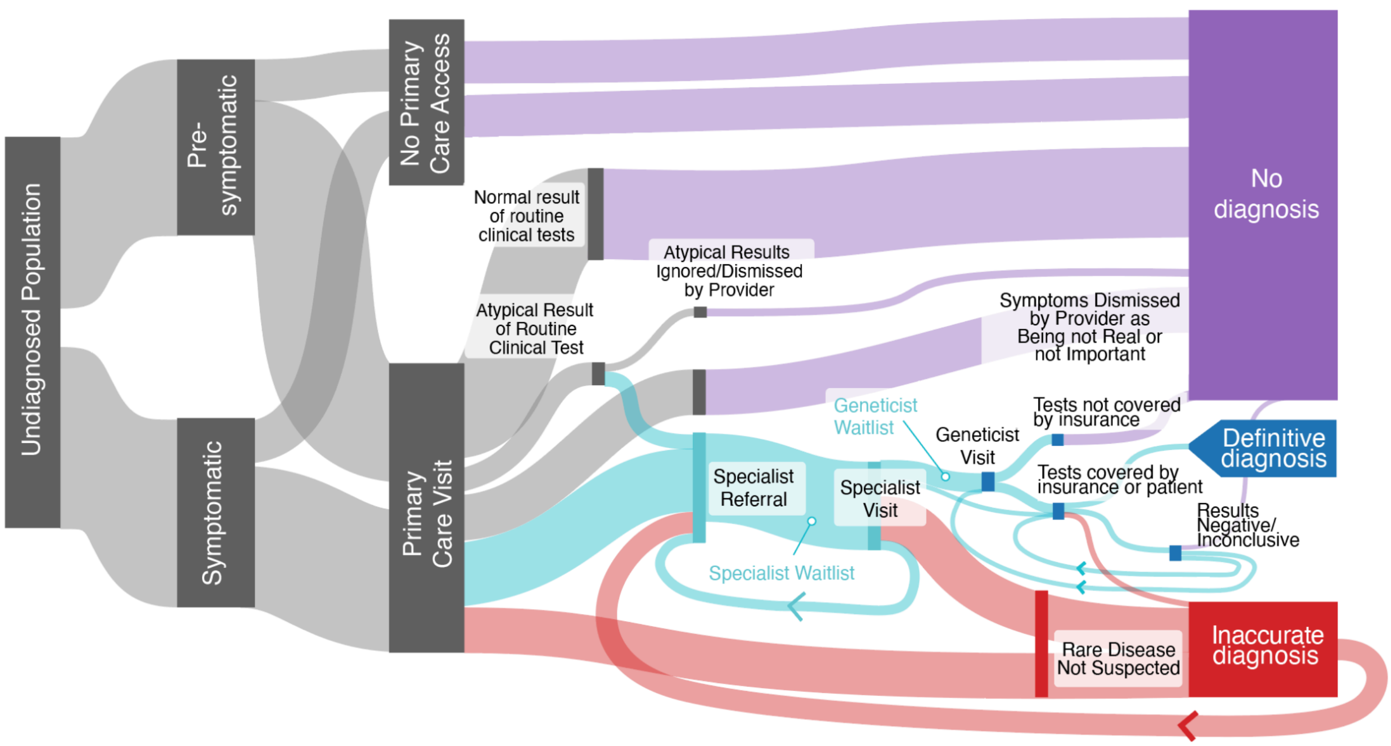
Different data modalities and lack of standardization. Data fragmentation is compounded by varying data standards and terminologies, creating significant bottlenecks to meaningful analysis and research. The challenge extends beyond mere standardization to the integration of multiple data modalities, including context and patient perspectives. Current EHR systems are not designed to integrate data — whether multiple modalities from an individual, or across individuals. All the data matter — whether clinical or research data: laboratory tests, phenotypes, genomic profiles, imaging data, patient-reported outcomes, critical symptomology documented in unstructured notes, as well as data collected from digital tools and wearables.

To address this, we need robust technical infrastructure development, including secure application programming interfaces (APIs), real-time data integration and cloud-based solutions for processing multiple data types and maintaining provenance. Deploying population-wide digital pre-screening that leverages multi-source/type data integration and AI-based triage for rare diseases is critical to improving time-to-diagnosis and therefore outcomes. Such methods can flag potential rare disease cases early and streamline patient referrals to specialized care or research studies. This approach reduces data fragmentation and fosters real-time insights, highlighting the importance of large-scale data harmonization and interoperability for effective patient-centric solutions. Centers of excellence should be networked with standardized data collection - and dissemination protocols - providing the framework needed to make prospective integration a reality. Similarly, legacy data and knowledge need to be integrated and made available for meaningful insights. Systems hosting rare disease data will need to become more standardized, use exchange mechanisms such as FHIR (Fast Healthcare Interoperability Resources), and participate in information exchange networks like TEFCA (Trusted Exchange Framework and Common Agreement). Additionally, these advancements will make it easier for patients to access and engage with their health data, ultimately creating a beneficial feedback loop between patients, healthcare providers, and researchers.

Lack of standardized and scalable technical infrastructure. Today, vital rare disease data collection and analytics happens primarily in custom systems outside of the EHR; to truly transform care and drive translational research, we need these systems to better integrate with each other and the EHR to enable real-time decision-making. The lack of interoperable systems also results in limited large-scale collaborative analytics and reuse of one another’s artifacts, insufficient tools for federated learning, significant challenges in defining or maintaining data quality at scale, and a need for improved, standardized processing of clinical notes and other data types such as signal data, imaging, and wearables. The solution is to build scalable data processing systems with user-friendly interfaces and automated harmonization tools. FHIR-based integration capabilities and standardized APIs can provide the foundation for system interoperability but need more sophistication for ensuring basic standardization across systems and contexts as well as managing new complex data types. While it is critical to have automated data extraction tools to maintain data at scale, data quality and lack of interoperability are frequently overlooked, especially in federated systems. Federated systems should include sampling and data compliance technologies to ensure data interoperability where system interoperability or data exchange is not permissible.

Diagnostic Delays: Building Bridges to Earlier Detection and Treatment Access

The diagnostic journey for rare disease patients remains arduous, characterized by non-specific initial symptoms, limited rare disease knowledge among primary care physicians, delays in appropriate testing and referrals, and insufficient insurance coverage or access to specialists. The [average number of years it takes for a rare disease diagnosis is 4-5 years] or longer. Only about [10% of rare diseases have an FDA-approved therapy] for their treatment, underscoring the imperative for earlier and more accurate diagnoses. These challenges are shown qualitatively below (there are no data available upon which to base scale). For illustration purposes, diagnoses are simplified into missing, wrong, and definitive; however, in reality, diagnoses occur on a continuum of probabilities that can be refined with additional information.



Realizing public health surveillance and improving diagnostic access. To accelerate diagnosis, we need a multi-faceted approach that combines technological innovation with human expertise. To drive the cost down for the whole diagnosis journey we need age-appropriate risk-based screening in routine care (Primary Care, Community Health Centers, and rural clinics) followed by referral to the right specialists in fewer "hops". Fine-tuned algorithms can predict some rare diseases from early indicators (symptoms, routine blood tests, and patient-reported data) even before any genetic tests. Importantly, this model provides a single 'front door' for patients, reducing multiple referral steps and aligning clinical trial matching or advanced diagnostics into a unified care pathway. This 'front door' strategy has been implemented by various clinical AI platforms — including those focusing on rare cancer subtypes and hereditary conditions — to ensure patients enter an appropriate diagnostic and treatment pipeline as early as possible. There are gaps in insurance coverage of advanced diagnostic technologies and specialist care. We recommend mobile diagnostic units for rural areas, telemedicine programs for specialist consultation, community health workers trained in rare diseases, and sliding-scale payment programs for diagnostic testing to bridge these gaps.

Opportunities for digital health technologies. Technologies such as wearables (e.g., smartwatches, biosensors), nearables (e.g., ambient sensors), sensor-equipped tools (e.g., smartphones for passive data collection), combined with self-reported symptoms and digital diaries, offer significant advantages in rare disease diagnosis and management. Multimodal data, such as movement patterns, speech, digital cognitive measures and others, combined with physiological markers can reveal novel disease markers previously undetectable. For rare diseases with small patient populations, this approach, combined with data science and AI methods, has the potential to accelerate the development of diagnostic and progression algorithms by using real-world evidence at scale.

Opportunities for Large Language Models and AI. While LLMs hold great promise for medicine, their application to rare disease diagnostics requires cautious application. Current best-in-class LLMs often struggle with strict rule adherence and may generate biologically implausible recommendations. This is particularly concerning in clinical rare disease contexts, where the margin for error is small. Rather than positioning LLMs as diagnostic engines, their strength lies in augmenting expert review — highlighting potential research gaps, summarizing complex data, supporting hypothesis development, and enhancing information from alternative sources such as a parent sharing information on social media. LLMs can also facilitate standards adoption through supporting data curation and integration. We can provide all care providers and consumer interfaces with low-cost private and secure GenAI-enabled screening tools to find at-risk patients and refer them to specialists.

Leveraging Biology. Beyond disease-centric frameworks, we must shift toward adoption of standards that help build mechanism-centric models that uncover shared biological pathways across multiple rare diseases. This approach enables the development of platform therapeutics targeting underlying mechanisms rather than individual conditions, maximizing efficiency and impact.

Enhancing recruitment to trials. Improved diagnostic rates also provide patients with earlier access to care and clinical trials. The ability to identify larger patient populations at early disease stages can greatly increase the statistical power of trials, accelerating equitable access to both novel and existing treatments. This not only attracts greater R&D investment from pharmaceutical companies but also improves the success rate of clinical trials and the approval of new drugs. With only about 10% of rare diseases having an FDA-approved therapy, this underscores the critical need for earlier and accurate diagnoses to address these unmet medical needs. Precision medicine trials also often face a 'catch-22' where lack of access to diagnostics in the healthcare system makes patient identification, and accurate prevalence estimates a challenge. This adds further uncertainty to developing treatments for patient populations that are not yet well served. Introducing clinical trials as a care option can bridge the gap between research and clinical care, providing patients with advanced and ground-breaking medical treatments.

Enhancing education. To tackle the aforementioned challenges, the next generation of researchers, clinicians and healthcare providers, data scientists, physician-scientists, regulatory experts, hospitalists, and policymakers must be trained. We also need to provide free patient education and assist at-risk patients in navigating the complex process of rare disease diagnosis and treatment.

Governance: Creating a Sustainable Framework

A robust and modern governance framework is needed to address multiple challenges: creating interoperable data and systems, sharing information across international boundaries (critical for rare diseases), managing varying regulatory requirements, ensuring robust privacy and security, implementing ethical frameworks for AI, and maintaining patient control over personal health information. Central for such a framework are advanced encryption, privacy-preserving record linkage, and federated learning capabilities.

Success requires standardized data use agreements, clear protocols for patient consent management, and templates for international data sharing. This infrastructure would furthermore provide a platform for incentivizing this critical data sharing; as too often, data that is held by organizations and individuals is guarded or restricted in data silos. A successful data-sharing framework would ensure that providers of data, including patients, are encouraged to share data, such as by receiving a portion of profits, access to a broader dataset, or advanced analytics.

Additionally, rethinking data ownership through patient-centric equity models that enable data dignity—where individuals directly benefit from discoveries rooted in their health data—can create sustainable incentives for data sharing. This paradigm not only fosters collaboration but also aligns ethical considerations with scientific advancement.

Emerging AI technologies can also be leveraged to optimize clinical trial design and accelerate approval timelines. By modeling regulatory decision-making patterns and identifying key risk factors early, AI can help de-risk development pathways and streamline submissions. This capability is particularly valuable in rare disease drug development, where small patient populations and limited precedent create additional regulatory uncertainty. Integrating predictive AI into the regulatory strategy can thus serve as both a scientific and operational advantage.

Regulatory inertia can delay the translation of innovative diagnostics and treatments in part because regulatory frameworks used for more common diseases are frequently inappropriate for rare diseases. Agencies like the FDA often struggle to keep pace with the rapid evolution of computational models and have been reluctant to accept them as primary evidence for therapeutic plausibility. To foster progress, there must be a parallel effort to modernize regulatory pathways.

Resource Requirements and Success Metrics

We will ultimately demonstrate a financially sustainable model for continuous improvement in rare disease diagnostic efficiency and treatment development and a flywheel of continued investment in and realization of value from precision medicine. As a field, we need to empirically demonstrate the clinical and economic outcomes of actual efforts to responsibly apply AI to accelerate rare disease diagnosis and care.

Demonstrating success will require a) substantial human capital, including data scientists, clinical research coordinators, patient engagement specialists, patient groups, health economists, regulatory specialists, and technical infrastructure engineers; b) Infrastructure needs to encompass cloud computing resources, secure data storage systems, integration platforms, and mobile diagnostic equipment; and c) U.S. Congress to mandate changes in insurance policies, EHR vendor support, and terminological implementation and reporting.

A delay in diagnosing a condition can allow the disease to advance beyond the treatment window. This leads to poorer outcomes for patients, as they face higher chances of dying or becoming disabled from otherwise preventable complications. Additionally, such delays often make treatments more complex and challenging to implement. Success metrics for each step (defined in the diagnostic bottlenecks figure above) could include, for example, facets relating to data interoperability, infrastructure, governance, and systems. Progress can be measured through:

Quantitative metrics such as reduction in medical costs (inpatient hospital or outpatient care, physician visits, medications), reduction in time to diagnosis, increase in data completeness scores, reduction in time between data generation and availability for reuse, number of institutions adopting standards, number of new treatments in trials and approved) and

Qualitative indicators (improved patient satisfaction, enhanced physician confidence in diagnosis, stronger research collaboration networks, increased international collaboration). Patient-generated metrics are critical, such as quality of life, access to care, and qualitative measures regarding healthcare engagement.

These solutions will need investment from varied sources, including government grants, industry partnerships, foundation support, and healthcare systems.

Recommendations to funding agencies, philanthropic organizations, and researchers

Several takeaways came about from the meeting. Although rare disease is a resource-poor area of medicine, there are passionate and dedicated people who — with modest resources — have already built some of the key components needed to transform the landscape. The RAPID funding opportunity could sustain some of these components and help them work optimally together. To proceed, we need systems-level thinking, an interconnected approach that enables sustainable transformation for the diagnosis and care of patients with rare diseases. Such systems-level approaches will be critical to the success of the elements within the RAPID program but also require the RAPID program to contribute to broad systematic change. Specific recommendations include the following:

- Develop a shared infrastructure to identify and define new diseases
- Advance omics as a true standard of care
- Develop platform approaches to therapy development for ultra-rare diseases, shared among pharma, nonprofits, FDA, and academia.

Conclusion

The challenges facing the rare disease community are significant but are neither unique nor insurmountable. The self-organized meeting amongst scientists, geneticists, researchers, technologists, and innovators, demonstrates the rare disease community's commitment to collaboration and innovation. Much like how the internet's shared protocols or the metric system revolutionized global connectivity and trade, unified rare disease data standards and processes will link fragmented cohorts into robust networks, streamline research workflows, and accelerate breakthroughs for precision medicine that provide the right treatment at the right time. By implementing these solutions to unify data, accelerate diagnosis, and establish interoperability and governance in a phased, systems-level approach that extends beyond RAPID, we will create diagnostic efficiency and treatment development for rare disease research and care. Success will require sustained stakeholder commitment, and create positive impact in patient outcomes, health systems, and reduce economic burden.