Redefining Rigor: Fit-for-Purpose Trials to Unlock Rare Disease Therapies

July 22, 2025

Patient Advocate Authors (listed alphabetically)

Kara K. Berasi, PharmD, MS, CEO, Haystack Project
Jocelyn Duff, Executive Director, CureCMT4J
Todd Durham, Sr. VP, Clinical & Outcomes Research, Foundation Fighting Blindness
Brandi Fesler, CEO, Sarcoma Foundation of America
H. Eric Hartman, Director of Advocacy, Choroideremia Research Foundation
Chris Moen, MD, President & CMO, Choroideremia Research Foundation
Tricha Shivas, MBe, Chief of Strategy, Foundation for Sarcoidosis Research
Dennis Sprecher, MD, Drug Development Project Lead, Cure HHT Foundation
Kristen Wheeden, MBA, President, United Porphyrias Association
Christine Zahn, Founder & Director, Arginase 1 Deficiency Foundation

Clinical and Academic Authors (listed alphabetically)

Shein-Chung Chow, PhD, Professor, Department of Biostatistics and Bioinformatics, Duke University School of Medicine
Amy Dickey, MD, Assistant Professor of Medicine, Harvard Medical School & Massachusetts General Hospital
Gregory M. Enns, MD, Director, Biochemical Genetics Program, Stanford University
Andrea Gropman, MD, Director Neurometabolic Translational Research, Center for Experimental Neurotherapeutics, St. Jude Children's Research Hospital
Byron L. Lam, MD, Medical Director of Clinical Research, Bascom Palmer Eye Institute, University of Miami
William D. Tap, MD, Chief of the Sarcoma Medical Oncology Service, Memorial Sloan Kettering Cancer Center
Jerry Vockley, MD, PhD, Chief of Genetic and Genomic Medicine, UPMC Children's Hospital of Pittsburgh
Janet Woodcock, MD, Board Member, Haystack Project, Former FDA Principal Deputy Commissioner

The authors appreciate the contributions of the following individuals (listed alphabetically) for their insights and participation in the March 12, 2025 Scientific Workshop:

Aniz Girach, Fmr. CMO, Nightstar Therapeutics
Sanjay Jalota, Fmr. Vice President, Global Regulatory Affairs, Aeglea BioTherapeutics
Allen Melemed, MD, Fmr. Senior Medical Director, Eli Lilly
Marshall Summar, MD, CEO, Uncommon Cures

Prepared for Haystack Project by:

Connect 4 Strategies (Saira Sultan, JD) and
Consilium Strategies (M. Kay Scanlan, JD)

About Haystack Project

Haystack Project (Haystack) is a 501(c)(3) non-profit organization enabling our growing membership of rare and ultra-rare disease and rare cancer patient advocacy organizations to coordinate and focus efforts to resolve systemic obstacles to innovation and access. Its core mission is to spur innovation and quality in care toward effective, accessible, and affordable treatment options for all Americans, regardless of the rarity of their condition. We strive to amplify the patient and caregiver voice in these disease states where unmet need is high and treatment delays and inadequacies can be catastrophic.

Advocacy is at the center of Haystack Project's mission. The Advocacy Empowerment Institute (AEI) arms Haystack Project member organizations with tools for effective self-advocacy and health policy navigation. Through a set of programs and knowledge-sharing workshops, Haystack enhances collaboration and information exchange among patient advocacy organizations in the rare and ultra-rare disease communities.

Haystack Project's advocacy efforts were also a driving force behind the Helping Experts Accelerate Rare Treatments (HEART) Act, a bipartisan bill introduced in 2022 to increase rare disease expertise and patient input in the FDA drug approval process. The HEART Act required the FDA and the Government Accountability Office (GAO) to examine and report on how the FDA incorporates patient experience data, consults rare disease experts, and encourages the Agency to include patient-identified experts in review meetings.
In 2024, Haystack Project renewed its focus on policy discussions and outreach efforts to highlight and ultimately resolve the challenges researchers face in bringing promising treatments for rare and ultra-rare diseases to the patients who need them.

Executive Summary

An estimated 30 million Americans suffer from a rare disease.[1] Decades of investment in biomedical research have revolutionized our understanding of human disease. Academic laboratories have identified the genetic and/or biochemical basis of hundreds of rare diseases and laid the groundwork for developing new treatment approaches. While these scientific advances have fueled hope within the rare disease patient communities, 95% of the approximately 10,000 rare diseases identified to date remain without a Food and Drug Administration (FDA) approved treatment option.[2]

Over the past several years, patients have seen over 1,000 rare disease research and development (R&D) programs put on the shelf[3] – many before they reach the human testing phase. Several Haystack Project patient advocacy organizations have experienced the disappointment and frustration of seeing that a promising investigational treatment will likely never reach patients. It is particularly troubling when the treatment candidate delivers tangible improvements in patient lives yet fails to meet primary endpoints and/or a research sponsor receives a Refusal to File or Complete Response Letter (CRL).

"We are having to rely on standard RCT designs that do not yield well-controlled studies. Either you commit to a prolonged study or you risk inconclusive data or a study failure.

And it is important to remember that these patients are often getting to the same place - mortality, disability, permanent organ damage - as with the rapidly progressive conditions where we would not do an RCT due to ethics concerns."*

--Workshop Panelist

* Note: Haystack Project convened its Scientific Workshop under the Chatham House Rule. This enabled us to capture the full value of the discussion and share insights without attribution.

[1] Wan EL, Elkaim Y, Gao W, Yoon R. Zebras Among Us: Advocating for the 30 Million Americans Living with Rare Disease. Med Sci Educ. 2023 Aug 15;33(5):1239-1242. doi: 10.1007/s40670-023-01856-2. PMID: 37886282; PMCID: PMC10597899.

[2] Gürkan H, Bilge Satkın N. The Importance of Genetic Diagnosis in Rare Diseases. Balkan Med J. 2025 Mar 3;42(2):92-93. doi: 10.4274/balkanmedj.galenos.2025.2025-270125. PMID: 40033553; PMCID: PMC11881545.

[33] Developing Drugs for Rare Diseases: A New Approach to Generating Clinical Evidence

The FDA's application of the "substantial evidence"[4] standard appears grounded in the assumption that the statutory requirement for "adequate and well-controlled investigations" is a mandate for randomized, placebo-controlled study designs. Our communities rely on the FDA to minimize the possibility of patient exposure to unsafe and/or ineffective drugs while enabling research that translates scientific advances into tangible improvements in available treatment options. Haystack Project is concerned that progress toward addressing unmet needs in rare diseases will be needlessly slowed unless regulators and policy leaders recognize that randomization is often an inadequate method for yielding a well-controlled study in rare diseases. This is particularly true within the context of interventional studies involving extremely small-population diseases with slow progression and/or moderate to high heterogeneity. In these diseases, an RCT might take decades to demonstrate a statistically significant benefit (p value below .05) – and that is ***if*** the endpoints are appropriate for the condition and enrollment is preserved over the course of the study.

Similarly, the Agency's efforts to leverage its review experience to ensure predictability and uniformity on study endpoints often ignore condition-specific nuances in rare disease mechanisms and complexity. Study designs and expectations that are not "fit for purpose," i.e., crafted within the context of the specific rare disease and its patient population, compound the risk that a treatment does not have a realistic path to FDA approval.

Over the past several years, Haystack Project has sought to gain a more granular understanding of the regulatory hurdles to rare disease R&D and identify pragmatic recommendations that align with and augment FDA's rare disease initiatives and maintain the scientific rigor our patient communities expect. These efforts included advocacy toward enactment of the Helping Experts Accelerate Rare Treatments (HEART) Act, which mandated reports on (1) FDA's strategies to ensure reviewers have the necessary expertise and use appropriate flexibility in their rare disease reviews,[5] and (2) learnings from the European Medicines Agency rare disease review mechanisms.[6]

In March 2025, Haystack Project convened a multi-stakeholder panel to further inform our work with FDA. The Scientific Workshop discussion centered on several case studies of rare disease R&D efforts that ultimately failed to clear FDA's evidentiary bar.

---

[4] Section 505(d) of the FD&C Act

[5] GAO-25-106774, RARE DISEASE DRUGS: FDA Has Steps Underway to Strengthen Coordination of Activities Supporting Drug Development

[6] Regulatory Processes for Rare Disease Drugs in the United States and European Union: Flexibilities and Collaborative Opportunities | The National Academies Press

In order to promote a candid discussion, Haystack Project convened the Workshop under the Chatham House Rule. This enabled us to capture the full value of the discussion and share insights without attribution. The discussion revealed common challenges including:

- Endpoints for clinical studies are often derived from FDA's prior experience reviewing treatments with different mechanisms of action impacting the same body system.
- Heterogeneity in disease symptoms can make it difficult to identify a single primary endpoint. Although studies may demonstrate positive results on secondary endpoints, it is difficult to power rare disease studies to demonstrate statistically significant results on multiple endpoints.
- Many rare diseases do not have enough patients to achieve an adequately powered RCT. These studies will fail unless the intervention is a "cure" -- something that rarely happens in drug development."
- The challenges research sponsors face are set in stone early in research programs. FDA does not include disease-specific expert opinion on what it would take to convince a disease "expert" that a specific treatment candidate is effective and/or that clinical effect from a clinical trial was due to the intervention.
- FDA's use of Real World Evidence (RWE) to augment clinical study data is not consistent.
- Patients and patient advocacy organizations that have engaged with FDA on specific R&D programs do not believe their input on meaningful outcomes impacts study design.
- The FDA is often hesitant to accept use of biomarkers as surrogate endpoints. Both researchers and study sponsors express a lack of clarity on what FDA requires in a "validated" biomarker.

Throughout the Scientific Workshop discussions, panel members and participants focused beyond "what went wrong" to consider "what could have been done differently" and, finally, "how do we get from where we are to where we need to be." These context-driven discussions yielded a set of recommendations that move FDA away from a "feasibility" approach that relies on flexibilities as exceptions to rules requiring RCTs and/or endpoints derived from FDA reviews of other conditions. The recommendations in this Report were informed by the Haystack Project Scientific Workshop as well as the GAO[7] and NASEM[8] Reports (Appendix 1 provides key insights from the reports). They are grounded in experience and were developed with the primary goal of ensuring that small population studies are both adequate and well-controlled.

[7] Rare Disease Drugs: FDA Has Steps Underway to Strengthen Coordination of Activities Supporting Drug Development, GAO-25-106774, Published: Nov 18, 2024.

[8] Processes to Evaluate the Safety and Efficacy of Drugs for Rare Diseases or Conditions in the United States and the European Union | National Academies

- Clarify an operational definition of the statutory term “well-controlled investigations” to reflect the rare disease reality that randomization may yield a poorly controlled study.
- Replace “regulatory flexibility” discretion with clear authority to consider alternative methodologies and innovative study designs that can constitute well-controlled investigations
- Reduce the burden and unpredictability associated with biomarkers as surrogate endpoints.
- Emphasize the Importance of disease-specific expertise in rare disease study design, endpoint selection, data analysis, and evaluation.
- Move rare disease R&D program frameworks away from the “hypothesis testing” approach that tends to create unattainable study goals in favor of alternative statistical analysis plans.

Introduction

Innovators seeking to navigate the regulatory framework toward approval of rare disease treatments face unique challenges, including limited patient populations, disease heterogeneity, and limited natural history data. These factors make the regulatory path for rare disease treatments unpredictable, costly and for many candidate therapies, so risky that it is simply not feasible. While basic science flourishes and promising treatment options seem within reach, the ever-growing number of shelved development programs is a clear signal. Halted programs represent more than a financial loss for sponsors. These are lost opportunities for patients with few, if any, alternatives; they highlight the need for new approaches that reduce the risks and uncertainties in rare disease research and development (R&D).

Multiple factors contribute to the failure of rare disease candidates to clear regulatory hurdles and discontinuation of industry-sponsored rare and ultra-rare disease drug programs. The bottom line is that these projects face unique scientific and logistical challenges that make them far more fragile than development efforts for common diseases. By understanding why past programs have been halted – whether due to drug failure, study failure, or business decisions – stakeholders can identify refinements likely to move the needle on delivering much-needed therapies to the millions of patients with rare disorders.

This Report was developed in parallel with Haystack Project's PROTECT Rare initiative, and builds on Haystack's efforts toward enacting and implementing the Helping Experts Accelerate Rare Treatments (HEART) Act. We have focused on empowering FDA with a framework that augments the set of "regulatory flexibilities" for rare disease interventions with clear authority to incorporate information from disease-specific experts early and throughout its processes and look beyond the RCT, traditional statistical analyses, and Agency precedent when determining that a study design is "well-controlled" and its endpoints are likely associated with clinical benefit.

Haystack Project convened a multi-stakeholder Scientific Workshop to better understand the challenges rare disease patients, researchers, and study sponsors encounter. Participants also identified a set of recommendations that would maintain (rather than compromise) scientific rigor through study designs that are fit-for-purpose within the context of the specific condition and the tested intervention.

In this Report, Haystack Project provides background information detailing how current paradigms for evaluating an intervention's safety and effectiveness are compounding the inherent challenges to R&D efforts in rare, and particularly in ultra-rare, conditions. We

provide a brief overview of the Scientific Workshop and then leverage the case studies to offer recommendations addressing the regulatory barriers identified throughout the workshop.

Background:

Rare disease innovators face substantial challenges navigating a viable regulatory pathway to approval.

Approximately 10,000 rare diseases have been identified to date, a significant portion (70%) of which emerge in childhood and have a profound impact on a child's health and well-being, often leading to chronic illnesses and premature death. Hundreds of orphan drugs have been approved since enactment of The Orphan Drug Act of 1983, and countless lives saved or significantly improved. Unfortunately, approximately 95% of rare diseases remain without an FDA-approved treatment option. This massive treatment gap underscores the formidable R&D challenges innovators face in converting rare disease research into new, accessible treatments.

Of the thousands of orphan drug designations granted in the U.S., only a small fraction (16%) have resulted in an approved treatment. Approximately 23% have been *classified as inactive* (development halted without approval). The remainder are still in progress or undetermined. The attrition rate for rare disease research is high - over half of rare disease trials started in the 2010s ended up discontinued or unpublished (i.e., did not yield actionable results).

Common challenges that pervade R&D for rare diseases include:

Very Small Patient Populations Impede Trial Recruitment: This makes it difficult to reach sufficient enrollment to achieve results meeting FDA's standard for statistical significance. This is the biggest hurdle in rare disease trials. In one analysis, insufficient patient accrual accounted for ~31% of discontinued rare disease trials.[9] Ultra-rare diseases are especially prone to recruitment failure.

Heterogeneity in Phenotype and Genotype: Rare diseases can have significant genotypic and phenotypic heterogeneity; patients with the same disease diagnosis might have very different mutations, symptom profiles, or disease progression. Studies designed as traditional one-size-fits-all clinical trials often fail to capture the intervention's clinical

[9] Rees CA, Pica N, Monuteaux MC, Bourgeois FT. Noncompletion and nonpublication of trials studying rare diseases: A cross-sectional analysis. PLoS Med. 2019 Nov 21;16(11):e1002966. doi: 10.1371/journal.pmed.1002966. PMID: 31751330; PMCID: PMC6871779.

benefit (or lack thereof). According to FDA's "LEADER 3D: Learning and Education to Advance and Empower Rare Disease Drug Developers PUBLIC REPORT OF EXTERNAL STAKEHOLDER ANALYSIS,"[10] interviewees described how heterogeneity in certain rare diseases complicates finding any one endpoint or surrogate that works for every patient.

Limited Scientific Knowledge and Natural History Data: For many rare diseases, the underlying biology and disease progression have not reached the level of broad scientific understanding needed to meet FDA's requirements on appropriate endpoints. Similarly, "natural history" data (often developed by patient advocacy organization or rare disease consortia) may be incomplete or not meet FDA's standards for evidence.[11] This deficiency is particularly impactful in rare diseases with slow, heterogeneous progression and can impede use of historical comparators as an alternative to a placebo (or standard of care) comparator.

Regulatory Uncertainties: Since each rare disease is unique, there is often no regulatory precedent to help sponsors understand how to meet FDA requirements. This lack of regulatory precedent also presents challenges for FDA staff.[12] Requests for unconventional trial designs and the need for creative statistical approaches can derail a program if FDA and the sponsor cannot agree on a feasible path to approval.

In recent years, patient advocacy organizations have increasingly co-funded early research for rare diseases, and in some instances have discovered drug candidates and handed them over to industry for development.[13] However, even these collaborative, patient-led efforts can fall through if the fundamental needs – studies that are fit-for-purpose, sustainable funding and a regulatory pathway to approval – remain unmet. At least two dozen patient-supported rare disease programs have been reported as either canceled or put on hold by their industry sponsors in the last couple of years despite the philanthropic funding support. This includes instances in neuromuscular and pediatric rare diseases where families raised millions to propel a therapy toward approval, only to see their industry partner later withdraw due to business decisions or regulatory hurdles.

---

[10] LEADER 3D: Public Report of External Stakeholder Analysis

[11] Liu, J., Barrett, J.S., Leonardi, E.T., Lee, L., Roychoudhury, S., Chen, Y. and Trifillis, P. (2022), Natural History and Real-World Data in Rare Diseases: Applications, Limitations, and Future Perspectives. J Clin Pharm, 62: S38-S55. https://doi.org/10.1002/jcph.2134

[12] ICER-White-Paper_The-Next-Generation-of-Rare-Disease-Drug-Policy_040722.pdf

[13] Patterson AM, O'Boyle M, VanNoy GE, Dies KA. Emerging roles and opportunities for rare disease patient advocacy groups. Ther Adv Rare Dis. 2023 Apr 24;4:26330040231164425. doi: 10.1177/26330040231164425. PMID: 37197559; PMCID: PMC10184204.

FDA Flexibilities and the Unpredictable Regulatory Landscape for Rare Disease Treatment Candidates

The 21st Century Cures Act of 2016 included provisions to accelerate development and approval of treatments, particularly for serious and rare diseases. The law was intended to streamline the FDA approval process by allowing more flexibility in evidence requirements – for instance, permitting "real world evidence" and summary data to supplement evidence from clinical trials. It also explicitly sought to facilitate the development of genetically targeted drugs for rare diseases. Congress' goal was to reduce the time and cost of bringing orphan therapies to market, thereby lowering one barrier that has historically led to program abandonment.

Since the Act's passage, the FDA has approved numerous gene therapies and novel orphan drugs. In addition, reauthorization of the *Rare Pediatric Disease Priority Review Voucher* has provided extra incentives to see rare disease drugs through to approval by offering vouchers that can be sold/transferred to a third party.

FDA has also exercised its regulatory flexibility authority by accepting surrogate endpoints or smaller trials for some conditions and granting accelerated approval for several orphan drugs. Lack of a feasible path to approval, however, remains a major issue for many rare and ultra-rare conditions. Over the past several years, the rare disease drug development landscape has reflected an increased perception of risk due, in part, to regulatory uncertainties. This likely stems from both the inherent challenges of rare disease research and recent policy uncertainties, including:

Rare disease studies should be fit-for-purpose.

"None of us wants to give a patient a drug that doesn't work. That's true across the board, but we've lost many drugs that work because we can't show it in a way people believe they need to see it. "

"We shouldn't be wasting precious resources and patients time on studies that are not going to be fit for purpose and won't inform anything and will break hearts frankly."

"When we just look at a p-value in rare diseases, there's no way - unless we did the trial differently - we could even try to meet that endpoint. So, we're setting ourselves up to fail from the beginning and walking into it knowing we're likely to fail. Well, that's what I want to bring a halt to."

--Workshop Panelist

Uncertainties on FDA Use of Regulatory Flexibility: Innovators have noted that the FDA's application of flexibility (e.g. acceptance of novel endpoints or trial designs) can be unpredictable.[14] For example, one sponsor noted an instance where one FDA center was "reluctant to accept a surrogate endpoint" for a disease despite previous acceptance of that endpoint for the same disease. Similarly, two patient advocacy groups observed that FDA sometimes requires long, traditional trials (with large populations or survival endpoints) even for slowly progressive degenerative rare diseases

[14] GAO-25-106774, RARE DISEASE DRUGS: FDA Has Steps Underway to Strengthen Coordination of Activities Supporting Drug Development

for which study durations of 20-30 years may be required to show statistically significant clinical benefit.

Similarly, the FDA appears to have increased its emphasis on randomized controlled trials for rare and serious diseases whenever ethical and feasible. Rare disease drug development often relied on single-arm studies, external controls, or surrogate endpoints due to challenges recruiting enough patients to demonstrate a statistically significant clinical benefit. While the FDA still acknowledges those challenges, it is encouraging sponsors to incorporate control groups and generate what is considers "more definitive" evidence.[15]

The 2023 FDA draft guidance for oncology (where many rare cancers qualify for orphan status) explicitly identified RCT designs as the *"preferred approach"* for trials supporting accelerated approval even for rare cancers.[16] The guidance suggests that sponsors can use innovative design (e.g. an adaptive trial that uses one part for accelerated approval on a surrogate and continues to collect long-term outcomes for full approval) rather than skipping a control arm. FDA's message is clear: whenever possible, a new therapy – even for a rare condition – should be evaluated against a comparator (placebo or standard of care) to provide interpretable evidence of efficacy.

Erosion of Policy Incentives and Clarity: Long-standing incentives (Orphan Drug Act benefits, tax credits, priority review vouchers, etc.) have been crucial in rare disease R&D. However, recent policy changes and political uncertainty have diminished some of these incentives, contributing to a more uncertain environment. Difficulties passing bills with new rare disease incentives and refinements to existing programs (e.g. reductions in orphan drug tax credits in 2018) have reduced the "upside" to rare disease R&D.

Knowledge Gaps and Evolving Standards: The lack of regulatory precedent for the thousands of unique rare diseases challenges both sponsors and the FDA. Although the FDA has expanded its efforts to engage patients and caregivers (e.g. patient-focused drug development meetings), review teams cannot be expected to understand disease mechanisms or the patient population for each rare and ultra-rare condition. The lack of disease-specific expertise can lead to very conservative or unclear requirements. Sponsors have also voiced concerns that developing and validating appropriate endpoints *"was not straightforward"* and urged the FDA to clarify the evidentiary requirement for surrogate endpoints in rare diseases.[17]

---

[15] Expedited Program for Serious Conditions--Accelerated Approval of Drugs and Biologics
[16] Federal Register :: Clinical Trial Considerations To Support Accelerated Approval of Oncology Therapeutics; Draft Guidance for Industry; Availability
[17] *supra*,note 14

Uncertainties on When and How Regulatory Flexibilities are Used Compounds the Inherent Challenges in Rare Disease R&D and Reduces Investor and Industry Interest in Developing Much-Need Treatments.

The factors identified above present a heightened risk of very high development costs that has recently deterred investment in rare disease treatments. Both financial data and industry actions reflect a cooling of enthusiasm compared to the 2000s–2010s. For example, a Deloitte analyses indicate that the average return on investment (ROI) for recently launched rare disease drugs has plummeted – orphan drugs launched since 2022 have an ROI of roughly one-third of what drugs launched a decade ago garnered.[18] A Tufts Center review concluded that rising development costs, longer timelines, and low expected returns manufacturers have "greatly diminished the attractiveness" of rare disease markets.[19] In addition, a study performed by an industry rare disease coalition reported that the market value of publicly-traded rare disease declined by approximately 7% per year over the past five years (compared to a 1.3% annual decline for non-rare manufacturers).[20]

Although industry executives have publicly reaffirmed commitment to rare diseases in principle (often citing the unmet medical need), they have also acknowledged the need to "recalibrate benefit-risk tolerance" and business expectations for rare and ultra-rare disease R&D programs.[21] In 2023, several large companies shelved gene therapies or ultra-orphan programs after clinical setbacks, and other manufacturers have signaled that they are prioritizing investments where the path to proving efficacy is clearer and shorter.[22] Ultra-rare diseases with moderate-to-slow progression and high heterogeneity represent a "perfect storm" of trial complexity and have seen less investment momentum as a result. As noted in a 2025 policy paper, *"[u]ltra-rare diseases… share a lack of commercial viability, making traditional drug development models inaccessible for these populations. Innovative, collaborative approaches are required to bring any therapy for these overlooked conditions to fruition."*[23] That sentiment was echoed in a rare-disease investor coalition's warning that if the regulatory environment does not stabilize and better accommodate the contextual reality of rare disease R&D, many ultra-rare programs *"will continue struggling to survive"* in the development pipeline.[24]

---

[18] The Hard Truth About Rare Disease and Gene Therapy Drug Development
[19] The Hard Truth About Rare Disease and Gene Therapy Drug Development
[20] Microsoft Word - RDCC final.docx
[21] UltraRareSummary.pdf
[22] Id.
[23] Vavassori S, Russell S, Scotti C, Benvenuti S. Unlocking the full potential of rare disease drug development: exploring the not-for-profit sector's contributions to drug development and access. Front Pharmacol. 2024 Aug 12;15:1441807. doi: 10.3389/fphar.2024.1441807. PMID: 39188954; PMCID: PMC11345155.
[24] *See, supra,* Note 21.

A significant portion of Haystack Project's ultra-rare patient communities have diseases with very high heterogeneity in disease manifestations, symptoms, and/or rates of progression. This variability complicates establishing endpoints and measuring a drug's effect. Similarly, while studies in non-rare conditions can reduce the "noise" patient heterogeneity introduces into RCTs by enrolling more participants, this is not a realistic option in many rare and most ultra-rare conditions. Haystack Project agrees with the conclusion from a 2024 Everylife Foundation policy forum that *"traditional clinical trial designs aren't feasible"* for many ultra-rare conditions with variable progression. We would, however, reframe that conclusion to reflect the mathematical reality that randomized study designs in these conditions are unlikely to yield well-controlled investigations.

Flexibilities such as use of single-arm studies with natural history data or patients as their own controls, adaptive trials, or Bayesian methods are therefore a means to the same "end" RCTs achieve in larger population studies. not an exception to the rule requiring that studies be well-controlled.

Haystack Project's Scientific Workshop

Overview

The GAO Report on FDA's strategies to ensure reviewers have the necessary expertise and use appropriate flexibilities in their rare disease reviews contained important insights on FDA's efforts to meet the R&D challenges in rare diseases. Its focus on rare disease programs that successfully secured approval, however, left significant gaps in the information required to craft meaningful regulatory refinements. Haystack Project's March 12, 2025, Scientific Workshop was convened to fill those knowledge gaps and enable a productive dialogue among impacted stakeholders.

The Scientific Workshop used a case study format with researchers, clinicians, and/or patients describing the R&D project, challenges faced, and observations. These presentations were followed by discussions among presenters, panelists, and participants to gain granularity on challenges and pressure-test potential "solutions."
Case studies were selected to represent recent orphan drug research programs that encountered significant obstacles in securing approval, as well as a cancer treatment that received accelerated approval and was later withdrawn from the market due to failed confirmatory studies. Case studies included:

- A monoclonal antibody for treatment of metastatic soft tissue sarcomas that was initially granted accelerated approval. The manufacturer later withdrew the product upon failure of confirmatory RCTs to meet study endpoints.

- An orally administered, selective melanocortin-1 receptor agonist studied in erythropoietic protoporphyria patients.
- An investigational recombinant AAV2 vector intended to address the underlying genetic cause of an inherited retinal disease (IRD).
- A modified version of the human arginase 1 enzyme to treat Arginase-1 deficiency (ARG1-D).

The workshop also included related vignette-styled discussions on the R&D challenges associated with sarcoidosis (a condition with significant heterogeneity impacting multiple organ systems) and Charcot-Marie-Tooth disease (CMT) (a group of inherited neurologic disorders with five main "types" characterized by different genetic mutations and nerve damage patterns). CMT has over 100 known causative genetic mutations.

Workshop participants explored:

- How trial designs other than the RCT can meet the standard of "adequate and ***well-controlled*** clinical investigations." Participants agreed that the current RCT design is poorly suited for use in small populations, and mathematical calculations show that this design will fail to find moderate-sized beneficial effects that are easily shown in large trials.

- Whether and how the FDA considers information conveyed by rare disease-specific experts early in (and throughout) drug development programs to ensure that endpoints are identified based on a deep understanding of disease and treatment mechanisms, and that factors such as heterogeneity in disease symptoms and progression are fully considered in designing well-controlled studies.

- Mechanisms for modifying FDA's existing "regulatory flexibilities"[25] approach into a more predictable, transparent, and effective framework that maintains the "substantial evidence" standard and enables rare disease development programs that are both scientifically rigorous rare and fit-for-purpose.

Appendix 1 contains a detailed Scientific Workshop Report. The workshop case study discussions are summarized below.

Case Study #1: a monoclonal antibody for treatment of metastatic soft tissue sarcomas.

[25] Chow SC, Pariser A, and Galson S (2025). The role of regulatory flexibility in the review and approval process of rare disease drug development. Journal of Biopharmaceutical Statistics. https://doi.org/10.1080/10543406.2025.2489290

The study sponsor conducted an RCT comparing the investigational drug plus doxorubicin to doxorubicin alone on patients with 17 different rare, very heterogeneous types of sarcomas. Although the intervention did not impact tumor response or other secondary endpoints, there was a statistically significant improvement in overall survival. FDA granted accelerated approval for treating the included sarcomas in combination with doxorubicin.

The confirmatory Phase 3 study (primary endpoint of overall survival) was conducted through 110 sites in 25 countries. The manufacturer withdrew the product after the confirmatory study failed to meet its primary endpoint.

This case study illustrated multiple challenges in rare disease R&D. An oncologist with expertise in sarcomas emphasized that "[t]here's such heterogeneity and differential biology that it's very difficult when we begin to parse the sarcomas to get enough experience in specific diseases to actually move the bar forward." It is possible that the observed overall survival benefit in the Phase 1/2 study was due to chance imbalances in assignment of patients with very different prognoses and rates of tumor progression to the two arms of the randomized trial. Similarly, the challenges within the Phase 3 study may have confounded the data. For example, there was substantial variability among international sites on previous treatments participants received as well as differences in prognosis and/or potential treatment response for the 30 different sarcoma subtypes included in the study. There is also a possibility that the Phase 3 study may have "diluted out" the patient population with patient types that did well on the Phase 1/2 study either unrepresented or under-represented in the confirmatory study.

> "One important thing we are missing here is the patient voice. In this case, you are looking at doxorubicin, which has significant long-term side effects that can create other diseases -- the patient isn't there to say, 'Okay, even if PFS was the same, I would want this.'
>
> --Workshop Panelist

Despite the failed Phase 3 study, the researcher presenting this case study remained convinced that some patients benefited from the drug. He emphasized the reduced toxicities with the studied intervention and the significant improvements in quality-of-life patients experienced. The failed Phase 3 study ultimately deprives future patients of a treatment option that might have been available if FDA and the sponsor had agreed on an alternative study design and/or inclusion of patient experience data.

The discussion among panelists, presenters, and participants explored "what could have been done differently" and included:

- Augmenting study data with real-world data, including patient experience data.

- Including within FDA's set of regulatory flexibilities the potential to prioritize what the disease community (clinicians ***and*** patient advocates) identify as important outcomes.
- Considering adaptive study designs to account for heterogenous patient populations. This would enable researchers to identify responding subpopulations and shift the target population without being forced to start over again.
- Incorporating alternative statistical plans that are suitable for achieving "statistical assurance" in small population studies.

EPP researchers face difficulties measuring outcomes.

Researchers must account for heterogeneity in symptom severity as well as inter- and intra-patient variability in:

- weather and light exposure
- protective clothing
- geographic location
- behaviors and conditioned fear of sunlight

It is impossible to quantify or control all these factors with respect to impact on disease symptoms or treatment effect.

Case Study #2: Erythropoietic Protoporphyria

Erythropoietic Protoporphyria is a very rare metabolic disorder characterized by very severe cutaneous photosensitivity. Patients have a defect in the last enzyme of heme synthesis causing a build-up of protoporphyrin in red blood cells. This excess protoporphyrin makes the skin highly sensitive to sunlight, causing extreme pain. The severity of reactions varies, with some individuals experiencing mild discomfort and others having intense pain and even blistering after brief sun exposure. EPP leads to liver injury in approximately 10% of EPP patients; 3-5% of EPP patients develop end-stage liver disease.

An orally administered, selective melanocortin-1 receptor agonist demonstrated clinical benefit in a Phase 2 study. The first Phase 3 study included both a "low" dose and "high" dose interventional arm and both doses failed to show significant improvement on the primary endpoint of change from baseline in average daily sunlight exposure time until the first prodromal symptom (the early warning tingling/burning that precedes severe pain). The trial was powered for the prodromal time endpoint, based on expectations of ~60 minute improvements demonstrated in the Phase 2 study. The improvement was, however, far more modest (20-30 minutes in the high-dose group) and inter-patient variability was high. The improvements on the primary endpoint did not reach statistical significance; high-dose group improvement in the pain-event reduction secondary endpoint was statistically significant (p=0.006), and the global impression impact was highly significant (p=0.001).

An EPP expert noted that the primary endpoint was chosen to avoid causing severe pain participants (particularly those in the placebo arm) would experience if the endpoint had

been time to burn. This ethical trade-off meant that the trial may not have measured the full extent of protection the drug might have provided if it delayed or stopped prodromal tingling from escalating to an EPP attack. The discussion highlighted the conundrum for EPP R&D efforts - there is no ”perfect” was of designing an RCT in EPP.

The disconnect between the objectively measured primary endpoint and patient experience data, combined with the data from the Phase 2 study signaled a clinical benefit and suggested that trial design and endpoint selection played a major role in the Phase 3 study failure. It is not clear whether FDA might have viewed this as a case for which regulatory flexibilities were appropriate. The sponsor decided to conduct an additional Phase 3 study (apparently, with the same primary endpoint)using only the optimized dose and include an open-label extension to collect long-term data.

Workshop participants discussed several options for study design and endpoint selection, including:

- Adaptive trial design with an interim analysis to assess which endpoints are showing signals in the data.
- A self-controlled study with a run-in period during which data would not be collected. “Starting at week 4, for example, you might measure things like pain, time to improvement, time to flare, or use the area under the curve measure.”
- A self-controlled study with a crossover design to reduce impact of inter-patient variability.
- Since the test is “would this convince an expert,” input from disease-specific experts should be used in cases like this where it could be most impactful. There should be some sort of input, not just from any expert, but disease specific experts and patients who can guide study design within the context of the disease.

Case Study #3: A gene therapy for an inherited retinal disease

The sponsor sought to address the underlying genetic cause of an IRD with slow progression and significant heterogeneity in rate of progression to blindness. The investigational intervention was a recombinant AAV2 vector delivering a functional version of the gene into the retinal pigment epithelium and photoreceptor cells with a single subretinal injection. The first symptoms of the condition usually emerge in the first or second decade of life as problems with night vision. Patients typically experience a gradual loss of peripheral vision in their 20s, but visual acuity is often preserved into the 30s or 40s creating a tunnel vision effect. As individuals age (50s to 70s), central visual acuity is also lost resulting in severe vision impairment or total blindness.

This case study highlighted the combined impacts of very small patient populations, slow disease progression, significant heterogeneity, and regulatory precedent driving selection of an unsuitable primary endpoint. Since this IRD is an ultra-rare condition without treatment options to halt or slow progression, patients may not be motivated to have frequent examinations by retinal specialists. There are, therefore, significant gaps in understanding the natural history of the disease, particularly with respect to its progression.

The Phase 3 study failed to meet its primary endpoint of "Percentage of Participants with a ≥15 -Letter Improvement From Baseline in Best Corrected Visual Acuity (BCVA) at Month 12 as Measured by the Early Treatment of Diabetic Retinopathy Study (ETDRS) Chart." Despite data signaling the potential significant benefit of slowing the disease's inevitable progression to blindness, the research program has been abandoned.

The Phase 3 study was initiated based on a prior Phase 1/2 study in which over 90% of treated patients maintained visual acuity at 24 months post-treatment. Twenty-one percent of treated patients with moderate to severe visual acuity loss experienced a gain in visual acuity of at least 15 ETDRS letters from baseline compared to 1% of untreated patients in a natural history study.

> "Children are losing their vision at about 7% a year. I had night blindness discovered by the doctors. at age 11. So, it's a pediatric disease."
>
> --Workshop Panelist

Discussions among participants focused on the multiple challenges – small population, slow progression, heterogeneity, irreversible retinal damage – that made it virtually impossible for the study to meet its primary endpoint during the 12-month study period. Participants agreed that this case study highlighted the need for refinements in how FDA evaluates rare and ultra-rare disease treatments, including:

- Ensure that endpoints are determined within the context of the disease rather than driven by historical precedent. In IRD research, studies should focus studies on the more viable treatment goal of slowed progression. Endpoints related to changes in the structure of the eye might be helpful in augmenting data.
    - If you ask the patient – "what is the difference to you that you will want to have surgery for gene therapy" they will say "roll back the clock by 3 to 5 years"

- There must be openness to alternative designs and to expert opinion about what matters in the disease, and for a degenerative disorder, stabilization by definition is improvement.

    - Given the slow disease progression and significant patient heterogeneity, the only way to execute a well-controlled study would be with enrollment numbers similar to what we see in common diseases – which is impossible in an ultra-rare disease. The only other way to mute the heterogeneity "noise" is time, and if you have to extend a study to 20 or 30 years to fit the ultra-rare "square peg" into the RCT "round hole," it is pretty clear that randomization is not the best way of executing a well-controlled investigation.
    - A run-in period within a self-controlled study may provide insight into clinical improvement on the meaningful endpoint, i.e., slowed progression.

- FDA needs to have clear authority and a directive to consider alternative statistical methodologies that are consistent with the statute

Case Study #4: Argomase-1 deficiency (ARG1-D)

Arginase-1 deficiency (ARG1-D) is a rare genetic disorder that can cause seizures, spasticity and intellectual disability in untreated children. Current treatment recommendations target reducing plasma arginine and ammonia concentrations by restricting dietary protein intake and administering essential amino acid [EAA] formula and vitamin and mineral supplementation to ensure nutritional requirements are met. This standard of care is often ineffective in fully controlling plasma arginine because (1) only approximately 20–25% of arginine is derived from diet and (2) it is often difficult for patients and their caregivers to maintain adherence to highly restrictive diets. There are no approved therapies for this disease, and poor outcomes with standard of care highlight a clear unmet need for clinically effective treatment options.

A biologic was developed to more effectively and reliably meet the goal of the existing standard of care -- reduce plasma arginine levels. The Phase 3 double-blinded RCT study met its primary endpoint with a 76.7% plasma arginine reduction. 90.5% of pegzilarginase treated patients achieved normal plasma arginine levels. The Phase 3 study, combined with the Phase 1/2 clinical trial and open-label extension study demonstrated that reductions in arginine levels were accompanied by a positive trend in Gross Motor Function Measure Part E, a measure of patient mobility.

The BLA submission included positive results from the double-blind, placebo-controlled Phase 3 study and its ongoing long-term extension study as well as data from a Phase 1/2 clinical trial and its open-label extension study. FDA, however, issued a Refusal to File letter In June of 2022 stating that additional data was required to support effectiveness and demonstrate that plasma arginine and metabolite reduction predicts clinical benefit in patients with ARG1-D. FDA suggested either an animal study or another clinical trial

demonstrating a treatment effect on "clinically meaningful" outcomes. The sponsor subsequently conveyed the asset to another manufacturer.

This case study represented a "perfect storm" of R&D challenges for ultra-rare conditions. The rarity of the condition made it necessary to enroll patients at all stages of this degenerative disease, some of whom had permanent brain damage. In these later-stage patients, the outcome measures on improvement in function were impossible to meet. The sponsor also encountered FDA hesitance to accept a biomarker that is clinically meaningful, i.e., it is used by disease-specific expert clinicians to ascertain the "success" of dietary restrictions constituting the existing standard of care.

Many of the recommendations participants suggested were relevant to and suggested within previously discussed scenarios and include:

- Sponsors should understand that they have recourse when FDA approves a study design early in product development and rejects the adequacy of the design after data is presented.

- FDA should incorporate opinions of disease-specific experts into early discussions on study design and relevant endpoints. Several experts conveyed their opinions in discussions with the FDA and did so without remuneration. They were left with the impression that FDA considered their opinions biased. Ultimately, the information was not incorporated into the FDA's review process.
    - One expert stated that ""I can't tell you how many hours that other experts and I spent with the. FDA and to discuss our clinical practice, our use of Arginine as a biomarker. I remember putting up numerous slides related to Arginine and downstream metabolism."

- Sponsors and FDA should consider the concept of an "open placebo" run-in period to control for any placebo effect. Since the functional tests incorporated into this study are effort dependent, there could always be a placebo effect or training effect. If you put participants on the open placebo while you "train" them, and track their trajectory, and symptoms, you have accounted for those effects and can then put participants on drug.

Recommendations from Haystack Project's Scientific Workshop

Clarify an operational definition of the statutory term "well-controlled investigations" to reflect the rare disease reality that randomization may be a poor control method in these circumstances.

There is no statutory requirement or FDA regulation requiring RCTs as the sole (or even preferred) study design for obtaining substantial evidence in support of a studied intervention's safety and efficacy. Instead, clinical studies supporting FDA approval must be adequate and well-controlled. Traditional randomized controlled trials (RCTs) face numerous challenges when applied to rare and ultra-rare diseases that can impact whether the "gold standard" study is, in practice, well-controlled. These challenges stem from the nature of rare conditions (very small patient numbers, slow disease progression, and/or high inter-patient variability in symptoms and/or progression).

One fundamental limitation of the RCT in rare disease R&D is the small available patient population and high potential that studies will lack statistical power to detect anything but a very large treatment effect. Since a single event or outlier patient can sway aggregate results, random chance can dominate outcomes. To achieve conventional significance (e.g. $p<0.05$) with adequate power (80–90%), a rare disease trial might require unrealistically large sample sizes that simply cannot be attained. The result is often inconclusive trials that fail to show a statistically significant benefit even if a therapy has a clinical effect.

Slow disease progression and/or significant heterogeneity compound the challenges for rare disease R&D sponsors. A key premise of RCTs is that random assignment will control for extraneous variables by evenly distributing known and unknown confounders between treatment and control groups. This principle assumes reasonably large sample sizes. The more heterogeneous the patients, the larger the sample needed for statistical significance. Unfortunately, sample size is precisely constrained in rare diseases, In the context of small, heterogeneous trials, randomization's ability to create comparable groups is severely weakened, undermining the "well-controlled" nature of RCTs.
The IRD gene therapy case study exemplified how a rigid endpoint (3-line vision improvement at 1 year) in a heterogeneous, slowly progressive retinal disease can lead to a technically "negative" trial, even if some patients benefited. Overall, IRDs underscore that in slowly progressing vision loss, RCTs often struggle to show a statistically significant difference within practical timeframes. The small patient numbers and variability in disease stage at enrollment further exacerbate this. In essence, the IRD study likely failed not because the gene therapy was ineffective for all – some patients clearly benefited – but because the RCT framework and endpoint were not well-suited to capture the therapy's value in a heterogenous, slow-moving disease

The practical impact of this is profound. A therapy that might help a subset of patients or slow disease progression for many might be abandoned because it does not meet a classical RCT endpoint. It argues for the field to recognize that the concept of a well-

controlled study in rare diseases must be assessed within the context of the specific condition and its patient population.

Replace "regulatory flexibility" discretion with clearly outlined examples of alternative methodologies and innovative designs that can constitute well-controlled investigations.

The FDA has signaled openness to alternatives to the classic RCT such as accepting a single adequate trial plus supportive evidence, or use of external controls, when a second trial isn't feasible (GAO). Sponsors, however, have expressed uncertainties on when and how FDA might apply its regulatory flexibilities. However, sponsors worry that flexibility is not applied consistently. This uncertainty is impactful. If researchers, study sponsors, and investors cannot predict whether the FDA will accept approaches such as a historical control, biomarker, novel study design, or expert opinion on an appropriate disease-specific endpoint, they may decline to put R&D resources to a treatment candidate. Patient communities are increasingly concerned that until there is greater certainty that a drug for a slow, heterogenous ultra-rare disease can reach approval in a reasonable timeframe, many companies will remain cautious in committing R&D dollars.

Given the inherent limitations and challenges in navigating a regulatory path to approval for rare disease treatments, researchers and regulators have been exploring alternative trial methodologies and innovative designs that are more suitable for assessing treatment efficacy in rare and ultra-rare diseases. These approaches aim to maximize information gained from each patient and to use external data or flexible designs to overcome the constraints of a small heterogenous populations that complicate traditional RCTs. Examples include:

Adaptive Trial Designs:[26] Adaptive designs allow pre-planned modifications to the trial based on interim data, without undermining the validity or integrity of the study. In a rare disease context, adaptive methods can allocate patients more efficiently and avoid wasting patient data. For example, response-adaptive randomization might assign a higher proportion of new patients to the better-performing treatment arm as early evidence emerges. This is ethically appealing when patient numbers are limited. Other adaptations include sample size re-estimation (adjusting the total enrollment if initial assumptions of effect size or variability were off) and early stopping rules for futility or efficacy. These features can shorten trial duration or reduce the number of patients on an inferior treatment.

[26] Adaptive Design Clinical Trials for Drugs and Biologics Guidance for Industry | FDA

Adaptive designs also offer mechanisms to counter the rigidity of a clinical trial primary endpoint while maintaining the rigor of a well-controlled study. Interventional studies in very small populations often fail based on not meeting the primary endpoint despite statistically significant results on multiple secondary endpoints. Adaptive designs can provide flexibility to learn and adjust during the trial so that fewer promising treatments "fail" due to study design. Regulatory agencies have shown openness to adaptive designs under Complex Innovative Design pilot programs, although sponsors must simulate and justify control of error rates up front.

Self-controlled trials: A self-controlled trial may employ multiple cross-over periods where the patient receives alternating treatments (e.g. experimental therapy vs. placebo) in a randomized order or may compare a long run-in period to a period on treatment. Outcomes from the different periods are compared within that same patient. This self-controlled trial is also known as a n-of-1 trial. While traditionally considered a tool for personalized medicine or for optimizing an individual's treatment, a series of such trials across multiple patients can collectively provide evidence of a treatment's effect. In the rare disease setting, self-controlled trials are valuable when patient populations are so scarce that a between-patient trial is not feasible. Each patient serves as their own control, which eliminates between-patient variability and addresses heterogeneity to some extent.

Use of Historical or External Controls: When a concurrent control group is not practical or ethical, researchers can leverage historical controls or external datasets as comparators. In a single-arm trial, all patients receive the experimental therapy and their outcomes are compared against what would have been expected based on retrospective data (such as patient registries, natural history studies, or previous trials). These external control arms (ECAs) constructed from real-world data can expand the effective sample size and provide context for the treated patients' outcomes. For rare cancers and genetic diseases, it is increasingly common to use its natural history as the control – if untreated patients historically have a certain progression rate or survival, an observed improvement in the treated group can suggest drug efficacy. The advantage is that every patient in the trial gets the active treatment (addressing ethical concerns), and the study can often be completed faster.

This approach is particularly useful in rare diseases with a well-documented natural history.

Bayesian and Other Innovative Statistical Methods: Bayesian trial designs have gained prominence as a way to make the most of limited data in rare disease research. A Bayesian framework allows investigators to formally incorporate prior knowledge (from earlier studies, expert belief, or related conditions) into the analysis and to continuously update evidence as data accumulates. In practice, this can enable *borrowing* of strength from

historical controls or from trials in similar diseases, while quantifying the uncertainty in that borrowing[27]. For example, Bayesian hierarchical models can be used to combine data from a small new trial with data from past observations, effectively increasing the statistical power without enrolling more patients.

These methods can also provide more intuitive evidence to clinicians and regulators, such as the probability of a clinical benefit given the observed data (rather than a binary p-value). Bayesian adaptive designs combine the benefits of both approaches – trials can adapt based on the posterior probabilities at interim looks, which often leads to more efficient use of data. Importantly, Bayesian designs allow flexibility in analyzing endpoints and can handle complex borrowing and extrapolation (for instance, extrapolating adult trial results to pediatric rare disease populations with appropriate adjustments).

During the sarcoma case study discussion, a participant noted "This is a methodological issue. This is an example where Bayesian adaptive design might have worked if you'd been able to adapt it based on response. And how can we iterate the process as it goes on? That's where you can also bring in prospective, real world data controls and patient experience." An external observer would periodically review ongoing data and Sarcoma case studied is an example where other adaptive design approaches might be used, including having an external observer who periodically reviews ongoing data and flag trends such as "this group is not responding" and study adjustments would be built into the initial design. The advantage would be avoiding instances where sponsors conclude a study, find it "failed" and either start over, perform a subset analysis, or abandon the program.

Use of a Multi Domain Responder Index (MDRI)[28] for studies involving very small, heterogeneous patient populations enables inclusion of participants at various stages of disease progression. MDRI employs a composite endpoint integrating individual patient outcomes across multiple pre-defined clinical domains. Each domain uses a threshold for meaningful change, treating each patient's improvement across multiple areas as a single, interpretable, measurable score or outcome.

"Open placebo" run-in period. This alternative study design was suggested within the discussion on ARG-1D. Participants would be "trained" on the muscle tests during an open placebo period during which the investigator assesses their trajectory, symptoms, etc., before putting them on drug.

---

[27] Garczarek U, Muehlemann N, Richard F, Yajnik P, Russek-Cohen E. Bayesian Strategies in Rare Diseases. Ther Innov Regul Sci. 2023 May;57(3):445-452. doi: 10.1007/s43441-022-00485-y. Epub 2022 Dec 24. PMID: 36566312; PMCID: PMC9789883.

[28] Tandon PK, Kakkis ED. The multi-domain responder index: a novel analysis tool to capture a broader assessment of clinical benefit in heterogeneous complex rare diseases. Orphanet J Rare Dis. 2021 Apr 19;16(1):183. doi: 10.1186/s13023-021-01805-5. PMID: 33874971; PMCID: PMC8054393.

Reduce the burden and unpredictability associated with biomarkers as surrogate endpoints.

Surrogate endpoints have become an indispensable tool for developing treatments in rare diseases, particularly in ultra-rare metabolic disorders. The FDA has actively encouraged use of biomarkers to expedite development and issued guidance in 2020 specifically to help sponsors demonstrate efficacy in "slowly progressive, low-prevalence rare diseases… caused by single enzyme defects." This guidance states that for diseases with well-characterized pathways, measurable changes in substrate deposition in relevant tissues can serve as evidence of effectiveness for a new therapy.[29] In other words, if a drug markedly reduces the buildup of the harmful metabolite in patient tissues, that may be acceptable proof of benefit – even if an actual clinical outcome (improved symptoms or survival) would take years to observe. This approach is crucial for ultra-rare metabolic conditions where waiting for clinical endpoints *"may require an extremely long time, even decades"*[30]

> "When you have a deficiency that leads to an abnormality and you correct the deficiency, that should be sufficient – period - right? And there's precedent for that in PKU.
>
> If FDA had the "permission" to stop looking for clinical benefit in these slowly progressing, highly variable enzyme deficiency diseases where we understand the biochemistry, I think that would provide clarity and help a whole group of diseases."
>
> --Workshop Panelist

The Accelerated Approval pathway is the most frequently used regulatory mechanism enabling approval on biomarker surrogate endpoints. Using surrogate endpoints allows much earlier approval enabling patients access to promising treatments. The use of surrogates in rare and ultra-rare diseases, however, is not without controversy or uncertainty. One challenge is the validation of these surrogate markers. A surrogate endpoint serves as a proxy for clinical benefit, and it is often difficult to conclusively prove that lowering a given biomarker will indeed improve patient outcomes, particularly in very rare conditions. A recent analysis noted that *"many rare diseases have not utilized accelerated approval due to the difficulty in gaining acceptance of novel surrogate endpoints"* for conditions where no precedent exists.[31] Sponsors have expressed frustration with the process of developing and validating surrogate or intermediate endpoints, noting that it is

[29] https://www.fda.gov/media/136058/download

[30] Id.

[31] Miyamoto BE, Kakkis ED. The potential investment impact of improved access to accelerated approval on the development of treatments for low prevalence rare diseases. Orphanet J Rare Dis. 2011 Jul 6;6:49. doi: 10.1186/1750-1172-6-49. PMID: 21733145; PMCID: PMC3149566.

not straightforward on the level of evidence required.[32]. Four of the seven sponsors interviewed in connection with GAO's 2024 report noted that even though FDA *can* be flexible, their experience *"raised questions about the FDA's application of flexibilities,"* specifically citing the Agency's reluctance in some cases to accept proposed surrogates.[33] In one example, a surrogate was accepted by one FDA review division but a different center within the FDA was unwilling to accept the same surrogate for the same disease.[34]

Although FDA has published multiple guidance documents (e.g. on natural history studies, efficacy evidence, and in late 2023 a general Rare Diseases drug development guidance to provide clarity to sponsors, help sponsors, and encourages early consultation to discuss potential endpoints, both industry and patient groups continue to call for more disease-specific or pathway-specific guidance, including guidance focused on biomarkers and surrogates for ultra-ultra-rare conditions.

The ARG1-D case study highlighted the uncertain path to FDA acceptance of a biomarker as a surrogate endpoint. During the discussion, participants acknowledged the FDA's need to apply appropriate flexibility (given the urgent needs and small populations) without compromising on demonstrating that a therapy truly benefits patients. Participants concluded that:

- The scientific advances over the last 40 years, and our knowledge of biochemistry have provided sufficient confidence that if the enzyme is shown to be active at the site of action, or there is a metabolite (or other biomarker) that is recognized as clinically important, it is a valid biomarker.
- FDA must have the "permission" to stop looking for clinical benefit in these slowly progressing, highly variable enzyme deficiency diseases and rely on these valid biomarkers.
- Similarly, if the existing standard of care is directed toward reducing – to use the case study example - arginine levels, there is an established level of consensus on arginine levels as a valid biomarker. It should be treated as a "clinically accepted" biomarker.

Emphasize the Importance of Disease-Specific Experience/Expertise in Rare Disease Study Design, Endpoint Selection, Data Analysis, and Evaluation

Given that the statutory standard of substantial evidence specifically calls for sufficient evidence to convince an expert, disease-specific expertise is relevant to the study design

[32] GAO-25-106774, RARE DISEASE DRUGS: FDA Has Steps Underway to Strengthen Coordination of Activities Supporting Drug Development
[33] Id.
[34] Id.

facilitating that evidence. Many difficulties discussed throughout the Scientific Workshop could have been avoided if input from disease-specific experts on relevant endpoints and appropriate study design were part of FDA's process in rare diseases.

The IRD case study is a clear example of a clinical study for which the primary endpoint was derived from precedent (in diabetic retinopathy) and had little relationship with the IRD disease course or clinical benefit for the studied population. An IRD expert noted that since the primary endpoint required assessment of improvement in individuals with visual acuity loss, the study self-selected patients with irreversible retinal damage in whom appreciable improvements were unlikely. "When we look at natural history studies, the vast majority of choroideremia patients are declared legally blind before they have an impairment of their best corrected visual acuity." To the extent it would have been possible to meet the visual acuity improvement endpoint, it would have likely taken decades to do so since the clinical benefit for many study participants was slowed progression. Moreover, "[I]f your visual field is only a few degrees of vision because of your continued disease progression, that nominal increase in visual acuity is not terribly meaningful. If you ask the patient – 'what is the difference to you that you will want to have surgery for gene therapy' they will say 'roll back the clock by 3 to 5 years.'"

> "So, what is to be learned by this? Partly -and I think we've talked about this a bit - I don't think you should accept unfeasible endpoints when the FDA proposes them, because you are wasting your time and the patient's time.
>
> I do think that should be a mantra going forward."
>
> --Workshop Panelist

An oncologist discussing the sarcoma case study emphasized that "[o]ne important thing we are missing here is the patient voice. How many times have we sat around the table with a drug that shows activity? And in this particular case, you're looking at doxorubicin, which has very long term side effects that can create other diseases -- the patient isn't there to say, 'Okay, even if PFS was the same, I would want this new drug.' We're asked to study different endpoints, and the only person in the room who knows what endpoints are important is the person with the disease, and we need to be able to incorporate that into this process."

The porphyria case study illustrated the need for disease-specific experts to provide FDA with information on the inherent challenges in studying this condition due to significant inter- and intra-patient heterogeneity on every potential endpoint. Clinicians and researchers conveyed that there was an opportunity for FDA to hear from disease-specific experts regarding the validity of the proposed biomarker and its clinical use in managing patients.

Both the GAO Study and NASEM Report confirmed stakeholder perceptions that FDA does not actively seek or rely upon information conveyed by disease-specific experts. The

NASEM Report noted that the FDA-perceived conflicts of interest frequently impede consideration of expert opinions and generally precludes their participation on advisory panels. According to the GAO study:

- FDA officials said that while each rare disease is unique, the agency focuses reviewer training on the drug development challenges that are common across rare diseases.
    - CDER officials said that the center's Rare Diseases Team works to connect different divisions and offices within CDER and other centers, to consult with reviewers on rare disease-specific approaches and precedent as needed.
- Consulting with experts outside the FDA was not identified as a common practice by agency officials.
    - One official noted that reviewers generally do not need to consult with individual outside experts given their ability to engage with patient groups and consult with other reviewers within the agency for specific expertise when needed.
    - According to the lead of CDER's Rare Diseases Team the most common way review teams consult with outside experts during a marketing application review is through advisory committees.

Scientific Workshop participants emphasized the importance of FDA consultation collaboration with disease-specific experts and patient groups to ensure appropriate study design, meaningful endpoint selection, and accurate interpretation of trial outcomes. These consultations should be implemented as part of the review process in rare diseases and particularly in ultra-rare diseases whenever the FDA teams do not have disease specific expertise.

***Encourage FDA to move away from the "hypothesis testing" approach that tends to "doom" many rare disease R&D programs in favor of alternative statistical analysis plans*.**

Under the FD&C Act (section 505(d)), the FDA requires "substantial evidence" of effectiveness from "adequate and well-controlled investigations." This means studies must provide statistical assurance that the observed effect is real and not due to chance. In practical terms, this means:

- **Control of False-Positive Risk -** demonstrate that the chance of claiming an effect when none exists remains acceptably low (conventionally 5%).
- **Estimability & Precision** - sufficiently precise estimates (e.g. narrow confidence or credible intervals) so regulators can judge the clinical relevance of the effect.
- **Reproducibility of Findings** - even in small samples, the study design and analysis plan must show that the results aren't artifacts of ad-hoc analyses.

Workshop participants noted that FDA processes are based on hypotheses testing; i.e., the null hypothesis that the drug is ineffective and the alternative hypothesis that the drug is effective. Because of this hypothesis testing approach, we need 80% power and a p value under 0.05 to demonstrate that an effect is statistically significant. While this works when studying common conditions, it does not account for the small sample sizes and high variability often seen in rare diseases. As one researcher noted "[s]o if we stick with the hypothesis testing, I can tell you that there is no way to go. If you have a small population, high variability, and compare it to the usual 5,000 patient trial in common conditions, you have over a 100 times chance of failure to demonstrate a significant improvement under the hypothesis testing approach. You need a huge treatment effect."

Participants agreed that:

- Methods for meeting the statutory "statistical assurance" requirement should, like study designs, be fit for purpose.
- We need to give FDA clear authority and a directive to consider alternative methodologies that are consistent with the statute.
- Alternatives might include a confidence interval approach in conjunction with "precision analysis."

Conclusion

The continuing unmet need in rare and ultra-rare diseases demands rethinking the traditional paradigm of evidence generation. The gold-standard RCT, while powerful in large common diseases, may not be fit-for-purpose in small, heterogeneous populations, particularly when disease progression is slow. Statistically, the combination of limited sample size and high outcome variability undermines the RCT's ability to detect realistic treatment effects. As we have detailed above, trials can be underpowered and subject to chance imbalances, yielding inconclusive or misleading results. The case studies strongly suggest that, from a design standpoint, conventional RCT features (parallel placebo control, short, fixed duration, single primary endpoint such an improvement threshold) can render a trial essentially unwinnable. This is especially likely in diseases where progression is variable and/or slow. The case studies explored during the Scientific Workshop illustrate

aspects of this problem – whether it's needing to split trials by subgroup, struggling to find sensitive endpoints applicable to all patients, or seeing a Phase 3 fail despite biological activity and observed clinical benefit.

This is not to say that randomized trials have no place in rare diseases – they can be done and have succeeded under the right conditions (for example, when the effect is large, the condition has a rapid progressive course, or an appropriate enrichment is used). Rather, the key point is that RCTs are not always adequate as traditionally implemented. Flexibility and case-by-case consideration informed by disease-specific experts are needed to define "well-controlled" investigations for orphan diseases.

Ultimately, the goal is to ensure that effective treatments for small patient populations can demonstrate their value without being lost due to statistical and methodological artifacts. This may involve using multiple domain responder index, adaptive study designs, composite endpoints, disease-progression models, Bayesian methods, or patient-level longitudinal analyses that make the most of every patient and every data point. Just as importantly, it requires close collaboration with patient groups and disease-specific experts to design studies that maintain rigor while ensuring that effective therapies are not discarded due to the inadequacy of the testing method rather than the therapy itself.

APPENDIX 1

Key Insights and Recommendations from NASEM Report: "Processes to Evaluate the Safety and Efficacy of Drugs for Rare Diseases or Conditions in the United States and the European Union"

The Academies acknowledged that rare disease patient access to effective treatments hinges on the inextricable link between regulatory processes, drug pricing and payer decisions that "have outsized impacts on the accessibility and affordability of treatments for rare diseases." Processes to Evaluate the Safety and Efficacy of Drugs for Rare Diseases or Conditions in the United States and the European Union | National Academies
The Report recognized that rare disease research is severely complicated by the very factors that are inherent to small populations - gaps in the knowledge of disease etiology, ethical concerns, severity of disease, small sample sizes, and unmet medical need. It emphasized *that rare diseases require additional methods of demonstrating substantial evidence of effectiveness and that "[n]ew approaches in study design and data analysis need not require lower regulatory standards."*

The Academies similarly acknowledged that although FDA appears committed to including patient communities in its processes, there is *substantial room for improvement*, and recommended that FDA, among other things:

- implement strategies to solicit the views of rare disease patients during the full range of regulatory review discussions.
- include caregivers and patient representatives throughout the review process (from initial review discussions through final regulatory decisions)
- provide necessary support and accommodation to enable full participation from underserved patients and under-resourced patient organizations
- include patient representatives with lived experience in advisory panel discussions to inform advisory committees on how primary or secondary outcome measures relate to functional status and quality of life.
- assist rare disease patient groups in creating and maintaining tools (e.g., patient registry, natural history data, translational tools) that can contribute to research and development

- Finally, the Academies acknowledged the barrier that FDA's conflicts of interest policies create with respect to ensuring that decisions are informed by true clinical/research experts (small disease populations are served by small groups of clinicians, most of whom are actively involved in clinical studies) and patient organizations (that may require and accept industry funding to enable activities). While the Report did not present solutions for resolving this barrier, its own conflicts of interest policy used by the Academies in developing Reports provides an excellent model that complies with federal regulations.

> "The conflict of interest policy of the National Academies of Sciences, Engineering, and Medicine (http://www.nationalacademies.org/coi) prohibits the appointment of an individual to a committee authoring a Consensus Study Report if the individual has a conflict of interest that is relevant to the task to be performed. An exception to this prohibition is permitted if the National Academies determines that the conflict is unavoidable and the conflict is publicly disclosed. A determination of a conflict of interest for an individual is not an assessment of that individual's actual behavior or character or ability to act objectively despite the conflicting interest."

Insights from GAO Study (GAO-25-106774, RARE DISEASE DRUGS: FDA Has Steps Underway to Strengthen Coordination of Activities Supporting Drug Development)

The Government Accountability Office (GAO) published its Report to Congress on November 18, 2024. The report described FDA's strategies to help ensure reviewers have the necessary expertise and use appropriate flexibilities in their reviews, as well as FDA's programs to support rare disease drug development. As part of its review, GAO interviewed FDA officials, drug sponsors of approved rare disease drugs, and patient advocacy groups representing individuals with rare diseases. While the Report is useful in clarifying FDA's views on regulatory flexibilities and gaining feedback from research sponsors and patients, its reliance on sponsors of rare disease-drugs that were ***approved*** between 2018 and 2023 provides an incomplete picture.

GAO's findings from interviews with FDA officials, industry stakeholders and patients revealed that:

- FDA officials said they are committed to using available flexibilities where appropriate to ensure safe and effective rare disease drugs are approved for marketing.

- In August 2024, FDA held the first meeting of CDER's new advisory committee on genetic metabolic diseases, which include very rare diseases that pose unique and complex drug development challenges.

    - 3 interviewed stakeholders stated that conflict interest rules can pose challenges to finding rare disease experts who can serve on advisory committees.
    - Note: In a hearing before the House Committee on Energy and Commerce on May 22, 2024, the director of CBER acknowledged that conflict of interest rules sometimes result in challenges to having the best possible experts participate on FDA advisory committees.
    - FDA guidance states that determining whether the statutory standard of substantial evidence has been met requires an element of expert judgment.

- Most of the sponsors interviewed noted concerns about the level of disease-specific knowledge among reviewers.

    - Some sponsors noted the effects this had on drug development.
    - One sponsor gave an example of a clinical reviewer insisting that the sponsor use patient survival as an outcome measure, which the sponsor said would require a 20- to 30-year clinical trial for the slow-progressing disease."

- A patient advocacy group said that patients were unsure who at FDA was in attendance at a patient focused drug development meeting or how the information from those meetings would be disseminated to the rest of the agency.

- Three drug sponsors and one patient advocacy group said it is not clear how FDA considers patient experience data in the drug approval process.

- One sponsor gave an example of an FDA center being reluctant to accept a surrogate endpoint for a disease even though the surrogate endpoint had been accepted by another center for the same disease.

- Two patient advocacy groups commented that FDA requiring long or randomized clinical trials with larger populations creates challenges for patients with degenerative and sometimes deadly diseases

- Sponsors reported the process of developing and validating endpoints was not straightforward and requested FDA clarify the evidence needed to support the use of surrogate endpoints"

- Three selected sponsors said that there is a need for regular and in person communication with FDA throughout the drug development and application process. These sponsors said they have seen an increase in the agency's use of written communications (known as written response only letters) in lieu of face-to-face meetings, which they said limits the quality of communication between sponsors and the agency

- FDA officials said that while each rare disease is unique, the agency focuses reviewer training on the drug development challenges that are common across rare diseases.

    - CDER officials said that the center's Rare Diseases Team works to connect different divisions and offices within CDER and other centers, to consult with reviewers on rare disease-specific approaches and precedent as needed.

- Consulting with experts outside FDA was not identified as a common practice by agency officials.

    - One official in CBER said that reviewers generally do not need to consult with individual outside experts given their ability to engage with patient groups and consult with other reviewers within the agency for specific expertise when needed.

    - According to the lead of CDER's Rare Diseases Team the most common way review teams consult with outside experts during a marketing application review is through advisory committees.

- CDER's director said that one of the core expectations for reviewers is to apply regulatory flexibilities to the maximum extent possible.

    - Although staff can have a natural hesitancy toward using flexibilities too much, leadership in CBER's Office of Therapeutic Products, which reviews most of the center's rare disease products, shares a vision of embracing the appropriate use of regulatory flexibilities for rare disease drugs and works to communicate this vision

- FDA noted that small molecule drugs, which are reviewed by CDER, often lack the high specificity to their targets. FDA officials explained that without a well-understood mechanism for how a drug engages with the target, and evidence to support a reasonably likely benefit resulting from that engagement, the evidence to support a surrogate endpoint, and thus the use of accelerated approval, is lacking."

- CDER's director said that to be able to use accelerated approval or other flexibilities, such as relying on evidence from clinical trials that are smaller or more limited, "translational science efforts around rare diseases must be enhanced."

APPENDIX 2

HAYSTACK PROJECT'S RARE DISEASE SCIENTIFIC WORKSHOP

Key Discussion Points

Commonalities in challenges emerged from discussion, including:

1. Endpoints for clinical studies are often derived from FDA's prior experience reviewing treatments with different mechanisms of action impacting the same body system.

    - Choroideremia (CHM) – Progressive, inherited retinal degenerative disease causing progressive vision loss and potential blindness. visual acuity endpoint (3 lines of improvement in visual acuity) originated in diabetic retinopathy studies conducted in the 1970's and 1980's.
        - Participant indicated that since CHM's primary impact on vision is the loss of visual field, starting at the periphery, visual acuity is not the most meaningful measure of treatment efficacy. This patient-centered feedback could have led to incorporation of endpoints directly relevant to patients.
        - Long-term follow-up and analysis of study results demonstrated that CHM gene therapy slowed disease progression, but less-relevant primary EP was not met.
        - Only patients with advanced disease were tested in the phase 3 trial as the primary endpoint was specifically limited to intervention impact on that subpopulation. Given the progressive and irreversible retinal damage in these patients, the bar of improvement on visual acuity primary endpoint was unattainable.
        - Patients view a delay in progression to blindness of 3-5 years as "success."

2. Primary vs secondary endpoints

    - Heterogeneity in disease symptoms can make small population studies with a single identified primary endpoint doomed to fail.
    - Treatments have demonstrated clinical benefit based on secondary endpoints while failing to hit "significance" bar on primary endpoint.
    - This challenge was discussed:
        - If you choose the wrong outcome for the wrong endpoint (or FDA's previous reviews drive the primary endpoint), the trial may fail to meet FDA's approval standards and may not even get to an Advisory

Committee to ask question of whether experts would be convinced of safety and efficacy.
    - This is highly problematic and something we need to solve.
    - We have seen too many cases (both presented at the workshop and in life) where a primary endpoint failed to achieve statistical significance (i.e. p-value), but multiple secondary endpoints are impressive.
    - Could alternative statistical designs help us move from the status quo of "primary is king and secondary doesn't seem to matter at all"?
    - Could an alternative approach to endpoints, such as incorporating secondary or multiple tailored endpoints, be developed specifically for rare diseases, rather than applying the current FDA Multiple Endpoints in Clinical Trials: Guidance for Industry uniformly across all?
        - Under adaptive trial design, it is possible to change the primary study endpoint to a co-primary endpoints (i.e., the current primary endpoint plus the most promising secondary endpoint) after the review of interim data to improve the probability of success of the intended trial in demonstrating safety and efficacy while maintaining the rigor of a well-controlled study.

3. Many rare diseases do have not have enough patients to achieve an adequately powered RCT and thus such trials are doomed to failure unless the intervention is a "cure" -- something that rarely happens in drug development."

    - This applies equally to accelerated approval drugs for which confirmatory study requirements often specify a large RCT if approval is based on a small single arm study.
    - A large RCT confirmatory study often requires multiple international sites and the enrollment of every available patient.
    - Enrollment requirements can exacerbate challenges associated with heterogeneity, including enrolling patients that are significantly different from those upon which accelerated approval was granted.
    - A strong signal should not require a confirmatory RCT. We often then lose the drug.

4. Many of the challenges research sponsors face are set in stone early in research programs. FDA does hear from experts but does not appear to take expert opinion on what it would take to convince a disease "expert" that a specific treatment candidate works and/or that clinical effect from trial was due to intervention.

    - Research sponsors also need to know that what FDA says at one point in time (e.g., that a surrogate endpoint is sufficient) will not be changed later in the development program.

- We need early FDA actions on clinical program design that is informed by experts in the specific disease and for sponsors to be able to rely on FDA's "marching orders."
- Many sponsors are also worried about challenging FDA recommendations, either due to risk of running a study that will not represent what the FDA wants (and what the Agency would require in order to grant approval) or that pushing the FDA on one asset could cause issues with other assets under review.

5. "RCT was victory over anecdote, not triumph over other clinical study designs."

    - Statute requires well-controlled clinical investigation(s)
    - RCT's don't always get the "right answer" (sarcoma drug demonstrating clinically significant benefit when multiple sarcoma subtypes were included in trial and subjects randomized w/o stratification on specific cancer, where confirmatory trial in certain of the sarcomas showed no benefit).
    - Each of the discussed conditions and research programs had to overcome this challenge and were not able to do so
    - What changes/clarifications do we need to make to ensure "experts" inform the choice of endpoints and study design that (a) constitute well-controlled investigations; and (b) will generate data capable of convincing experts that the drug is (or is not) safe and effective?

6. Our work on HEART Act demonstrated FDA hesitation to utilize disease-specific experts. The GAO study pursuant to the HEART Act revealed that the FDA rarely consults outside experts based on belief that advisory committees suffice when FDA does not have internal expertise with the specific condition.

    - Experts do not have to clear conflicts unless they are serving on advisory committees as special government employees. Nothing precludes the FDA from otherwise hearing from experts on a specific disease.
    - Nothing should prevent disease specific experts from providing their opinions, and those opinions should not be limited to the short time limits for public input.

7. A big takeaway was FDA not consistently using Real World Evidence (RWE)[35]
    - One recent example of using a natural history study as an external control for an ultra-rare disorder is in Fibrodysplasia Ossificans Progressiva (FOP) (prevalence 1/1 million).
    - Early Phase 2 studies were placebo controlled and appeared to show a preferred dosage regimen for Phase 3.
    - FOP had no approved therapy,

[35] Data Standards for Drug and Biological Product Submissions Containing Real-World Data | FDA

- 10% of the known global population of FOP patients had participated in the natural history study
- Because of the debilitating nature of the disease, the Phase 3 study was an open label trial with a contemporaneous natural history cohort serving as an external control.

a. The objective measure was heterotopic bone volume by whole body CT (excluding head) which was assessed by a central radiology reader in a blinded fashion- for the participants in the open label Phase 3 study, and the participants in the natural history cohort.
b. The treatment eventually achieved FDA approval (palovarotene, SOHONOS) by the FDA, not the EMA.
c. This "success story" illustrates how the FDA review division realized the high unmet need, did their own independent data analyses, and were open to collaboration on what would constitute an adequate and well-controlled investigation.

8. How do we get FDA to accept what a *community* believes are meaningful outcomes.

    - Would the FDA accept opinion or consensus papers from experts in a specific rare disease or rare disease group.
    - Would it be helpful to have?
        - regulatory standards around acceptable real-world evidence that can actually be used or
        - a transparent, clear framework for ensuring a nexus between specific RWE and clinical benefit that enables a disease-specific contextual approach.

9. Participants agreed that we must focus on what the statute says, "substantial evidence that would convince experts." We are not looking for a "lower" bar to approval; we need to use trial designs that are fit for purpose to rare and ultra-rare diseases.

    - A participant pointed to the distinction between diagnosis and phenotype, suggesting that this makes Bayesian adaptive model necessary.

    - Another participant noted that the key question is "How do we get to what's practical and achievable?"

    - This is at the core of what we are trying to figure out.
        - We also have to be mindful of the need to find a pragmatic, balanced approach that is not only going to resound with the. FDA but that is

capable of convincing payers that well-controlled investigations were sufficient to convince experts that a particular drug delivers a treatment effect proven or likely to drive a clinical benefit.

- **For the Payer Workshop:** Is there any alternative to the current pricing process, such as incorporating additional information (e.g., improving quality of life, functionality, maintaining independent living (if relevant), etc., to better inform the pricing of novel treatments for rare diseases? (For example, the first approved ocular gene therapy (Luxturna) is priced at approximately $800K for bilateral treatment, and payers may approach the pricing of future therapies with increased concern).

10. The porphyria case study illustrated the need, in some disease states, for patients to serve as their own controls within adaptive study designs

    - People perceive pain differently so that a "scored" pain endpoint can only be "controlled" if the population is large enough to blunt patient variability on perception/scoring of pain
    - People live in different areas of the US where sunlight (and, therefore, episode triggers, are quantitatively and qualitatively different)
    - Study design should consider a set of endpoints initially and then narrow down target endpoints as the study progresses (Bayesian)
    - Again, we must consider how we can include patient preferred outcomes and treatment impacts that are meaningful to patients.

11. MDRI- multi domain responder index[36]

    - Is this a potential solution or something to be considered?
    - This analysis tool may help capture a broader evaluation of clinical benefit in heterogeneous complex rare diseases.
    - It allows analysis of a broad array of clinical features and sums up the thresholds of change required to demonstrate clinical benefit.

12. Biomarkers as endpoints - Enzyme Replacement Therapy

    - Consideration that if we know a drug is active at the site where it normally operates, that biomarker should be sufficient to demonstrate clinical benefit
    - Could this avoid having to pursue accelerated approval and the resulting noncoverage from some payers?

[36] Tandon, P. K. The Multi-domain responder index: a novel analysis tool to capture a broader assessment of Clinical benefit in heterogeneous complex rare diseases. Tandon and Kakkis Orphanet J Rare Dis (2021) 16:183 https://doi.org/10.1186/s13023-021-01805-5

- What about biomarkers that are medically accepted as clinically meaningful (i.e., used by clinicians to assess disease progression, impact of interventions like dietary restrictions)?
- For these research programs, is an RCT appropriate?
- How can we consider approval based on either open label studies or a combination of open label studies and long-term follow-up?

Case Studies

Case Study #1: a monoclonal antibody for treatment of metastatic soft tissue sarcomas.

Background

- The sponsor conducted a randomized controlled trial in the U.S. of antibody plus doxorubicin versus doxorubicin alone. More than 17 different rare, very heterogeneous types of sarcomas were represented in the trial.

- The results showed a numerical improvement in progression free survival (the primary endpoint) that was not statistically significant (target hazard ratio was 0.67)

    - ***Improvement in overall survival was highly statistically significant.***
    - There was little difference in objective tumor response or other secondary endpoints between the two arms.

- Based on this study, the drug was approved via accelerated approval for treatment of the included sarcomas in combination with doxorubicin.

- A confirmatory Phase 3 study with an overall survival primary endpoint was conducted through 110 sites in 25 countries.

    - Data from this study failed to show an improvement in survival between patients receiving study drug with doxorubicin and those taking doxorubicin alone.

- The product was subsequently withdrawn by the manufacturer.

Discussion

- “There's such heterogeneity and differential biology that it's very difficult when we begin to parse the sarcomas to get enough experience in specific diseases to actually move the bar forward.”

- It is possible the observed overall survival benefit in the Phase 1/2 study was due to chance imbalances in assignment of patients with very different prognoses and rates of tumor progression to the two arms of the randomized trial. While statistical analysis techniques are employed to assess the possibility of this happening, they cannot rule it out.

- It is also important to note that the Phase 3 study faced similar challenges that may have confounded the data, including variability among international sites on previous treatments participants received and differences in prognosis and/or potential treatment response for the 30 sarcoma subtypes included in that study.

- The confirmatory study may have “diluted out” the patient population, i.e., the patients who did well on the earlier study were not represented in the later study.

What could have been done differently?

Augment study data with real world data, patient experience data.

- “If you can add real world data or look at the patient experience data . . . I think that is something that's worth a discussion, because had we been able to do that in this trial in particular, then we might have had a different outcome with that Phase. 3.”

More fully incorporate patient preferences into clinical studies and reviews.

- “I think one of the things that I would be most excited about is, how do we allow flexibility of the FDA to accept what a community, not only clinicians but patient advocates, feel is important when looking at outcomes of a drug and allow us to explore that.”

- “One important thing we are missing here is the patient voice. How many times have we sat around the table with a drug that shows activity? And in this particular case, you're looking at doxorubicin, which has very long term side effects that can create other diseases -- the patient isn't there to say, ”Okay, even if PFS was the same, I would want this new drug.’”

Incorporate endpoints reflecting patient-centered outcomes, including QoL.

- “We're asked to study different endpoints, and the only person in the room who knows what endpoints are important is the person with the disease, and we need to be able to incorporate that into this process.”

- “It's fair to say that in this biz [cancer research] the strategy is strictly death. There is no quality of life. Well, no, it's the mandate of overall survival. I think it’s one of the worst outcomes that we could do. We have so many diseases where we've lost drugs because I can't show worth to a pharmaceutical company, because I can't get their tumor to shrink 10%. But I have a patient who couldn't breathe. They say “I just walked up that flight of stairs.”

- “We are urgently trying to develop appropriate symptom-based outcomes. But the regulatory bodies view these as a distant exploratory endpoint. Show the tumor shrinks, and then they will look at the symptom-based endpoint. We have to “order” our endpoints.”

Consider adaptive study designs

- “This [heterogeneity in small population diseases] gets to why we need to think about adaptive designs where you can make that analysis [identify responding subpopulations] and shift your population to the target population without losing all that information and being forced to start over again, which is the “frequentist” method.”

- “This is a methodological issue. This is an example where Bayesian adaptive design might have worked if you'd started a large trial and been able to adapt it based on response. But how can we iterate the process as it goes on? And that's where you can also bring in prospective, real world data controls and patient experience.

- “We should also consider other adaptive design approaches such as having an external observer who breaks in and says ‘Okay, this group is not responding.’ That is built into the initial design so you don’t get to the end and say ‘oh, we failed. Now we’re going to do a subset analysis.’”

The regulatory requirement of “statistical assurance” does not mandate power analysis and p value below .05. Small population diseases with significant heterogeneity cannot achieve “statistical assurance” without alternative statistical plans.

Case Study #2: Erythropoietic Porphyria

Background

- This disease is very rare and characterized by very severe cutaneous photosensitivity. It results from a defect in the last enzyme of heme synthesis, and a toxic substance builds up in the blood protoporphyrin, which is light, sensitive, and leads to very severe pain.

- Patients have to change their lives to avoid light exposure and prevent the extreme pain due to EPP reactions. The symptoms can last for days, and there's no medication to treat the pain once it starts.

- Patients must cover up and avoid sunlight and sometimes indoor lights depending on the patients.

- Patients are sensitive to blue light, so sunscreen doesn't help. It doesn't have any effect in EPP.

- In addition to light sensitivity, EPP can also cause liver failure requiring liver transplantation.

- An oral medication was studied for EPP.
    - The medication increases melanin to improve light tolerance.
    - It blocks a broader range of wavelengths than sunscreen.

- In a successful 2020 phase 2 study published in the New England Journal of Medicine, it showed great benefits in patients

- In the 1st phase, 3 trial, it did not demonstrate significant improvement on the primary endpoint even though it showed significant improvement on multiple secondary endpoints

Discussion

- FDA, the investigators, and patients all believe it works, but because it was not successful on the one pre-specified endpoint it failed.

- Past clinical trials have also struggled before this.

- Researchers have thought about different models and assume legislation would be needed for adaptive trial designs.
    - After the first part of the study, you do an endpoint analysis to see which endpoints are showing signal in the data.

    - You then comb down the rest of the trial to those endpoints and carry the data forward from patients in the 1st part of the trial.

- The study sponsors and investigators did not like the endpoints that were proposed but were told that those endpoints were required as they had been used previously.

- They did not feel like there was any flexibility.

What could have been done differently?

Alternative trial designs

- "I can see using a self-controlled study with a run-in period and not collecting data while the drug is getting in equilibrium. Starting at week 4, for example, you might measure things like pain, time to improvement, time to flare, or use the area under the curve measure."

- We did think they were doing something different, because the 1st approved drug had a primary endpoint of time to burn. The patients did not know if they had drug or placebo and were told to go out in the sun until it hurts. For this trial we created the validated endpoint of time to prodrome, to that beginning signal of tingling but not pain. But there is a lot of variability in that across patients.

Incorporate input from disease-specific experts early in the process.

- "It seems to me that since the test is "would this convince an expert?" that input from disease-specific experts should be used where it is most impactful. In some ultra rare conditions and rare conditions, people may have different symptoms. There may be different disease trajectories, etc., but if the question is would an expert be convinced that the response is due to the drug, there should be some sort of input, not just from any expert, but disease specific experts and patients who can guide study design within context of the disease and say "Hey, look if the study data showed X,Y, and Z, then I would say, that's a winner. "'

- "Okay, what we're doing now is saying, what was used before is appropriate. But what we heard on the oncology example is that experts in the disease, not in cancer in general, but pediatric sarcomas, understand what is the most important thing."

- "I want to point out the importance of expertise. I'm familiar with interactions with the agencies. Ophthalmologists are not subspecialists in inherited retinal degenerations. And we know so very little, relatively speaking, about natural history. what's available in the published literature represents a fraction of what's known. So it's actually the expertise of clinicians seeing patients with the disease actually accounts for a lot of what could be known about these conditions. So if the

expertise is the regulator who has general ophthalmologic experience, for example, it is not really much expertise at all. I think that's very relevant to EPP and we find that in all rare diseases."

Case Study #3: a gene therapy for an inherited retinal disease (IRD)

Background

- The investigational recombinant AAV2 vector was intended to address the underlying genetic cause of choroideremia by delivering a functional version of the human choroideremia gene into the retinal pigment epithelium and photoreceptor cells with a single subretinal injection.

- The Phase 3 study failed to meet its primary endpoint and, despite potential that the treatment might provide significant benefit to patients by slowing the disease's inevitable progression to blindness, the research program has been abandoned.

- The condition is a rare, degenerative, X-linked inherited retinal disorder that leads to blindness.
  - A diagnosis can be confirmed with genetic testing.
  - In males, symptoms typically occur in the first or second decade of life when problems with night vision occur.
  - In the 20s, gradual loss of peripheral vision begins but visual acuity is often preserved into the 30s or 40s creating a tunnel vision effect.
  - As males age into their 50s - 70s, central visual acuity is also lost resulting in severe vision impairment or total blindness.
  - As visual acuity decreases, most patients have asymmetrical progression with vision loss in their "worse" eye accelerating more rapidly than the other.
  - There are no FDA-approved therapies for this IRD and current management is symptomatic and supportive.

- Since this is an ultra-rare condition without treatment options to slow progression, patients may not be motivated to have frequent examinations by retinal specialists. Consequently, there are gaps in understanding the natural history of the disease, particularly with respect to its progression.

- Progression can be highly variable across patients as well as with respect to a patient's left and right eyes.

- The primary endpoint was "Percentage of Participants with a ≥15 -Letter Improvement From Baseline in Best Corrected Visual Acuity (BCVA) at Month 12 as Measured by the Early Treatment of Diabetic Retinopathy Study (ETDRS) Chart"

    - In a condition that causes retinal damage as it progresses, improvement in vision would be unlikely in patients that have progressed significantly. The inclusion criteria were set to require a level of progression at which retinal damage was likely.

- The study was originally initiated based on phase 1/2 studies, showing that at month 24, over 90% of treated patients maintained visual acuity unlike expected acuity loss in natural history of choroideremia.

    - Twenty-one percent of treated patients with moderate to severe visual acuity loss experienced a gain in visual acuity of at least 15 ETDRS letters from baseline compared to 1% of untreated patients in a natural history study.

- A recently published study in *Nature Medicine* noted that although the study failed to meet its primary endpoint, future trials might look at the goals of outcome success through a different lens.

Discussion

- The FDA has approved one gene therapy to date for a different inherited retinal disease.

    - The approved therapy is for a kind of inherited retinal disease in which the retina structure is intact. In most inherited retinal diseases, the retinal structure isn't intact and it's a "tall order" to think that treating a malfunctioning gene can improve visual function.
    - "In the Luxturna trial, there was an immediate reaction and sight seemed to be immediately restored, or within a short period of time, but in our case we just want to stop the disease."
    - A more viable goal of treatment in this IRD would be to slow progression

- Participants selected for the study had far more advanced disease than those in the earlier trials because of the low BCVA entry requirement. This limited the BCVA gain potential.

- The study was originally initiated based on phase 1/2 studies, showing that at month 24, over 90% of treated maintained visual acuity unlike expected acuity loss in natural history.

    - Twenty-one percent of treated patients with moderate to severe visual acuity loss experienced a gain in visual acuity of at least 15 ETDRS letters from baseline compared to 1% of untreated patients in a natural history study.

- Any time you have improvement beyond one line of vision you know there's something there. There may be some differences in patients, but 3 lines is purely from diabetic retinopathy 30 years ago it is ancient, and if someone improves by one or 2 lines, it's great.
- Unless we can really tackle the thinking about these historical outcome measures that are not really applicable to inherited renal disease, then it's probably very difficult to advance a dialogue about composite outcome measures and other adaptive clinical designs. Unless we overcome this, it's going to be very difficult.

What could have been done differently?

Endpoints must be determined within context of the disease, not based on historical precedent.

- This is a slow progressing disease. So, we need to study the changes in structure of the eye as an endpoint."
- A more viable goal of treatment [than improvement in visual acuity] would be slowed progression.
    - They clearly showed that the placebo group was deteriorating faster, even though they did injections into the retina.
    - "There must be openness to alternative designs and to expert opinion about what matters in the disease, and for a degenerative disorder. stabilization by definition is improvement."
- "So, what is to be learned by this? Partly, I think, we've talked about this a little bit - I don't think you should accept : unfeasible endpoints when the FDA proposes them, because you are wasting your time and the patient's time."
    - "I do think that should be a mantra going forward."
    - , and you know there are even appeals, mechanisms that you can do to protest, and I know this is a long time ago, and water under the bridge and everything. But I do think that ought to be a mantra going forward.

Study design challenges in endpoints also push inclusion/exclusion criteria away from younger patients.

- "If I were going to do a clinical trial and had no regulatory body to be concerned about, I would go and treat a lot of younger patients."
- "I would follow them for 5 or 7 years and look at every measure and develop a composite index, and then I would go ahead and do an interim analysis to see if this treatment is effective."
- "If you target a treatment to a younger age range, you might maintain that person's ADL and allow them to be "normally functioning," drive a car and things of that sort."
- "Is it fair to say that when you show what you showed, even for one line, that it would suggest that there is not a continued progression at the same rate?"

    - "Perhaps for one line you may need a longer period of time, and another measure would have been the OCT - optical coherence tomography which can be utilized as a structural measure, that you follow over time.
    - The only caveat to the OCT is that when you do a subretinal injection you can affect it. But you can also take the post-operative OCT as the baseline as you move forward and plot out a rate after the treatment compared to without the treatment.

Non-RCT study designs might provide better controls for heterogeneity.

- If you were to do a study differently without the traditional outcome measures, maybe you would choose, for example OCT of a certain size of retained retina, perhaps certain other parameters that would be more carefully selected
- You can't really compare to the other eye, because the other eye of the patient may have slightly different progression
- A run-in period within a self-controlled study may provide insight into clinical improvement on the meaningful endpoint, i.e., slowed progression.

The "hypothesis testing" approach does not account for small sample sizes in rare diseases and can lead to "failure" of effective drug candidates.

- "FDA processes are based on hypothesis testing.- the hypothesis that the drug is ineffective and alternative drug is effective. Because of this hypothesis testing, we want to make sure we have 80% power."
- "Then, we have to make sure that the P value is less than 0.05."
- Unfortunately the p-value actually does not really reflect the sample size or variability at all. Small patient populations and huge variability are common in these rare diseases.
    - "So if we stick with the hypothesis testing, I can tell you that there is no way to go."
- if you have a small population, high variability, and compare it to the usual 5,000 patient trial in common conditions, you have over 100 times the chance of failure to demonstrate a significant improvement under the hypothesis testing approach. You need a huge treatment effect.
- ***The solution may be to consider a confidence, interval approach in conjunction with the so-called precision analysis*.**
- We need to give FDA clear authority and a directive to consider alternative methodologies that are consistent with the statute

Consultation with disease-specific experts and patients on appropriate study endpoints in these very rare conditions is essential.

- "When we look at natural history studies, the vast majority of these IRD patients are declared legally blind before they have an impairment of their best corrected visual acuity."
- "To answer the question that somebody had posed earlier about what sort of impact visual acuity improvement has to a choroideremia patient - it may be negligible, because if your visual field is only a few degrees of vision because of your continued disease progression, that nominal increase in visuality acuity is not terribly meaningful."
    - If you ask the patient – "what is the difference to you that you will want to have surgery for gene therapy" they will say "roll back the clock by 3 to 5 years"

Case Study #4: Argomase-1 defocoemcu (ARG1-D)

Background

- Arginase-1 deficiency (ARG1-D) is a rare genetic disorder that can cause seizures, spasticity and intellectual disability in untreated children.
    - Occurs in approximately 1:950,000 births
    - Most infants with ARG1-D do not exhibit any clinical manifestations during the first year of life
    - ARG1-D becomes evident between the ages of 1 and 3 years; onset of symptom varies but all patients progress over time and many experience loss of motor function and neurological impairment

- Current treatment recommendations target reducing plasma arginine and ammonia concentrations.
    - restricting dietary protein intake
    - administering essential amino acid [EAA] formula
    - vitamin and mineral supplementation to ensure nutritional requirements are met
    - nitrogen scavengers in some patients.

- Recommended approaches are often ineffective in fully controlling plasma arginine because only approximately 20–25% of arginine is derived from diet and highly restrictive diets are often difficult to adhere to.

- Recommended standard of care for ARG1-D is based on case reports and limited clinical studies and does not address the issue of endogenous arginine production.

- The lack of approved therapies and poor outcomes with standard of care for ARG1-D highlight a clear unmet need for clinically effective treatment options for patients with ARG1-D.

- Regulatory progress and actions
    - FDA granted Rare Pediatric Disease, Breakthrough Therapy, Fast Track and Orphan Drug designations
    - BLA submission included positive results from double-blind, placebo-controlled Phase 3 study and its ongoing long-term extension study as well as a Phase 1/2 clinical trial and its open-label extension study.
    - According to manufacturer, "the totality of data" demonstrated that the drug is able to rapidly and sustainably lower arginine levels and is accompanied by improvements in mobility."
    - The Phase 3 clinical trial met its primary endpoint with a 76.7% plasma arginine reduction.
        - 90.5% of treated patients achieved normal plasma arginine levels.
        - The arginine lowering was accompanied by a positive trend in Gross Motor Function Measure Part E, a measure of patient mobility.

- In June of 2022, the FDA issued a Refusal to File (RTF) letter
    - requested additional data to support effectiveness, such as evidence showing that plasma arginine and metabolite reduction predicts clinical benefit in patients with ARG1-D OR
    - clinical data demonstrating a treatment effect on clinically meaningful outcomes.
- The European Commission subsequently granted marketing authorization for the treatment ARG1-D in adults, adolescents and children aged 2 years and older.
- The manufacturer re-submitted the BLA and the BLA was accepted on November 5, 2024.

Discussion

- One disease expert said that the failure to meet FDA's bar of clinical benefit may have been due to fact that many of the 32 patients were older and had been diagnosed symptomatically only after severe and irreversible brain damage had occurred

- Clinicians collected video evidence showing that participants in the extension trials have deteriorated since going off of treatment. "In one case, a patient had not gotten particularly better, but when he got off the enzyme, he really just collapsed."

- Given the rarity of this disease, some of what FDA wanted for the study was not feasible. They wanted to limit the study to younger individuals but many of the patients who were enrolled were older and had more disease progression.

- There is no detailed information about how patients progress and how long it takes for them to stop walking. We have robust data in the Urea cycle consortium, but we have very few patients with Arginase deficiency.

    - Nobody's carefully done any type of biomechanical measures of gait, etc.

- The double blinded phase 3 was 32 patients, and some patients already had irreversible brain injury.

- The study met the primary endpoint with a 76.7% plasma arginine reduction and 90.5% of the patients achieved a normal plasma arginine level.
    - The ongoing long-term extension of the Phase 1/2 study showed that it was sustainable.
    - There was also a positive trend in a gross motor function. The families really saw improvement in the patients, as did the investigators at the different sites.

- "If you have a deficiency that leads to an abnormality and you correct the deficiency, that should be sufficient – period - right? And there's precedent for that in PKU.

What could have been done differently?

Sponsors should understand that they have recourse when FDA approves a study design early in product development and rejects it after data is presented.

- "The FDA completely ignored their own advice on the study design, and that is the part that that I see as a huge problem. It happened in another trial where a study design was approved, and then in the end, they look at it, and they say, "well, we don't like that study design."
    - I don't know why the company didn't appeal.

- "Researchers from the Urea Cycle Consortium actually wrote a letter to the FDA and had several phone calls.
    - We did this as individual investigators.

Study designs with biomarker endpoints in these conditions are not inferior with respect to scientific rigor.

- I think we can say that the scientific advances over the last 40 years, and our knowledge of biochemistry would show that if the enzyme is shown to be active at the site of action it is a valid biomarker.
- If FDA had the "permission" to stop looking for clinical benefit in these slowly progressing, highly variable enzyme deficiency diseases where we understand the biochemistry, I think that would provide clarity and help a whole group of diseases.
- Another point: if the standard of care is directed toward reducing arginine levels and success is measured by arginine levels, there is already a level of consensus on arginine levels as a valid biomarker. It is a "clinically accepted" biomarker.

- “I can't tell how many hours that other experts and I discussed our clinical practice, our use of Arginine. I remember putting up numerous slides related to Arginine and downstream metabolism.”

Consider concept of “open placebo” run-in period.

- “I don't know whether any of you are familiar with the idea of open placebo, however, it's very effective, and it removes this generally subjective placebo effect, which is what everybody's worried about, right?

- Most muscle tests are effort dependent, so there could be always a placebo effect or training effect. You put people on the open placebo while you “train” them. You follow them; you figure out their trajectory, their symptoms, and then you put them on the drug.

- They shouldn't have too much of a new placebo effect, because they already have their placebo effect. And this has been proven that the open placebo causes a placebo effect.
    - It doesn't affect objective measures whatsoever.

- You would look at the rate of progression of the disease, could do that individually, and once the individual has had a discernible pattern that you could measure, you could switch them over.