Molecular Genetics and Metabolism 137 (2022) 107–113

Contents lists available at ScienceDirect

Molecular Genetics and Metabolism

journal homepage: www.elsevier.com/locate/ymgme

Commentary

The transformation of drug development for the 21st century: Time for a change

Emil D. Kakkis *

Ultragenyx Pharmaceutical Inc., USA

The sequencing of the human genome and therapeutic technology advancements have enabled so many diseases to be diagnosed, understood, and potentially treated, that we have rapidly advanced past our ability to develop and assess treatments effectively for diseases that are now being diagnosed routinely. This is particularly true for those complex neurogenetic diseases which were so difficult to diagnose in the past, and whose symptoms are complex, variable, and irreversible at times. The major deficiency for drug development is that the advances in precision medicine concepts for the treatment of the underlying condition are not being matched with a precision medicine equivalent for accurate measurement of disease activity. The original requirement of improvements in how a patient feels, functions, or survives as the gold standard remains an often quoted and unquestioned belief in drug development and regulation [1] and now also in reimbursement [2]. The fact that a meaningful medicine should make a patient better in some immediate measurable way certainly makes sense in establishing a basis for why they are treated and why we should pay for the medicine. However, the science of treating has advanced past the simplest rapidly reversible diseases that can readily fit that paradigm, and our decisions on what is worth paying for needs to advance to an understanding that treating the underlying diseases before they are manifest as clinical conditions is essential and desirable. Furthermore, the complexity of the "Feels, Functions, Survives" standard is that downstream variable pathophysiologic responses to diseases caused by the intrinsically variable genotypes of human populations create substantial noise that thwart efforts to measure clinical changes and these responses are, at times, misleading and not actually consistent with treating the underlying disease. For diseases like progressive neurological diseases in which the brain tries to compensate and maintain function, the change in clinical manifestations occurs very late in the course of disease at which point the system can no longer compensate and finally breaks down, leaving little room to improve the underlying disease. The relatively late diagnosis of these diseases makes it very hard to study these diseases at the most effective early point of disease. As an example, my first patient with adrenoleukodystrophy was a bright capable 7-year-old who went from having trouble seeing the blackboard to serious visual deficiency in weeks. Yet his very first MRI showed evidence of completely destroyed white matter in his brain, a shocking finding that made it hard to understand how his clinical symptoms had only just begun. The disease was long in progress, though his brain adapted to keep him functional to the last moment. The biology of many neurologic diseases, and frankly many complex diseases like renal diseases, are also not convenient for clinical endpoints. More importantly, the strong emphasis on clinical endpoints tends to bias development of successful treatments to effects on late-stage consequential clinical manifestations rather than underlying disease, and ultimately distracts us from treating the disease correctly. Instead of keeping all the neurons intact before they are lost, we are just making the last few neurons left perform a little better. Is that what society wants?

We need to start focusing on treating underlying disease before the final breakdown associated with clinical manifestations, and to do this, we first need to define a precision medicine term for disease measurement. Since precision medicine is targeted to the source of disease, we need to develop a class of biomarkers that are a direct measure of the core underlying primary biology or as close to it as possible. By being close to the source of disease, noise from secondary genetic variation in downstream pathophysiologic responses to disease is minimized and the biological plausibility is far greater. In combination with a direct treatment to the underlying disease cause, like an enzyme or gene replacement therapy, the biologic plausibility is much higher. When studying the primary biomarker, the impact of a variable degree of the pathophysiological progression of the underlying biology on clinical outcome would become less consequential to the success of development. We are able to measure what is happening early on in the disease course and will not be confused by the variable degree of irreversible disease and rate of progression of clinical manifestations. This newly defined category of biomarker is intended to be distinguished from downstream biomarkers (e.g., arrhythmias in patients with underlying heart disease or inflammatory markers associated with a chronic lysosomal disease) that are less directly representative of the primary disease process and more prone to error, especially if a treatment is not designed to solve the underlying genetic condition. This commentary describes a precision medicine term for disease measurement, *Primary Disease Activity Biomarkers* (PDAB) and particularly their roles in the study of treatments for rare genetic diseases and in a revolution in how we think about drugs, diseases, and treatment measurement that can transform our development of treatments and finally truly open the door to treatment of the underlying cause in rare genetic, neurologic and other diseases.

* Corresponding author at: 60 Leveroni Court, Novato, CA 94949, USA.
E-mail address: ekakkis@ultragenyx.com

https://doi.org/10.1016/j.ymgme.2022.07.015

1. Background for the 20th century paradigm

1.1. Feels, functions, survives: a paradigm for the beginning when efficacy was added to the requirements for approval

The 1962 Kefauver-Harris amendments to the federal Food, Drug and Cosmetics Act provided the first critical requirements for drugs to be effective and the standard that there be substantial evidence of effectiveness [3]. These changes put the appropriate pressure on the development of drugs to assure we are using drugs that actually do something useful and are reasonably safe. The standards for what are appropriate clinical endpoints was the often-quoted Robert Temple description of how a patient "feels, functions or survives" [1]. This standard makes perfect sense in a setting with common, reversible clinical diseases with or without much understanding of underlying causes or pathophysiology. If the patient is better, then the drug works. There may be many debates on what is truly clinical, especially with regard to how one measures a patient's function. However, for many simple common diseases with the ability to rapidly improve over weeks to months and with some consistency in pattern of disease symptoms, this standard works very well. For diseases with small populations, slower progression, complex variability, progressive irreversibility, and delays between disease tissue impact and clinical manifestation of disease, the "Feels, Functions, Survives" standard can be extraordinarily difficult if not impossible to navigate the development of treatments that affect underlying causes. For example, in Alzheimer's disease, the use of cognitive function has led to approval of drugs like donepezil, which can improve cognitive function of the residual but damaged nervous system despite having no impact on the actual disease cause [4]. The same situation exists for Parkinson's disease treated with L-DOPA, wherein treatment can enhance function for a while until the degenerative disease finally fully overtakes the patient [5]. Both drugs are useful, but they will not treat the underlying disease which progresses well before the symptoms appear and is advanced by the time a person is diagnosed.

For a disease like AIDS, HIV infection has a long course that could take 7 years or more to become manifest as AIDS, and secondary causes like other infections and general health could have substantial ramifications on the risk for progression and the speed and degree of outcomes. In this situation, what type of drugs would we create if we focused on improving clinical immune function, for example reducing opportunistic infections as a clinical endpoint? Likely, we would find drugs that stimulate the remaining immune system to function a bit better, like gamma-interferon or IL-2, but would never treat the underlying cause of disease. This is how an extreme focus solely on clinical endpoints can actually fundamentally mislead the clinical research and drug development. The long time frames, variable irreversible symptoms, and complex biology with delays between the cause and the outcome challenge us to consider how "Feels, Functions, Survives" can be effectively used to create treatments and whether a new paradigm is needed. While HIV infection may seem distinct from rare genetic diseases, they share a common characteristic: a clear, primary cause in the presence of a genetic entity that is not normal and from which all of the disease is sourced. It is also similar in how infection precedes the progression of disease by many years and in the tremendous variability in disease course based on other genetic and environmental factors.

1.2. Accelerated approval: Conceived in the urgency of the AIDS crisis leads to new insights on drug development

The accelerated approval regulations promulgated in 1992 and later ratified in law [6], were a response to the AIDS crisis and the demand for more rapid and effective drug development by AIDS activists and others. The concept was to approve drugs first on a biomarker before confirmatory studies, often expected to be ongoing, were completed. This was essentially an attempt at speeding access by months or years, but not intended to transform drug development. At the time, many at FDA and outside pundits charged that the new Accelerated Approval (AA) regulations would destroy drug development and lead to disastrously ineffective drugs, thereby distracting patients from real drugs developed with clinical endpoints. The truth was that the exact opposite happened. The first HIV drugs that were approved and gained reimbursement with CD4 cell counts as the biomarker were later followed by better drugs based on improved and an earlier biomarker of disease activity, namely HIV viral load. During the 16 years after the 1992 regulations went into effect, 29 drugs were approved with multiple mechanisms of action against the virus, including four multiple combination drugs that led to highly active anti-retroviral therapy (HAART) and the transformation of HIV infection from a death sentence to a chronic survivable disease [7]. The importance was that viral load as a precision medicine measure of the amount of virus present could allow a drug developer to rapidly assess the impact of a new drug target on the virus in a matter of a few weeks and look at combinations of directly acting antiviral drugs more quickly to drive the virus levels down through direct action on the virus. If AA had not been available, the process of running clinical trials would take years with variable endpoints like rate of opportunistic infections or mortality that would be difficult to power, highly variable, imprecise, confounded, and untenable. No HAART could have been developed without AA. The pundits were proven wrong. AA not only accelerated the development of many effective drugs against HIV, it also the transformed drug development process in a fundamentally more effective direction that was not just expedient, but far more insightful and effective. The same thing later occurred for Hepatitis C, where HepC viral load led to rapid improvements, obsolescence of early drugs, and finally a cure [8]. This pattern exists because the linking of multiple drug approvals for the same disease allowed the rapid improvements and medical evolution of treatment that would never have been possible with longer and slower clinical endpoint measurements, where the required improvements to exceed prior impact on clinical endpoints would necessitate huge studies that are very difficult to conduct with low probability of success. Thankfully, the payers at that time did pay for the biomarker-approved HIV drugs to reduce viral load without any evidence of clinical benefit at first, which is an essential part of the medical evolution of HIV treatment. If the payers had concluded that drugs without proven clinical benefit should not be paid for and should be considered "experimental" as recently occurred or threatened via the Centers for Medicare and Medicaid Services decision on Aduhelm (plus other payers' recent pushback on paying for biomarker-approved drugs), the AIDS epidemic would have been far more disastrous and possibly never controlled. Some may claim we knew more about HIV then, but in fact many commonly held assertions like it being a slow virus, were actually wrong. We learned about HIV by treating it and while the science may look obvious now, it was not obvious then.

The AA process and the use of viral load as the primary disease activity biomarker opened the door to not only drug development, but to a revolution in the treatment of HIV with multiple competing drugs and combinations. It is critical to appreciate that the change is not just speeding each step, but a transformation of the process and the concept of treating disease at its root and measure it directly, not just at its clinical manifestations. The "Feels, Functions, Survives" as a paradigm could not achieve this result. The clinical benefit data for these HIV treatments took time to develop, but it became clear that patients on long-term treatment were not progressing to AIDS and this clinical benefit required longer periods of follow-up which was not well suited to randomized trials in general, unlike the viral load studies which were all randomized but shorter studies [9]. There are other key learnings from the HIV legacy: 1) the first biomarker does not have to be perfect and can evolve as scientific knowledge grows; while CD4, the first biomarker, was imperfect and perhaps too far downstream in the pathophysiology of the disease; the science advanced and the viral load biomarker was established [10], 2) the development of superior drugs

was enabled, easily obsoleting the first drugs. This may have been very difficult to do in a head-to-head clinical endpoint study with smaller differences in clinical outcomes, thereby requiring larger and longer studies, and 3) longer follow-up on major clinical outcomes can be conducted and demonstrate clinical benefit. This combination of factors meant that HIV treatment development breaks the paradigm of drug development, taking it in a more potent and effective direction, the exact opposite of what the detractors predict.

2. Qualifying biomarkers for accelerated approval

The qualification of individual biomarkers to reach the "reasonably likely to predict" standard has been extremely slow in rare genetic diseases despite the particularly strong direct biological relevance of the biomarker and the treatment strategies that address the underlying cause [11]. The reason for this difficulty is unclear, but the FDA has not provided any specific guidance that would allow researchers to understand what data are required to qualify a biomarker, despite the passage of law such as the Faster Access to Specialized Therapies (FAST) act passed within the PDUFA framework a few years ago. That law specified that FDA must provide a guidance indicating how to qualify a biomarker using only pathophysiologic (e.g., animal data) and pharmacologic (pharmacokinetic and pharmacodynamic clinical data) criteria when other types of data (e.g., clinical outcome data) are impossible or impractical to collect. Other biomarker qualification programs at the FDA currently include the CDER biomarker qualification or drug development tools program [12] but this program is designed for common diseases and biomarkers not associated with a particular drug. This pathway does not work for rare genetic diseases, and an attempt to use this pathway for the development of a biomarker for MPS 7, Sly Syndrome, in an enzyme replacement therapy program was specifically rejected by that group [E. Kakkis, Personal Communication: Mepsevii Development Program].

3. The case for a new concept for drug development: measuring underlying primary disease not consequences of disease

To apply this new concept to inborn errors of metabolism (IEM) or other genetic or neurologic diseases, we would have to develop a better approach to defining and qualifying biomarkers than currently exists. Proposed qualification criteria have been developed publicly [10] and put forth [13]; however, qualification of new biomarkers even for genetic diseases has generally been extraordinarily difficult for regulators, with guidance only providing the vague "case by case" language [14]. The standards of data being required will basically ensure that drugs get approved first on clinical endpoints before any biomarkers would be qualified, thus defeating the value of AA and delaying the process.

Biomarkers as primary endpoints for AA, also called surrogate endpoints, have been criticized for failing to predict clinical benefit and led to backlash toward their use in the AA pathway [15–17]. More specifically, the dramatic results for encainide and flecainide - antiarrhythmic drugs that increased mortality were considered a cautionary tale toward the use of any biomarkers [16]. Most examples of failed biomarkers in the past have been downstream pathologic biomarkers close to the end-stage of clinical pathophysiology, and the treatments being tested were not addressing the underlying disease at all, but rather the symptoms. For example, treating arrhythmia itself to reduce arrhythmia, but not the diverse underlying causes of cardiac disease that can result in said arrhythmia fails to solve the underlying disease. Treating congestive heart failure by evaluating cardiac output with vascular active agents rather than measuring and treating the original cause of congestive failure will also be prone to failure. When a treatment is truly directed toward the underlying disease biology, like replacing an enzyme that is missing rather than a downstream pathophysiologic process, then the biomarkers with a more direct plausible link to the original cause of disease will have a much higher probability of predictive success, and these biomarkers are being termed Primary Disease Activity Biomarkers, or PDAB. This type of biomarker has a higher probability of success, as has been observed with historically approved drugs using PDABs at a time before greater stringency and obstructions to their qualification occurred. None of the drugs approved with biomarkers in rare genetic diseases have been subsequently deemed ineffective nor have any of these drugs been withdrawn (Table 1). Downstream pathophysiologic markers can seem closer to the disease course or clinical outcomes, but they are highly complicated by secondary processes and distant genetic loci that will confound their interpretation, especially in smaller rare disease studies. For genetic diseases, the biomarker covering the first step to assess a treatment designed for the underlying cause, will have much stronger biologic plausibility and reliability. These considerations and the supportive science have been well described previously in Kakkis et al. [13].

3.1. Primary disease activity biomarkers: a new biomarker definition

This biomarker type should be distinguished from others that have been described that represent downstream pathophysiologic or clinical outcome processes [18] in order to inform a scientific approach to defining, evaluating, and qualifying these particular biomarkers. The simplest examples of the PDAB come from biochemical genetic disorders. In phenylketonuria (PKU), a defect in the PAH gene prevents the liver from properly oxidizing phenylalanine (Phe), which builds up in the bloodstream and reaches the entire body including in the brain. In the brain, Phe is toxic to the neurons as well as to myelin-producing cells, and causes a distinct severe microcephalic developmental delay that is dependent for severity on the degree of PAH deficiency, the levels of Phe obtained, and the time and age of the patient to its exposure [19,20]. Serum Phe level reflects the liver content, and exposure to elevated Phe generally increases the brain Phe exposure that impacts the brain in vivo, but also its function in vitro. Phe level is a very good PDAB on the basis of its being a direct product of the enzyme block and its proven in vitro toxicity to various neural cell types. On a pharmacologic analysis, Phe can be lowered by dietary restriction and leads to a reduction in brain disease over many years, dependent in part on the lowering of Phe level and the time and age of Phe exposure. The more and earlier in life Phe is lowered, the less toxin exposure occurs and the less damage from the toxin occurs. In a young baby, this can save their brain from injury. In an adult exposed to high Phe from a young age, the damage is complete and the impact of reducing Phe is much more limited, relating to behavior, but intellectual function is not recoverable. Between the two extremes are a range of severity, and other genetic loci clearly have an impact on the degree of effect of different amounts of exposure [21]. Fortunately for patients with PKU, unlike many other untreated neurogenetic diseases, the historical dietary treatment data collected over decades of time was sufficient to gain acceptance of Phe as an endpoint in the first drug approved for PKU, sapropterin [22]. This drug enhances residual PAH activity in the liver which is a plausible natural biological mechanism for reducing Phe in a manner that should improve the brain. A second drug, pegvaliase, is an enzyme that does not restore the normal path but diverts Phe by cleaving it into nontoxic compounds in the bloodstream, thereby preventing Phe levels from getting high. Given the biologic knowledge of Phe and its actions, removal of Phe by PAH or by another enzyme diversion are also effective. From this case, a PDAB should have a direct biochemical/pathophysiologic origin, a direct impact on the causal pathway, and be close to the origin of disease. Pharmacologic studies in models should show that the reduction of the biomarker by that method would improve the disease state in a predictable manner and that the compartment sampled for the biomarker reflects the target tissue of disease. These principles are part of an extensive guideline published by stakeholders that could help guide the development of effective biomarkers [13].

Table 1
Biomarker used in rare genetic disease approvals: drugs, types and comments.

No.	Primary Endpoint Variable Used for Approval	Disease Indication	Established Name	Biomarker type	Comments
1	Hemoglobin concentration	Gaucher disease Type 1	Alglucerase 1992, Imiglucerase, Velaglucerase alpha	PDAB Tissue impact	First approval on two biomarkers and 12 patient open label trial
2	Spleen volume	Gaucher disease Type 1	Alglucerase 1992, Imiglucerase, Eliglustat, Taliglucerase alfa	PDAB Tissue impact	Spleen load of storage indicative of total body and has impact on comfort/platelets
3	Plasma ammonia levels	Acute hyperammonemia of urea cycle defect diseases	Phenylbutyrate 1996, Phenylacetate +benzoate Glycerol phenylbutyrate Carglumic acid	PDAB Ammonia accumulates from block in urea cycle	Ammonia is a neurotoxin in elevated amounts due to inadequate liver metabolism
4	Plasma uric acid (PUA) level	Chronic gout	Allopurinol 1966, Pegloticase	PDAB Uric acid from 1° block	Classic enzyme deficiency leads to toxic/precipitating compound
5	Serum ferritin levels	Thalassemia Syndromes Sickle cell disease	Deferoxamine 1968 Deferiprone, Deferasirox	PDAB Excess iron as oxidant	Iron builds up as transfusions are given and red cell iron is recirculated and absorbed
6	Percent change in LDL-C	Homozygous Familial Hypercholesterolemia (HoFH)	Lomitapide, Mipomersen, Evanicumab	PDAB High levels accumulate in blood vessels	High LDL-C levels directly "toxic" by being taken up by vasculature leads to vascular disease
7	Cystine in WBC	Cystinosis	Cysteamine 1994, Cysteamine bitartrate	PDAB Cystine accumulates in white cells	Cystine accumulation in white cells reflects cystine in diverse tissues in the body
8	Reductions in HbA1c, fasting glucose, and triglycerides	Leptin deficiency complications	Metreleptin	Mix of endpoint types	Leptin deficiency leads to a type II diabetes like phenotype and hence same biomarkers
9	Renovascular storage by biopsy/histology	Fabry	Agalsidase beta	PDAB Tissue impact	Storage in vasculature leads to renal failure eventually after many years. Marker for vascular disease in general
10	Phenylalanine level	Phenylketonuria	Sapropterin diHCl, 2007 Pegvaliase-pqpz	PDAB Phe builds upstream of the block in the liver	Phe is classic neurotoxin type and clearing Phe from bloodstream will protect the brain
11	Rickets severity scale	X-linked hypophosphatemia	Burosumab	PDAB Tissue impact	Low phosphate directly causes rickets, which leads to long term clinical manifestations like pain, bowing, dysfunction
12	Urinary oxalate	Primary hyperoxaluria (PH)	Lumasiran	PDAB Direct product of block	Treatment of PH had been stalled previously due to requirement for kidney function endpoint. Use of Oxalate dramatically changes the tractability of studies and source of stones

HoFH, homozygous familial hypercholesterolemia; LDL-C, low density lipoprotein cholesterol; PDAB, primary disease activity biomarkers; PH, primary hyperoxaluria; PUA, plasma uric acid; WBC, white blood cells.

Extrapolating beyond Phe, there are many endpoints in biochemical genetics that have a direct, well known genetic basis, the same pathophysiologic key role and evidence from pharmacologic studies that demonstrate the relationship between the biomarker and the result (Table 1). In Table 1, there are many examples of biomarkers used as primary endpoints where the value is clear. In fact, most of these examples are part of approved drugs over the last 30 plus years, and no biomarker-based approval of a biochemical genetic disorder has ever been withdrawn for failure of the drug to work. Many of these biomarkers were used in products first approved in the 1990s before there were as many challenges and sentiment against biomarker approvals [11]. The majority of the biomarkers in Table 1 are consistent with being a PDAB, as noted. It is not surprising from a scientific perspective that they were chosen, but also important that they have not failed, despite their first use in the 1990's, with minimal or no formal research to qualify their use. What is always true is that even for those biomarkers that work well, there are exceptions or variations in outcome or degree of clinical disease related to other environmental and biology aspects, including the subject's genetics; however, the standard of being reasonably likely to predict benefit is still achieved for the populations treated even if predictive value is not a linear proportional relationship to clinical disease.

XLH is another important disease for the consideration of PDAB development, as there are two biomarker endpoints that could be used. The one used in the approval of burosumab was a rickets score on bone X-rays. An X-ray of bones is not a measure of "Feels, Functions, or Survives," but it is easy to imagine that if the bones are bad, then the patient will have problems. This represents a tissue-damage biomarker of primary effect that would predict benefit from treatment. Phosphate level, which is low in these patients, could also be a credible PDAB biomarker endpoint that is more upstream in the disease process, although this was not accepted at first as there was less insight into what the ideal phosphate level needed to be. Since treatment improves phosphate into the lower normal range, and since this does impact the bones based on the X-ray rickets score endpoint, it is clear that normalizing phosphate is a good PDAB as well and may be better since it is upstream at the root cause of the biological problem. The value of phosphate as a PDAB over tissue-injury primary biomarker like X-rays is that phosphate level also impacts the muscles. Therefore, while X-rays tell us about one target tissue, bone, and should be acceptable, the true PDAB is phosphate, which tells us more about the primary cause of the problem and better represents the impact of the disease on the whole body, including bones and muscles. This is important because with two burosumab regimens that have nearly the same effect on the bones, the one that maintained phosphate steadily in the normal range rather than cycling up and down and had much better impact on the walking and functional activity of the patients [23]. Therefore,

phosphate was a better predictor of overall clinical function. However, for each patient, different phosphate levels could have different precise outcomes, which can create confusion but does not change the thesis. Getting patients to the low normal phosphate range was sufficient in all patients to generate important benefit, and phosphate remains an effective tool for regular monitoring of outcome.

When it comes the brain, an acceptable PDAB may be harder to define for some diseases. Some examples of PDAB that should be accepted, but are not currently, are shown below. PDAB could be biochemical measures in CSF or imaging measures of tissue impact. For lysosomal diseases like Tay-Sachs or Sanfilippo Syndrome, there are substrates that accumulate and are released to the CSF that are directly in the line of the pathophysiology. These substrate levels can be reduced by enzyme or gene therapy in disease models and show that the CSF samples decrease in a dose-dependent fashion. Based on this animal model data and clinical data from transplants, it is possible to demonstrate that CSF substrate levels can adequately support the impact of reducing the substrate accumulations and pathology in the brain. Given the large degree of variability and irreversibility in the MPS and other lysosomal disorders with CNS manifestations, the only effective way to demonstrate an impact on the disease is via a reduction in the PDAB substrate. To look at clinical outcomes will take multiple years of controlled studies and treatment before or early in the course before irreversible destruction has happened. Without newborn screening, this is hard to do. If not a controlled study, patients would need to be followed for many years to compare with sufficient change in the natural history, which makes for unsustainable development programs and creates a challenge for both ethical trial conduct and efficient insight into how to treat patients. This is a distinct situation from rapidly progressive diseases like late infantile Battens disease (CLN2) or Pompe disease (GSDII), in which the time course of symptoms to devastation occur in less than 1 year to 2 years and therefore, natural history controlled studies have been more tractable and have led to two approvals (Brineura and Myozyme) [24,25]. There are no slower progressive neurologic disorders with approvals in the lysosomal space. In fact, multiple programs are close to being abandoned because of the resistance of regulators to qualifiable PDAB biomarkers and pressure to adhere to clinical endpoints only, despite the variable irreversibility of these diseases at the time of diagnosis and treatment in studies. If these type of programs continue to be canceled [26,27], this will result in a further delay of a decade or more before we figure out the right way to do this. We need to change our approach right now if we ever expect to treat the myriad of diseases that affects patients.

3.2. The case for IEM as a key place for advancement of a new approach

Biology of many IEM require treatment before or early in symptom development. Clinical changes occur late in the course, far past the time of onset of definitive damage, due to the plasticity and adaptability of neurologic function to delay symptoms until too late. The consequences of the disease impact can often be irreversible and progressive. Many examples of these diseases and their biomarkers are shown in Table 2. In all cases, the accumulation of a biomarker in the CSF is a direct outcome of cellular lysis and release in the brain. In the cases of the MPS diseases, the cellular origin for GAG storage can be demonstrated because of distinctive chemical changes on the terminal sugar chains that can only happen in lysosomes [28]. Alternatively, imaging biomarkers can look at the primary tissue impact of disease where a biochemical marker is not available. In all cases, animal models show that intrathecal, IV, intraparenchymal by gene or enzyme replacement, or other strategies in mice and dog models will reduce storage in the brain tissue and thereby reduce the CSF substrates. These data provide sufficient support for understanding the degree of reduction required to have an important pathological effect on the brain. In situations where neurologic function can be measured (e.g., MPS 7 mouse cognition studies for enzyme replacement therapy), these reductions in pathology do have neurologic correlation when enough animals are studied over a long enough time [29]. Given the strength of the genetics, the biological plausibility of replacing a missing enzyme, and available data from animal models and human studies, the probability of biomarkers for approved systemic therapies being wrong is minimal, which means they should meet the standard of being reasonably likely to predict benefit. The exact amount of optimal reduction can be debated and can evolve as treatment progresses, but the data on these biomarkers can be used to estimate the magnitude of change likely needed to improve tissue pathology and set minimum thresholds for successful outcomes [30].

3.3. Impacts on the efficiency and effectiveness of drug development

If we are able to qualify PDAB-type biomarkers and construct development programs around them, there will be many improvements in our development efficiency. The PDAB should be qualified by work along the lines of that previously described [13] to assure that a quality validated assay is developed that can precisely measure the PDAB, and that the pharmacology and biology of the disease and drug is understood. If the Faster Access to Specialized Treatments (FAST) Act previously passed by Congress in the Prescription Drug User Fee Act (PDUFA) V [31,32] had been properly implemented by the FDA, a guidance would have been issued that included language to qualify biomarkers for accelerated approval based only on pharmacologic and pathophysiologic criteria when other types of data were impossible or impractical to collect. With that Act, the work for the ultra-rare diseases would have accelerated instead of stalled when there is no other way to get a drug developed for an ultra-rare disease.

PDAB-driven trials can be done in randomized controlled or open label formats of reasonable length and be successful in smaller sizes for ultra-rare diseases. Small, short studies could be used to establish the dose response and set up randomized studies of a tractable size. The efficiency and predictability will allow all comers to be enrolled instead of finely divided cohorts of subsets with a particular clinical problem at a particular stage of disease. If a randomized controlled trial cannot be done, the analysis of the PDAB can also be blinded and controlled with independent specimens from those not in a trial, to assure objectivity in the assays similar to what was done for rickets scores from X-rays in the burosumab pediatric XLH trial [23]. The cost of these development programs can be at least 70% reduced based on analyses we have done before, and this does not include the potential improvement in successful conduct or fewer inadvertent failures [33].

With the approval of a new drug by a PDAB program, high-quality post-marketing studies are required of sufficient length and design to capture the longer-term outcomes to demonstrate clinical benefit and certainly the FDA should be able to enforce that these studies are conducted. Traditional type registry data is often not of appropriate quality, with too much missing data and incomplete conduct. To solve this problem, a new design has been proposed called the Disease Monitoring Program, or DMP, that provides a high-quality GCP-compliant format that is fully sponsored to assure data are not missing and can continue collection for 10 years. The longer design is important to capture clinical outcomes that can take years to evolve and so there should be no expectation that post-marketing is completed within a short time. The design follows a larger number of patients then in approval trials, all on commercial-sourced drug, along with some patients who can be followed even if they have not opted to be treated. This design has been accepted by FDA for four approved programs to date. The approval of the first therapeutic products will also enable the establishment of newborn screening to find these patients earlier in life and treat prior to clinical onset. Finally, after approval, the one most important outcome is that the first approval will set a biomarker bar for others to improve on using appropriate therapeutic strategies and will result in more rapid improvements. Laronidase was approved in 2003 on clinical endpoints after the biomarker-driven first study was rejected for

Table 2
Diseases that have reasonable likely to predict PDAB biomarkers that are dependent on accelerated approval.

No.	Disease Indication	PDAB proposed	Analyte Specifics	Science supporting	Comments
1	Sanfilippo MPS III A, B, C, D	Heparan sulfate	Abnormal fragment in CSF evaluated by LC/MS/MS or NRE method	Significant animal model work and clinical work ongoing in gene therapy and enzyme replacement	A disease that has severe irreversible findings in all patients at symptom diagnosis
2	Hurler or Hurler-Scheie MPS I Hunter, MPS II	Dermatan sulfate Heparan sulfate	Dermatan for systemic disease measured in urine or blood. Heparan is for brain disease measured in CSF	Extensive trials in animal models and human clinical trials for urine, serum, and CSF GAG	CNS disease in severe MPS I and II are like MPS III in terms of irreversible disease and advanced state of disease at common diagnosis age
3	GM1 gangliosidosis	GM1 ganglioside	LC/MS/MS assay of CSF or serum/urine for systemic disease	Significant animal model work and ongoing gene therapy work in humans.	Another cumulated substrate that is released into the body fluids from sites of storage
4	GM2 gangliosidosis	GM2 ganglioside	LC/MS/MD assay of CSF	Significant animal model work in different types	Severe patients die so quickly that treatment is difficult with onset <1y and death in 1y. Less severe types are the ones needing the biomarker, since diseases can progress over years
5	Canavan	*N*-acetylaspartate (NAA)	Accumulating in CSF and measurable directly or by MRS	The missing enzyme that metabolizes NAA and leads to its excessive rise and multiple associated impacts	NAA may have complexities in its biological impact, but it is clearly the toxin in excess. Replacing the enzyme via gene therapy should be a direct context for monitoring improved NAA
6	Niemann Pick C (NPC)	24-hydrocholesterol	Released into CSF on treatment from storage in neuronal lysosomes	24-hydroxycholesterol accumulates in neurons and other cells; the drugs can release this storage so it comes out of the cells or goes up.	NPC1 and NPC2 are complex neurological diseases, but strong animal model data show that reducing the storage by gene or by solubilization/flipping using cyclodextrin can have an important effect on neurologic disease and storage
7	Alexander	Glial fibrillary acidic protein (GFAP)	Gain of function mutations that result in high GFAP levels	A gain of function mutation causes a rise in GFAP, which is toxic.	A gain of function is more difficult to treat than a missing enzyme, but the advent of numerous knockdown strategies makes this plausible today
10	Propionic or methylmalonic acidemia	Propionic acid Methylmalonic	LC/MS/MS measured in serum or elsewhere. Can measure one step derivatives as well, propionyl-carnitine, methyl-citrate	Strong data in patients and animal models showing propionic or methylmalonic acid is toxic to urea cycle, bone marrow, and brain.	The organic acidemias are well known toxic disease and there are tissue specific issues for systemic or brain content that both likely need to be addressed for success
11	Arginase deficiency	Arginine	Measured in serum	Both arginine and ammonia can be used to assess disease level.	Unlike other UCD, where ammonia is dominant toxin, ARG deficiency has elevated arginine which has its own neurological effect on spasticity

CSF, cerebral spinal fluid; GFAP, glial fibrillary acidic protein; GAG, glycosaminoglycan; MPS, mucopolysaccharidosis; MRS, magnetic resonance spectroscopy; NAA, *N*-acetylaspartate; NPC, Niemann Pick C; PDAB, primary disease activity biomarkers; UCD, urea cycle disorder.

approval (see Saving Ryan, a book describing this story [34]). Presently, Laronidase has been on the market for 19 years with all patents expired, and not a single improved product has gotten approved despite more than 10 companies working on next-generation therapies. This is entirely because of the challenging requirement to show improvements using clinical efficacy endpoints over an existing therapy in an ultra-rare and variable disease. This is tragic, as enzyme therapy was a reasonable temporary solution that should have started a process of medical evolution and yet this never proceeded, precisely because of the emphasis only on clinical endpoints. If the biomarker-based approval had happened originally, subsequent improvements would have been approved by now, just as occurred with HIV drugs.

3.4. The future of drug development in the 21st century: Time for a change if we are to treat the myriad of diseases with the multitude of technologies

The field of precision medicine, both in diagnosis and treatment, has provided a new opportunity to treat so many diseases that we have been unable to impact in the past. We will not do this well if we do not change the drug development paradigm and our own mindset and laws around defining what constitutes drug benefit. We know the causes of many diseases at precise levels, leading to the development of treatments targeting underlying pathophysiology, and can measure the impact of these treatments in precise biochemical ways to demonstrate an effect. The ability to compare a drug with a new modality or to develop combinations or a generic version would be dramatically improved. We need to define, as a society, that treating the underlying disease is treating the disease and not just an experimental result not worth approving nor paying for. The “Feels, Functions, Survives” standard will not enable the treatment of all diseases. It would be a tragedy to know that the great science we have created had been lost for procedural reasons, when on that most terrible day, genetic lightning strikes one of our own and no treatment is available.

Disclosures

EK: Employee and stockholder of Ultragenyx Pharmaceutical Inc.

References

[1] R.J. Temple, Clinical Measurement in Drug Evaluation, Wiley, New York, 1995.
[2] B. Bruen, E. Docteur, R. Lopert, J. Cohen, J. DiMasi, A. Dor, P. Neumann, R. DeSantis, C. Shih, The Impact of Reimbursement Policies and Practices on Healthcare Technology Innovation, 2016.
[3] S.W. Junod, Quick Guide to Clinical Trials, Bioplan, Inc., Washington, 2008.
[4] J.S. Birks, R.J. Harvey, Donepezil for dementia due to Alzheimer's disease Cochrane Database, Syst. Rev. 6 (2018) CD001190.
[5] N. Tambasco, M. Romoli, P. Calabresi, Levodopa in Parkinson's disease: current status and future developments, Curr. Neuropharmacol. 16 (2018) 1239–1252.
[6] Food and Drug Administration Modernization Act of 1997, S. 830, 105th Congress, 1st Session1997.
[7] E.D. Kakkis, The History and Challenges of Biochemical Genetics Markers and the Development of Qualification Criteria for Surrogate Endpoints Intended for Use in Accelerated Approval, Rare Disease Workshop Series, Washington, DC, 2011.
[8] W. Spearman, G.M. Dusheiko, M. Hellard, M. Sonderup, Hepatitis C Lancet, 394, 2019 1451–1466.
[9] L. Barr, R. Jefferys, A landscape analysis of HIV cure-related clinical trials and observational studies in 2018, J. Virus Erad. 5 (2019) 212–219.
[10] E.D. Kakkis, Accelerated Approval (AA) in Rare Diseases: Review of a White Paper Proposal, Rare Disease Workshop Series, Washington, DC, 2013.
[11] E.D. Kakkis, S. Kowalcyk, M.G. Bronstein, Accessing the accelerated approval pathway for rare disease therapeutics, Nat. Biotechnol. 34 (2016) 380–383.
[12] US Food and Drug Administration. CDER & CBER Drug Development Tool Qualification Project Search2021.
[13] E.D. Kakkis, M. O'Donovan, G. Cox, M. Hayes, F. Goodsaid, P.K. Tandon, P. Furlong, S. Boynton, M. Bozic, M. Orfali, M. Thornton, Recommendations for the development

of rare disease drugs using the accelerated approval pathway and for qualifying biomarkers as primary endpoints, Orphanet J. Rare Dis. 10 (2015) 16.
[14] US Food and Drug Administration, Expedited Programs for Serious Conditions––Drugs and Biologics, Guidance Document2014.
[15] T.R. Fleming, Surrogate endpoints and FDA's accelerated approval process, Health Aff. (Millwood) 24 (2005) 67–78.
[16] T.R. Fleming, J.H. Powers, Biomarkers and surrogate endpoints in clinical trials, Stat. Med. 31 (2012) 2973–2984.
[17] T.R. Fleming, D.L. DeMets, Surrogate end points in clinical trials: are we being misled? Ann. Intern. Med. 125 (1996) 605–613.
[18] R.M. Califf, Biomarker definitions and their applications, Exp. Biol. Med. (Maywood) 243 (2018) 213–221.
[19] S.E. Waisbren, K. Noel, K. Fahrbach, C. Cella, D. Frame, A. Dorenbaum, H. Levy, Phenylalanine blood levels and clinical outcomes in phenylketonuria: a systematic literature review and meta-analysis, Mol. Genet. Metab. 92 (2007) 63–70.
[20] E.D. Kakkis, The Development of Phenylalanine Blood Levels as a Surrogate Endpoint in Phenylketonuria, Rare Disease Workshop Series, Washington, DC, 2011.
[21] C.R. Scriver, The PAH gene, phenylketonuria, and a paradigm shift, Hum. Mutat. 28 (2007) 831–845.
[22] US Food and Drug Administration Center for Drug Evaluation and Research. Kuvan NDA 22–181 Approval LetterDecember 13, 2007.
[23] T.O. Carpenter, M.P. Whyte, E.A. Imel, A.M. Boot, W. Hogler, A. Linglart, R. Padidela, W. Van't Hoff, M. Mao, C.Y. Chen, A. Skrinar, E. Kakkis, J. San Martin, A.A. Portale, Burosumab therapy in children with X-linked hypophosphatemia, N. Engl. J. Med. 378 (2018) 1987–1998.
[24] US Food and Drug Administration Center for Drug Evaluation and Research. Myozyme NDA 125141/0 Approval LetterApril 28, 2006.
[25] US Food and Drug Administration Center for Drug Evaluation and Research. Brineura NDA 761052Orig1s000 Approval LetterApril 27, 2017.
[26] Takeda, Takeda Delivers Strong FY2021 Results; Continued Growth Momentum Expected in FY2022. Press Release, Available at: https://www.takeda.com/newsroom/newsreleases/2022/takeda-delivers-strong-fy2021-results-continued-growth-momentum-expected-in-fy2022/ 2022.
[27] Abeona Therapeutics, Abeona Therapeutics Announces Strategy Update and 2021 Financial Results. Press Release, Available at: https://investors.abeonatherapeutics.com/press-releases/detail/229/abeona-therapeutics-announces-strategy-update-and-2021 2022.
[28] R. Lawrence, J.R. Brown, K. Al-Mafraji, W.C. Lamanna, J.R. Beitel, G.J. Boons, J.D. Esko, B.E. Crawford, Disease-specific non–reducing end carbohydrate biomarkers for mucopolysaccharidoses, Nat. Chem. Biol. 8 (2012) 197–204.
[29] M.S. Sands, C. Vogler, J.W. Kyle, J.H. Grubb, B. Levy, N. Galvin, W.S. Sly, E.H. Birkeinmeier, Enzyme replacement therapy for murine Mucopolysaccharidosis type VII, J. Clin. Invest. 93 (1994) 2324–2331.
[30] E.D. Kakkis, D. Marsden, Urinary glycosaminoglycans as a potential biomarker for evaluating treatment efficacy in subjects with mucopolysaccharidoses, Mol. Genet. Metab. 130 (2020) 7–15.
[31] Food and Drug Administration Safety and Innovation Act, S. 3187, 112th Cong, 2nd Sess2012.
[32] Faster Access to Specialized Treatments Act; H.R. 41322012.
[33] B.E. Miyamoto, E.D. Kakkis, The potential investment impact of improved access to accelerated approval on the development of treatments for low prevalence rare diseases, Orphanet J. Rare Dis. 6 (2011) 49.
[34] E.D. Kakkis, Saving Ryan, Imposivity Media, 2022.