Clinical Outcome Assessment in Rare Disease Clinical Trials - Emerging Good Practices

INTRODUCTION:

A rare disease (RD), also known as an “orphan disease,” is one that affects only a small number of people within a population, although no universally accepted definition has emerged (Richter et al. 2015). The definition is most often based upon the prevalence of a disease within a specific country or geographical region (Richter et al. 2015). For example, in the U.S., an RD is defined as one with a prevalence of fewer than 200,000 persons (Rare Diseases Act 2002). The European Union (EU) defines an RD as one affecting fewer than one in every 2,000 persons within a population (European Parliament 2000).

RDs represent a wide variety of disorders and constellations of clinical signs and symptoms. Many are catastrophic, causing chronic or progressive physical degeneration, disability or premature death. Approximately 80% have a genetic etiology. An estimated 50-75% of RDs affect children (NIH 2013; Yaneva-Deliverska 2011). Moreover, 30% of children with a rare disease do not live to five years of age (PhRMA 2013). Presently, about 7,000 conditions are recognized as rare, although there may be many more (U.S. National Library of Medicine). On average, five new rare diseases are described every week in the medical literature (European Medicines Agency (a)).

Although few patients have any specific RD, taken as a whole, these conditions impact a significant number of people throughout the world. Recent estimates show that there are approximately 350 million people with a RD worldwide (The Global Genes Project). In the U.S. alone, there are approximately 30 million RD patients, with another 30 million residing in the EU (The Global Genes Project).

Patients with RDs typically have many unmet medical needs and few treatment options (Sibelius 2001). Diagnosis is often delayed by many years due to lack of health care providers with relevant clinical experience of these disorders (Kole and Faurisson, 2009). Appropriate treatments, if they exist, can be difficult to access (Kole and Faurisson, 2009). Moreover, they can be very expensive; the average list price per-person per-year cost has been estimated at approximately $137,782 USD (Picavet et al. 2014). (Note that this estimate does not include several very high-cost drugs such as Kalydeco and Glybera and may not reflect true costs of treatment after discounting).

Historically, small markets for RDs have discouraged research and development of targeted medications for these conditions. Several factors are changing this landscape, however. The rise of RD patient advocacy organizations and patient groups has increased awareness among policy makers and in pharmaceutical research communities regarding the paucity of effective treatments for RDs. As a result, many countries now offer special programs, e.g., financial incentives, rapid evaluation, and prolonged marketing exclusivity, to encourage development of RD medications. Medical innovation programs, not necessarily specifically
aimed at RD drug development, have also provided incentives for research and development of medicinal products for RDs.

These incentives, coupled with the advances in basic, translational and scientific research, particularly in genetics, have led to a surge in new treatments for patients with RDs (Meekings et al. 2011). Since the Orphan Drug Act was enacted in the U.S. in 1983, the U.S. Food and Drug Administration (FDA) has approved over 500 drugs for rare conditions (FDA Law Blog 2015). Almost 20% of all FDA drug approvals between 2000 and 2013 were for orphan drugs. In the EU, 96 new RD treatments were approved during the same time period (Tufts Center for the Study of Drug Development 2014). Many more treatments are currently in development (PhRMA 2013).

The increased focus on treatments for RDs is complemented by the recent efforts to include the patient perspective in all areas of medical research. Medical product efficacy is directly related to the impact of treatment on survival and/or on how patients feel and function in their daily lives (Patrick et al. 2007). Given the high cost of RD treatments, patients, as well as providers and payers, will need high-quality information on treatment benefits in order to make optimal care decisions and to ensure that each patient has access to the treatment most appropriate for him/her.

Evaluating the effectiveness of RD treatments from the patient’s point of view is therefore necessary to understand how to improve patient care and provide information that will allow patients to access the most appropriate treatments (Perfetto 2015). Clinical trial efficacy and safety outcomes need to be meaningful to patients, especially when patients and providers are faced with important choices about treatment strategies. These patient-centered outcomes are defined as, “those outcomes important to patients’ survival, function, or feelings as identified or affirmed by the patients themselves, or judged to be in patients’ best interests by providers and caregivers when patients cannot report for themselves” (Patrick 2013).

Patient advocacy organizations are supporting patient-centered outcome endpoints to measure treatment benefit in RD clinical trials. The recent position paper by the European Organization for Rare Disorders (EURORDIS) on patients’ priorities and needs for RD research emphasized the need to assess treatment options from the patient perspective, especially in terms of their impact on patients’ daily lives and functioning. Patient quality of life (QOL) was listed as a major research priority for patients with RDs (EURORDIS 2011). Furthermore, EURORDIS called for developing and validating patient-reported outcome (PRO) tools to support evidence of treatment benefit, as well as an increase in funding for research on patient QOL and a patient-centered approach to care.

There are significant challenges to developing, modifying or selecting clinical outcome assessments (COAs) for the evaluation of RD treatments. While all of these challenges may also be present for non-rare conditions, they tend to be magnified in RDs due to the small size of each RD patient population. Also, while one or two of these challenges may arise in non-rare disease populations, there is a greater likelihood of multiple limitations of this kind in RD populations. In addition to the size of patient populations these obstacles may include:

1. Heterogeneity of disease presentation within individual diseases;
2. Unknown or incomplete disease natural history for many RDs;
3. The large number of RDs that affect vulnerable populations, especially young children.

Additionally, RDs may be associated with significant, often progressive, disability and cognitive impairments (de Chalendar et al. 2014).
These obstacles make it more difficult to obtain relevant, valid and reliable information about the patient’s own experience with their condition and its treatment.

The debilitating, often progressive nature of many RDs and the young age of many RD patient populations pose challenges to collecting data on benefits of treatments directly from patients in the form of PROs. RD patients may not be able to self-complete questionnaires due to age, cognitive or physical impairment. In such cases, it is necessary to collect information regarding treatment benefit from clinicians, parents or others who have direct knowledge about the functional status or daily life of the patient. Thus, while use of PROs in RD studies may be desired this may not be practical. In such circumstances it will be necessary to use well-validated observer-reported outcomes (ObsROs), clinician-reported outcomes (ClinROs) and/or performance-based outcomes (PerfOs) to assess treatment benefits meaningful to patients.

This paper therefore addresses challenges inherent in identifying, selecting, developing, adapting and implementing clinical outcomes assessments (COAs) including not just PROs but also observer-reported (ObsRO) clinician-reported (ClinRO), and performance-based (PerfO) outcomes instruments for use in clinical research of RDs.

While there are considerable obstacles to implementing COAs in clinical trials of treatments for RD populations, some suggested solutions and methodological innovations have been posed (Leidy 2011; Benjamin 2013; Benjamin 2014; Vernon 2014; Perfetto 2014; Burke 2014). Where feasible, these and other sources to provide recommended approaches to identifying, adapting or developing COAs for RDs are included.

In summary, standard methods and strategies of COA identification, development, validation and implementation, including those recommended by regulators, need to be operationalized in the context of the special challenges associated with RD populations. This paper provides emergent good practices for adapting such methods and strategies in light of these challenges. The practices outlined in this paper can be considered a starting point for future empirical research to build a set of valid, scientifically sound and accepted best practice standards for capturing the patient perspective in RD studies.

**Task Force Objective**

The task force’s objective was to develop emerging good research practice recommendations for measuring COAs in RD clinical trials. The task force co-authors note here and emphasize throughout this paper the following:

1. Each RD drug-development program will present its own unique challenges to selecting and/or developing and implementing COAs;
2. There is no single solution for addressing the myriad and diverse difficulties that will arise with implementation of these measures and the development of evidence to support their use within various RD contexts of use;
3. This paper presents possible solutions to address common obstacles that arise when working in RD populations. Not all solutions will be appropriate or pragmatic for a given context of use; and
4. Within any specific RD drug development program, innovation, creativity, and flexibility will be required on the part of investigators to optimize outcome selection, evidence development, and study design.
Organization of the Task Force Report

Measuring how patients survive, feel and function is an integral component of the development, review and regulation of new treatments. For example, the goal of the U.S. FDA Clinical Outcomes Assessment (COA) (formerly a part of the Study Endpoints and Labeling Development (SEALD) staff) is to ensure that meaningful medical product information is available to health care providers, caregivers, patients and families through the advancement of innovation and excellence in clinical trial measurement of treatment benefit. This includes the development and implementation of standards for COAs used as effectiveness endpoints (FDA. Guidance for Industry and FDA Staff Qualification Process for Drug Development Tools. USDHHS; January 2014).

The FDA has published a roadmap to Patient-Focused Outcome Measurement in Clinical Trials (see Appendix 1), and the European Medicines Agency (EMA) is currently developing a reflection paper on the use of PROs in oncology studies (European Medicines Agency (b)). The FDA COA Roadmap outlines each step in the process for understanding the effects of a treatment, taking into account the condition of interest and the desired outcomes from the patient perspective, incorporating this knowledge into the selection, adaptation or development of appropriate endpoints reflective of outcomes meaningful and important to patients.

The organization of this paper draws from the FDA COA Roadmap, using this as a framework to discuss the challenges that arise in COA identification, modification, development and implementation in RDs according to each Roadmap column. Examples are provided to illustrate the RD-specific challenges associated with each column. Potential solutions are suggested to guide investigators through the process of choosing measures that are appropriate for the given context of use and for developing the evidence of the instrument’s validity and reliability necessary to support these endpoints within the context of a RD clinical trial.

Issues and challenges generic to development or selection of COA measures are described in other ISPOR Task Force Reports and will not be covered here (Walton et al. 2015; Matza et al. 2013). This paper will focus on those challenges specific to or magnified by developing or modifying COAs in the context of RDs.
**Figure 1. Roadmap for Implementing COA Endpoints in Rare Disease Clinical Trials**

### 1. UNDERSTANDING THE DISEASE OR CONDITION?

- **What is known about the condition?**
  - Natural history data may be limited
  - Heterogeneity in clinical manifestations over time and by disease subtype

- **How is it treated?**
  - Disease-specific treatments may not exist
  - Treatment variation across regions, age, groups, payers, subgroups

- **How does condition impact patients and caregivers?**
  - May differ by disease stage, subtype, age, region
  - Little data may exist

### 2. CONCEPTUALIZING TREATMENT BENEFIT

- **What constitutes meaningful treatment benefit?**
  - What is the concept(s) that a COA score needs to represent, i.e., the concept of interest (COI)?
  - Is the definition of treatment benefit similar across all subgroups that will be enrolled in the clinical study?
  - Can treatment benefit be measured directly?

- **How will the clinical study be designed, i.e., the context of use (COU)?**
  - What are the inclusion/exclusion criteria?
  - What is the hierarchy of endpoints (primary, secondary, exploratory)?

- **Which COA types are needed?**
  - PRO
  - ClinRO
  - ObsRO
  - PerfO

### 3. SELECTING/DEVELOPING OUTCOME MEASURE

- **Are there any extant COAs that are appropriate?**
  - Does the score represent the concept of interest (COI)?
  - If no, can the COA be modified to fit the COU?
  - If no, are there time and resources available to develop a new COA?

- **How to develop or adapt the COA for context of use?**
  - Traditional methods may not be feasible
  - No one size fits all solution!
  - Take advantage of all opportunities for patient engagement and qualitative research to optimize COAs for use in the clinical study

---

*Adapted from Roadmap to Patient-Focused Outcome Measurement in Clinical Trials (Food and Drug Administration - Roadmap to Patient-Focused Outcome Measurement in Clinical Trials).

**Column I: Understanding the Disease or Condition**

Typically, in COA instrument selection or development, an understanding of the natural history of disease and relevant patient-centered outcome concepts is obtained through review of the literature, expert clinician interviews, and qualitative interviews with patients and/or their caregivers. However, when working in RDs, there are often large gaps in knowledge of the etiology of the disease or condition. The small number of patients and caregivers available to participate in qualitative studies to understand the patient experience, in
combination with lack of natural history data, can make it difficult to develop an understanding of the relevant disease experiences and important outcome concepts.

Frequently, the full spectrum of symptoms and clinical manifestations of an individual RD is not well described in the medical literature. Given the rapidly expanding knowledge base about RDs, publications might lag behind the scientific body of knowledge. In many cases, symptoms and clinical manifestations may not be well understood by the medical community, or even by the few expert clinicians that have experience with the disease.

For example, genetic diseases typically have multiple phenotypes, ranging from severe, rapidly progressive forms that often present in early childhood, to forms that progress slowly and that may not come to medical attention until adulthood. A related problem is that patients are diagnosed at different stages in their illness, and often not until the condition is advanced. These patients are frequently misdiagnosed or undiagnosed for a significant time after the first symptoms appear. Knowledge is quickly accumulating and often progress in disease definition diagnosis and treatment precedes publication. As RD is a global concern finding information requires a global perspective.

While there is a dearth of peer-reviewed research publications on diagnostic delays in RDs, two large survey studies have documented this problem. A recent report found that accurate RD diagnosis took on average about 6-8 years (Shire 2013). A survey of almost 6,000 patients or caregivers in 16 European countries found that, median time from first symptoms to correct diagnosis ranged from 15 months to 28 years for 75% of the sample (EUORDIS 2009). A study of 233 Hereditary Haemorrhagic Telangiectasia (HHT) patients found a mean of almost 26 years between first symptoms and definitive diagnosis (Pierucci et al. 2012). Thus, the early stages of these conditions can be difficult to understand and are often not well described.

Depending on the patient population, different treatment goals or clinical assessments may be more appropriate for one subpopulation versus another. One example is Pompe disease, a lysosomal storage disorder with clinical manifestations that are largely due to progressive muscle impairment. The infantile-onset form is rapidly progressive. If left untreated, it will result in death before 18 months of age for most patients, primarily due to cardiorespiratory failure. The adult-onset form is slowly progressive with gradual declines in motor and respiratory function, but with no cardiac involvement, and patients may live into their 6th and 7th decades.

A careful description of the full range of the clinical manifestations, timelines for disease progression, and clinical outcomes for the different phenotypes are extremely important in order to assess and select outcome measures that are meaningful to patients with different disease subtypes or at different stages of their condition. Another common scenario in RDs is that a group of diseases may be clustered into a syndrome based on similarities in clinical manifestations, although they are actually different conditions with important differences in their genetic origins, symptom profiles and patient experience of disease.

Batten disease is one such disorder. A progressive neurodegenerative disorder, it can be caused by one of at least eight different genetic mutations to the CNL gene. The individual disorders have similarities in their clinical manifestations, notably progressive neurocognitive impairment. However, there are also important differences between them that have significant diagnostic, prognostic and treatment implications. Thus, a careful description and understanding of the clinical course of the individual disorders is necessary to support the selection of outcome tools and to evaluate whether or not a common COA tool is appropriate for all conditions grouped within the syndrome.
Without sufficient available information on the disease or condition (Column 1), it can be problematic to conceptualize treatment benefit, to design a clinical study to test those benefits (Column 2), and to select or develop the appropriate outcome measure to evaluate these benefits (Column 3). A variety of information sources may be necessary to fully understand the disease or condition and to construct an appropriate COA measurement strategy.

Table 1 provides a summary of potentially available data sources to evaluate and review when developing an understanding of RDs with limited natural history data. The difficulties associated with understanding rare conditions and their treatments are listed in the first column, and the potential solutions are presented as column headings. Check marks in individual cells indicate a potential solution relevant to a specific obstacle.
### TABLE 1. Potential Information Sources for Rare Conditions

<table>
<thead>
<tr>
<th>CHALLENGES</th>
<th>SOURCES OF INFORMATION</th>
</tr>
</thead>
<tbody>
<tr>
<td>A. Incomplete Understanding of Rare Disease Clinical Factors</td>
<td>Local Policies &amp; Regulations</td>
</tr>
<tr>
<td>• Incomplete natural history of disease</td>
<td>Claims Data</td>
</tr>
<tr>
<td>• Misdiagnoses/Inappropriate treatments</td>
<td>Expert Clinician Interviews</td>
</tr>
<tr>
<td>• Difficulty distinguishing between RD signs and symptoms and those due to the consequences of inappropriate treatments</td>
<td>Clinical Trial Data</td>
</tr>
<tr>
<td>• Varying time to diagnosis</td>
<td>Patient/Caregiver Interview</td>
</tr>
<tr>
<td></td>
<td>Observation Studies, Registries</td>
</tr>
<tr>
<td></td>
<td>Simulation &amp; Modeling</td>
</tr>
<tr>
<td></td>
<td>Post-Hoc Analysis of RCT Data</td>
</tr>
<tr>
<td></td>
<td>Patient Advocacy Groups</td>
</tr>
<tr>
<td>B. Unknown Subpopulation Factors</td>
<td></td>
</tr>
<tr>
<td>• Highly heterogeneous disease manifestations</td>
<td></td>
</tr>
<tr>
<td>• Unknown range of severities along disease trajectory</td>
<td></td>
</tr>
<tr>
<td>• Disease variations caused by unknown factors</td>
<td></td>
</tr>
<tr>
<td>C. Varied Health Care Environment</td>
<td></td>
</tr>
<tr>
<td>• Variation across nations and regions</td>
<td></td>
</tr>
<tr>
<td>• Lack of standards of care or disease management guidelines</td>
<td></td>
</tr>
</tbody>
</table>

- Local Policies & Regulations
- Claims Data
- Expert Clinician Interviews
- Clinical Trial Data
- Patient/Caregiver Interview
- Observation Studies, Registries
- Simulation & Modeling
- Post-Hoc Analysis of RCT Data
- Patient Advocacy Groups
The following sections provide further details regarding RD challenges and approaches for each aspect of understanding the natural history, health care environment and patient perspectives for specific RDs.

Column I: Understanding the Disease or Condition: Challenges and Solutions

1. Challenge: An Incomplete Understanding of the Disease Natural History

RDs generally have poorly or incompletely understood natural histories. This poses challenges for COA selection, development, and/or implementation because the relevant concepts of interest are not completely apparent without this information. While this limitation is also a concern for non-rare diseases, the newness of many recently described RDs, along with the small patient populations, disease subtypes, heterogeneity of severity and symptoms, and the frequently times to correct diagnosis makes this especially problematic for these conditions. In addition, natural history data is rarely available for recently described RDs.

Suggested Solutions:

i. Use all available sources of information to understand natural history, including observational studies and registries, case reports, opinion of clinical experts and literature reviews as well as patient journey interviews with patients and caregivers. These sources can be starting points for building a disease knowledge base, and can help to identify basic features of the disease, distinctions between different genotypes/phenotypes, prevalence and any geographic clustering of disease subtypes or treatment patterns

ii. The RD community is highly motivated to advance science, research and patient care practices. Engaging with RD patient advocacy groups to recruit patients for patient journey studies, disease registries, or other natural history studies can assist with developing a deeper understanding of the disease or condition and is encouraged wherever possible.

iii. Given the lengthy time to diagnosis associated with many RDs, it may be especially important to understand timeframe from first disease symptoms to diagnosis. Improving the understanding of early disease progression can be crucial for reducing time to correct diagnosis and developing effective strategies for early intervention.
2. **Challenge: Diversity in Disease Presentation (Heterogeneity) and Patient Experience**

There can be substantial heterogeneity in RDs, both between and within disorders. For a variety of reasons (Table 2), this heterogeneity can result in a wide diversity in initial presentation, clinical manifestation and progression of the disease between patients with the same underlying condition. This heterogeneity, along with regional and cultural differences in how the RD is perceived and treated, can lead to different perspectives within a single RD patient population regarding what constitutes a meaningful treatment outcome. This poses unique challenges to identifying the most appropriate patient-centered concepts of interest and parameters for measurement of the treatment outcomes most salient and important to all patients.

### Table 2. Reasons for diversity in RD presentation

<table>
<thead>
<tr>
<th>Reason for Diversity</th>
<th>Consequences of Diversity</th>
</tr>
</thead>
</table>
| Phenotypic diversity | • Variations in treatment effectiveness  
|                      | • Variations in the type, timing and severity of symptoms for patients with the same underlying disease (Lesko 2011) |
| Disease progression  | • Core signs and symptoms of the disease may not be stable over course of disease |
| Delayed diagnosis    | • Variations in types and patterns of treatments  
|                      | • Variations in disease trajectory due to differences in disease progression |
| Incorrect diagnosis  | • May result in inappropriate treatments  
|                      | • May result in iatrogenic illness  
|                      | • Signs/symptoms associated with inappropriate treatment can be hard to distinguish from the RD itself  
|                      | • May change trajectory of RD |
| Few standards of care| • Variations in treatment patterns by geography or payer  
|                      | • Variations in patient health status, treatment outcomes |
| Regional and cultural differences in meaning, significance of RDs | • Patient perception of disease experience may differ based on:  
|                      | Geography  
|                      | Culture  
|                      | Clinical management  
|                      | Significant life events |

**Suggested Solution:**

1. Focus on the most common symptoms and impacts of the condition of interest that: 1. are most important to patients; and 2. would be expected to change with treatment given the therapy's (hypothesized) treatment effects and target product profile (as per the COU).

**EXAMPLE:** All three major types of Gaucher disease involve a symptom cluster that includes spleen and liver enlargement, anemia and thrombocytopenia. These symptoms can appear at any time from early infancy through adulthood and can range in severity from mild to very severe, but they are considered hallmarks of the condition and are consistently found throughout the course of the disease (Mignot et al. 2013).

2. In order to identify and focus on the most prevalent symptoms and impacts of the condition, the following recommendations may be helpful:
a. Develop an understanding of the disease natural history using all available sources of information (See Column I, Challenge I).

b. Conduct concept elicitation interviews with patients and/or caregivers and clinicians to understand disease experience over time. When gathering data from patients and caregivers, obtain information about patient treatment and wider clinical management history. This will be helpful in differentiating signs/symptoms of disease from those due to other causes.

If possible, select a range of patients and caregivers, representative of various ages, disease stages and relevant national or demographic subgroups, to ascertain differences and commonalities in terms of how the RD is experienced, the importance and impact of symptoms, and societal perceptions of/reactions to the condition.

iii. To understand how regional diversity may impact RD presentation, the following suggestions may be helpful:

a. Engage clinical experts from different regions with knowledge and experience in diagnosis and treatment of the specific RD to understand geographic differences in patient characteristics and disease management strategies.

b. Other sources of data such as registries or national surveys, claims databases and specific payer policies or decisions can also be helpful in understanding regional or geographic variations in patient experience, standards of care, access to clinical expertise or treatments, optimal management, etc.

Work with local patients and patient advocacy organizations to obtain a representative picture of treatment patterns, from first symptoms to definitive diagnosis and beyond, and to understand the patient experience throughout this trajectory.

iv. It is also useful to understand the consequences of diagnostic delays on patient outcomes. Those diagnosed early in their illness may have a longer time to disease progression, fewer or less severe symptoms, and lower risk of iatrogenic illnesses. Obtaining information about time from first symptoms to correct diagnosis and treatment is important to document when developing the conceptual model of disease and subsequent COA measurement strategy.

3. Barriers to Obtaining Input of Patients and Caregivers for Studies of Instrument Validation or Development

Due to the rarity of the condition, it can be difficult to identify patients and caregivers to approach for engagement in studies. Once patients or caregivers are identified, there may be obstacles or practical factors that make their participation in research activities difficult. For example, the small size of the target population makes it impractical to organize focus groups or conduct face to face interviews in an efficient manner. Practical considerations (e.g., transportation, need to modify work or treatment schedules, finding time away from caregiving to participate, etc.) can also influence a patient’s or family’s ability and willingness to participate in a study.

Moreover, as the number of new potential treatments for RDs increases, there is a danger of participant fatigue, especially where patient populations are very small and there is competition for their participation in
COA researchers must be sensitive to these issues and design studies that are feasible for the target population and they must ensure transparency about the potential risks or benefits of participation in a study (Godlew and Furlong 2012).

Suggested Solutions:

i. Where they exist, partner with a patient advocacy organization (PAO) to design the study as well as support recruitment efforts. PAOs can frequently broker the relationship between the researcher and patient communities, either directly or through their work with clinicians. However, negotiating this relationship can be a sensitive proposition. Several strategies may help to overcome PAO’s potential concerns and help to ensure a successful research partnership:

a. Before contacting a potential PAO, understand their focus and scope. Does the PAO have direct access to patients? Is there a focus on working with researchers and assisting with study design, recruitment, etc., or is the organization largely focused on supporting families and patients in finding clinical care resources, or providing a platform for the patient community to exchange experiences and information? Do they have a well-defined social media presence? Do they work with an existing umbrella organization? All of these factors may impact on a PAO’s ability and willingness to support patient engagement and mobilization for participation in research initiatives.

b. Partner with the PAO in designing the study and/or providing substantive review of study materials (interview guides, protocols, recruitment messages) or study results. This may increase buy-in and commitment to the research project and will ensure that the study is relevant and appropriate for the target population.

c. If there is an umbrella organization for the RD community of interest, it may be useful to contact this group rather than attempting to directly contact a specific local PAO. The umbrella organization usually has good contact with local organizations, which can be quite small. They also have the most resources and can help put out the word through email blasts, social media etc. The umbrella organization can provide guidance on which PAOs are best suited for research purposes. Umbrella organizations can also assist in brokering the relationship between the research team and the local PAOs.

While not all RDs have such umbrella groups, many do. For example the European Gaucher Alliance (https://www.eurogaucher.org/) includes member PAOs in 41 countries across the EU, Israel, Canada and the UK. The OrphanNet website includes a very comprehensive list of PAOs, searchable by disease. (http://www.orpha.net/consor/cgi-bin/SupportGroup.php?lng=EN) or Rare Diseases https://www.rarediseases.org/patient-orgs/current or EURODIS http://www.eurordis.org/content/european-disease-specific-federations

ii. Consider collecting at least some data using internet-based resources. Analysis of relevant blogs or “member stories” can enrich understanding of the patient or caregiver experience, and it can provide a richer and more comprehensive picture of patients across geographic regions or demographic characteristics than would be possible in a small qualitative study with targeted recruitment.

Sites such as RareConnect.org (https://www.rareconnect.org/en) provide a forum for a large number of RD communities and may contain a wealth of patient and caregiver comments and information. Before implementing such a strategy it is advised to contact the site sponsor or community site
moderator to alert them to your presence and purpose, and to confirm that there are no restrictions on these activities by a non-patient participant as your actions may be considered an intrusion.

Note that a limitation of these forums is that there is no independent clinical verification of the information provided by the participating individuals and no method of confirming the RD diagnosis they claim to have. Further, Internet access, motivation to blog or communicate in an open forum, and technical ability these individuals possess may not typical of most patients, and certain experiences may be modified or altogether absent depending on the needs and circumstances of the writer. Thus, data based on this source experiences cannot be considered representative of all patients with the RD. However, these sources can be helpful in generating a hypothesis and informing future, targeted research with patients with a confirmed diagnosis and known clinical characteristics.

iii. Many pharmaceutical companies developing products for RD indications have one or more liaisons responsible for engagement with patients and/or patient groups. If the COA development study is conducted under their auspices, typically this designated individual engages patients and/or groups.

iv. Partner with an expert center or dedicated clinical care network to identify and recruit potential study participants. Such organizations can have access to patients beyond their geographic region as the lack of local clinical expertise results in the need for patients to travel to obtain needed care.

Column II Challenges & Solutions: Conceptualizing Treatment Benefit and the Measurement Context

Meaningful treatment benefit common to all patients with a specific RD can be especially difficult to identify due to several different factors that may influence how patients experience their illness and their expectations for how a treatment can improve their condition. The small size of target patient population frequently necessitates including participants from multiple geographic regions as well as a wide range of age groups. What constitutes “normal” functioning at any one stage of life may vary, and expectations for improvement will also be influenced by the degree of patients’ impairment (Asadi-Lari et al. 2004).

There can also be differences in illness-related cultural values, beliefs and expectations across various ethnicities or regions. Phenotypic heterogeneity within a single RD may make it difficult to arrive at a single concept of treatment benefit across all subtypes of the condition; patients with the same RD can experience different symptoms, ages of onset, or severities of illness.

An example of the complexities involved in defining treatment benefit due to these factors can be found in amyloidosis, a condition caused by misfolded proteins that can deposit in organs and tissues, disrupting normal functioning in a variety of systems (see box appendix X). In all, 28 different proteins are associated with amyloidosis, 14 of which have systemic manifestations (Brambilla et al. 2013). It may therefore be necessary to include a wide variety of patient subgroups in a study although there may be only few individuals in each subgroup. Subgroup analysis in such situations can be difficult if not impossible.

Achieving concept saturation to identify the main symptom and impact themes can be similarly challenging.

A high percentage of RDs affect children where there are rapid and often variable cognitive and physical developmental changes both within and across age groups (Papadopoulos 2012). Other challenges related to identifying COIs can include patients’ difficulties in distinguishing between symptoms due to iatrogenic or comorbid conditions from those associated with the RD itself. Thus, the main strategy for identifying appropriate patient-centered outcomes in RDs is to focus on the core signs, symptoms and impacts of the RD that apply to all (or most) patients with the condition, and which can be measured across a range of...
ISPOR COAs in RD Trials Task Force: DRAFT for Review Only

Ages and severities of disease (Papadopoulos 2012). If this is not possible, a modification of the COU may be necessary to address the specific concerns of this population and range of disease types and severity.

Expectations regarding treatment benefit can be affected by a combination of:

1. What constitutes “normal” or “usual” functioning at any one stage of life;
2. Degree of patients’ impairment (Asadi-Lari et al. 2004);
3. Disease progression - patients in more advanced stages of disease may conceptualize acceptable benefit in a very different way from those who are not as impaired; small improvements may be acceptable for more impaired patients - a common problem in defining treatment benefit across the disease continuum.

For some patients, stable disease with no, or slowed, progression, is a desired outcome; others may desire clear symptom improvements and better physical functioning. In the case of systemic amyloidosis patients where different subtypes have different kinds of organ involvement, treatment benefit may constitute very different symptom improvements for someone with cardiac involvement and someone else with liver involvement.

In order to generate meaningful information about treatment benefit, the data gathered as described in Column I is used to conceptualize the optimal outcomes from the patient’s perspective (the concepts of interest, COI) for testing in a well-designed clinical study. An adequate COA must generate a score or rating that represents the outcome stated in the study objectives. Therefore, the treatment benefit must be conceptualized in the context of the future study. Establishing the COA (e.g., the identification, modification or creation of the COA, as described in Column III) cannot take place until a meaningful treatment benefit is determined (i.e., the COI is defined) and the study context of use (COU) is established. It should be noted that neither COU nor COI identification can be completed independently of the other. Identification is an iterative process.

Specific impediments to defining the COI and COU in RDs can arise from: 1. Lack of understanding of the natural history of the condition; (2) Heterogeneity of disease presentation; (3) Incomplete understanding of treatment benefits due to difficulty in accessing patients or caregivers to provide a comprehensive perspective on the most prevalent and important signs/symptoms of the disease to assess in the clinical trial. While all of these obstacles were discussed previously under Column I, the challenges and potential solutions specific to the establishment of the COI and COU are discussed in this section as outlined in Table 2.

### Table 2. Challenges and Suggested Solutions: Conceptualizing Rare Disease Treatment Benefit and the Measurement Context

<table>
<thead>
<tr>
<th>CHALLENGES</th>
<th>SOLUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Focus on measureable core signs, symptoms, and impacts that are experienced across all patients</td>
<td>Utilize study population selection criteria that focus on exclusion of other diseases</td>
</tr>
<tr>
<td>Group patients by similar treatment patterns to control for regional context variations</td>
<td>Collect additional data on environment to ensure correct stratification parameters for analyses</td>
</tr>
<tr>
<td>Utilize electronic data capture to collect data across regions</td>
<td>Conduct concept elicitation study to evaluate differences in patient experience between disease</td>
</tr>
</tbody>
</table>
### 1. Challenge: Identifying Concept(s) of Interest for Meaningful Treatment Benefit

As described above, it is often difficult to ascertain a clear and consistent understanding of treatment benefit in RDs due to the heterogeneity of the condition itself as well as the wide variation in patient experiences in treatment and cultural attitudes. The small size of the target population further complicates matters because it is necessary to include patients with a wide range of ages, degrees of disease progression and expectations for improvement in their condition.
Suggested Solutions:

i. Because the importance of core disease characteristics may vary by patient subgroup or change over time, conduct concept elicitation studies with as many relevant patient/caregiver groups as possible/feasible to identify those most salient to all patients.

ii. Consider what the most important differentiating characteristics are in terms of disease subtypes, age group, ethnic groups, degree of progression or other specific factors that may lead to variation in disease experience and perceived treatment benefit.

iii. Consider severity, nature, timing and degree of involvement of affected body systems.

iv. Use natural history information to understand how symptoms may vary across different ages and stages of illness.

v. Consider developing a COI definition that takes into account the important symptoms in a variety of subtypes, with a COA strategy that includes measurement of a variety of outcomes in a single scale or symptom battery.

2. Challenge: Difficulty Distinguishing RD-specific Clinical Signs and Symptoms from Iatrogenic Clinical Characteristics

As stated previously, diagnosis can be complicated by iatrogenic illness resulting from earlier accurate or inaccurate diagnosis and consequent appropriate or inappropriate treatments.

Suggested Solutions:

i. Focus on core symptoms and impacts.

ii. Select well-defined signs and symptoms based upon what is known about the disease (as per the COU), taking into account variations due to disease subtypes.

iii. Work with clinical experts who can differentiate possible treatment-related symptoms (iatrogenic as well as appropriate treatment side effects) from disease symptoms.

3. Challenge: Understanding the Relationship between Primary Clinical Endpoint and COA Endpoints

The most appropriate primary endpoint(s) in a clinical trial may be as yet unidentified, not widely accepted or not validated. Surrogate endpoints are frequently used in RD clinical trials, where the patient population is too small and/or the time to observe the true clinical outcome is too long, i.e., the inability to achieve the primary clinical endpoint within the trial duration, or due to the relative rarity of the event of interest (EMA Guidance on Clinical Trials in Small Populations 2006).
**Suggested Solutions:**

i. Work with clinical experts early in the design of the clinical trial program to align COA strategy with expected impacts of selected primary endpoints and hypothesized treatment benefit/effect size. Make sure that the relationship of primary endpoints to COA measures is as completely defined and explained as possible.

ii. Conduct a concept elicitation study to understand the meaning and importance of all clinical endpoints to patients and/or caregivers, especially regarding the relationship of endpoints to patient/caregiver treatment experience, social, emotional and physical function and RD-specific symptoms.

iii. COAs are especially important in helping to validate whether a surrogate endpoint (e.g., a biomarker intended to substitute for a clinical endpoint, or an objective indication of a biologic process or treatment response; Aronson, 2005) reflects the actual clinical endpoint adequately, e.g., it can predict treatment benefit. In this situation, it is crucial to ensure that any change in the COA is correlated as closely as possible to the surrogate endpoint as well as other endpoints in the study where this makes clinical and logical sense.

### 4. Challenge: Developing COA Measurement Strategy Appropriate for Non-Traditional Clinical Trial Designs

Because many RD drugs are fast-tracked through the regulatory approval process due to the severity of the indication and the lack of available treatment options, there is increased regulatory emphasis on alternative approaches to studying drugs with RD indications. This includes consideration of non-traditional clinical development programs for faster, more efficient data collection (e.g., adaptive trials, N of 1 trials, etc.), and the utilization of alternative study designs, endpoints, and statistical analyses (Hamburg 2011). The RD COA strategy must be feasible within the context of these alternative study designs.

**Suggested Solutions:**

i. Focus on measurement strategies that emphasize short-term outcomes for the trial and then collect longer-term outcomes in post-approval observational studies or in patient registries.

### 5. Challenge: RD Clinical Trials Are Likely to Be Multinational and Include Patients in Multiple Age Ranges

Due to small patient populations, RD clinical trials are typically multinational and often include individuals across a broader range of demographic characteristics, e.g., age groups. The progressive nature of many RDs may lead to a substantial variation in the effect size displayed in response to the treatment under investigation across patients in different stages of disease, as well as across patients with varying disease progression during the course of the study.

**Suggested Solutions:**

i. Build on information obtained in Column 1 to plan COA measurement strategy that addresses the differences in patient demographic, health care environment and perceptions of treatment benefit, risk and disease impact.
ii. Where possible, stratify patients by relevant characteristics (nationality, age group, disease severity at baseline, etc.) to understand systematic differences in outcomes, and enable standardization of measurement of treatment benefit.

iii. Consider a measurement strategy that includes multiple types of measures (PROs, ClinROs, and ObsROs) of the same construct(s) to capture outcomes data for both children and adults.

**6. Challenge: RD Presentation Can Change Over the Trajectory of the Illness**

Signs, symptoms, and feelings may not be stable over the course of the trial or across different age groups. Disease manifestations can vary across the lifespan and changing developmental characteristics (e.g., verbal ability) can influence patients’ ability to provide self-report. Patients with progressive disease may worsen over time and be unable to complete self-report instruments at later points in the study, or those with more progressive or severe disease may be unable to self-report while those with less severe involvement may be capable of completing a PRO. Children age and can become able to independently complete a PRO whereas at the start of the study they were unable to do so.

For example, eosinophilic esophagitis is an example of a chronic disease with signs and symptoms that differ by age. In infants, food refusal is commonly observed. School-age children often suffer from gastro-esophageal reflux-like symptoms, vomiting and abdominal pain. Adolescents and adults experience mostly dysphagia with solids and food impaction (Liacouras et al. 2011; Noel et al. 2004).

**Suggested Solutions:**

i. In longitudinal studies, or in studies that include a wide range of patient age groups or disease severities, it may be necessary to administer measures in different formats. Consider using parallel forms of instruments measuring the same constructs may be required, for example the CHO-KLAT was developed to measure symptoms in pediatric patients with severe hemophilia-A (Young et al. 2004). Several different forms were constructed and validated for use in different age groups, including an instrument specifically aimed at younger children, another for adolescents, and an ObsRO version for completion by parents of very young children. (See appendix X for a case study example).

   a. Focus on measures of symptom relief (e.g., reductions in severity, duration or frequency) if treatment benefit does not include slowing or preventing progression of disease (EMA Guidance on Clinical Trials in Small Populations 2006).

   b. Focus on core symptoms that are characteristic of the RD across age groups and over time, and that are expected to be affected by the treatment MOA. While some symptoms in COAs may be age-specific, a core set that appear to be most common across the trajectory should be kept consistent.

ii. In the case of young children or those with cognitive impairment who cannot provide valid and reliable self-report, ObsROs, based upon observable indicators of patient functioning (e.g., signs or behaviors) as reported by a parent, caregiver or other observer can provide indirect evidence of treatment benefit (FDA 2009).

iii. Focus on measures of activities of daily living (ADL) or social functioning for patients who will remain severely disabled, even with treatment (EMA 2006). However, be aware that, in studies that include patients with a range of disease severities, stages of progression, or developmental ages, the
measures must contain measures of ADL or social functioning reflective of the range of patient abilities to avoid floor and ceiling effects.

iv. Behavioral manifestations of the same symptom may vary by age group. Conduct interviews with clinicians, caretakers and/or patients to understand how to measure such symptoms over time or in study participants of different ages, adapt COAs for these variations.

v. Prioritize core concepts in the COA measurement strategy.

**Column III Challenges & Solutions: Selecting/Developing the Outcome Measure**

Once the COI and the COU have been established the process of selecting, adapting and validating the COA measure(s) for a specific application can begin. Few disease-specific COA instruments are available with published evidence of measurement properties for use in most RDs. Thus, when embarking on evaluating whether there is an existing COA instrument that is suitable for use in a specific patient population, special attention needs to be given to establishing the most precise definition of the concept of interest (COI) and the context of use (COU) as possible.

Once a list of potential COAs for measuring the COI has been determined, a standardized evaluation process begins whereby the COAs are evaluated to establish whether they are “fit for purpose” and thus appropriate for inclusion as important efficacy endpoints to demonstrate treatment benefits meaningful to patients. Defining the COU also assists with identifying other clinical data that would help in validating and interpreting results (Trivedi et al., 2013).

In addition, we recommend taking into account from the beginning that most applications in RD involve multinational clinical trials in order to achieve minimum sample sizes. Where the selected instrument has not been translated into relevant languages using industry-standard methods, simultaneous cross-cultural validation is advisable as discussed in ISPOR translation and cultural adaptation task force reports (Wild et al. 2005; Wild et al., 2009). These measurement challenges and types of potential solutions are listed in Table 3.
### TABLE 3. CHALLENGES AND SUGGESTED SOLUTIONS: SELECTING/DEVELOPING THE OUTCOME MEASURE

<table>
<thead>
<tr>
<th>CHALLENGES</th>
<th>SOLUTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Search for existing COA measuring COI in COU</strong></td>
<td>Use already developed and validated measures, including item banks</td>
</tr>
<tr>
<td></td>
<td>Increase validation study sample size by including patients with related conditions</td>
</tr>
<tr>
<td></td>
<td>Include a generic measure as well as a disease-specific instrument if appropriate</td>
</tr>
<tr>
<td></td>
<td>Conduct a hybrid concept elicitation / cognitive debriefing study</td>
</tr>
<tr>
<td></td>
<td>Document consensus across multiple sources of information</td>
</tr>
<tr>
<td><strong>Begin COA development /Complete COA development</strong></td>
<td><strong>Small patient populations with few, if any, condition-specific validated outcomes measures available.</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Difficult to reach concept saturation in qualitative work due to the small number of patients available for study.</strong></td>
</tr>
<tr>
<td></td>
<td><strong>Rarity of condition increases probability of multinational clinical trial to obtain maximum sample size possible.</strong></td>
</tr>
<tr>
<td></td>
<td><strong>In RDs, the available patient population is not large enough to conduct two different studies for concept elicitation and cognitive debriefing.</strong></td>
</tr>
</tbody>
</table>
1. Challenge: Selecting an Existing Clinical Outcome Assessment that Measures the Concept of Interest in the Appropriate Context of Use

Use existing and previously validated measures where possible. Selecting an existing tool to measure the COI is a practical solution given the obstacles associated with development of a de novo COA for use in RD populations (e.g., small sample sizes, lack of understanding natural history of disease, limited time and resources to engage in a full PRO development program).

However, while existing previously validated COA instruments can be used without modifications, the evaluation process will typically indicate that modifications of existing instruments are necessary to enhance their relevance to the target population and their ability to detect treatment differences. When evaluating whether an existing COA instrument can be selected and/or modified for use in a specific clinical development program, following the methods outlined in the ISPOR task force report on the use of existing PRO instruments and their modification (Rothman et al. 2009) is recommended. The report provides very detailed recommendations for evaluating content validity when choosing an existing or modifying an existing instrument for use in a clinical development program.

A. Few or no Condition-Specific Validated Outcomes Instruments are Available

Suggested Solutions:

i. Consider using item banks to select relevant or appropriate items, these items typically have been cognitively tested, which increases confidence in patient comprehension of the items. The ability to select already-tested items that are relevant to the condition of interest to develop a disease-specific measure has great advantages. However, the resulting instrument will still require at least content validation in order to demonstrate that it is fit for purpose.

ii. Select instruments closest to the COI that can be disaggregated to include only those subscales or items relevant to the COI. Again, additional validation will be needed, as per Rothman et al (2009).

iii. If there are other RDs with clinical characteristics similar to those evaluated in the clinical trial, consider whether or not it is clinically valid to increase the instrument validation study sample size by including patients with these related conditions. For example, symptoms of Gaucher Disease are similar to those observed in patients with Pompe Disease and Niemann-Pick Disease.

iv. If appropriate and feasible, consider validating in patients with the condition of interest who are ineligible or do not wish to participate in the clinical trial for reasons unrelated to clinical characteristics. However, when choosing an instrument validated on another patient population, be aware that there is risk of using a potentially uninterpretable measure in a novel therapeutic area without the full assessment of its measurement properties. Small pilot tests with even one or two patients may be helpful in determining the measure’s fitness for purpose in such situations.

v. Consider including a well-accepted generic COA instrument, if appropriate. Although less sensitive to change in general, these measures can provide a “standard assessment” and comparative data that can be taken into account by regulatory and coverage/payment bodies (Ralluy-Callado et al. 2013; Lyons et al. 2013). This strategy can also allow for comparisons across age groups, ethnicities, cultural or regional groups, and may aid the interpretation of results by permitting comparisons with the general population.
Well-validated generic instruments are also more likely to include valid translations and may increase the probability of regulatory acceptance (especially in non-U.S. markets). However, generic instruments may be less sensitive to disease-specific parameters, and the potentially small signal size and statistical significance seen when generic instruments are used in RD populations may not be sufficient to build a compelling story for reimbursement purposes.

B. Heterogeneity Impacting Measurement Ability across Disease Spectrum

When there is substantial heterogeneity in RD symptoms within the patient population; discrete outcomes measurable across the entire spectrum of disease may not be possible (Basch and Bennett 2014).

**Suggested Solution:**

i. Consider using a multi-attribute questionnaire or a questionnaire battery that can be customized to the individual patient’s symptom profile using skip patterns. Electronic data collection is especially useful in such situations as it can minimize potential problems due to inappropriate or missed responses that pose a risk in paper-based instruments with skip patterns. Navigating study participants only to the items can also reduce respondent burden.

2. Challenge: Begin COA development /Complete COA development

A. Difficulties Developing an Instrument - Methodology

Where a careful evaluation of extant instruments reveals that no tools exist that can address the COI and planned COU, it will be necessary to develop a new instrument. As previously stated, due to the size and nature of RD patient populations, the standard methods of adapting or developing COAs may not be feasible. These methods must be modified to allow for successful development of valid and reliable endpoints in planned RD clinical trials. This is especially true when conducting studies to document the instrument’s content validity (both qualitative and quantitative) and to evaluate the measurement properties (reliability, construct validity) of the instrument.

Saturation may be difficult to achieve given the small sample size of the likely validation study sample, especially when there is heterogeneity in the RD, where there is great regional variability in patient experience, or where patients in a variety of disease progression states are included in the study. The likelihood that the clinical trials will be conducted in multiple countries further complicates the process. As stated previously, instruments will thus need to be evaluated for cultural equivalence and validity within each country or culture from which participants are enrolled in the clinical trial(s) (Wild et al. 2009).

**Suggested Solutions:**

i. Because the standard validation methods required may not be feasible in small RD populations, there have been some alternative strategies proposed to achieve cross-cultural equivalence in such situations. For example, it may be possible to carry out a single forward translation by a bilingual clinical expert in each target country, followed by a multi-disciplinary expert committee review that compares back-translated versions.

ii. A cognitive interview study with even less than 10 patients, instead of the more typical 30 to 50 patients, is another option as this may be sufficient to reach saturation (Price et al. 2009). In as far as possible, investigators should try to ensure that patients representing the variability in patient disease characteristics (symptoms, severity, etc.) are included in the validation sample.
B. Difficulties Adapting an Instrument - Qualitative

When adapting tools from similar disease states, strict methodology would require two separate patient populations; one for concept elicitation and one for cognitive debriefing. In the rare disease field, the population typically may not be large enough to conduct two different analyses.

Suggested Solution:

i. Consider conducting a hybrid concept elicitation/cognitive debriefing study to document concept validity for a novel patient population (prior to implementing study, work with regulators to assure there is acceptance of this approach). Such studies combine a brief concept elicitation interview with a cognitive debriefing to confirm that the main concepts and domains in the selected measure are relevant to the COU and appropriate for the new population, and that no additional concepts are needed to assess outcomes in these patients. This solution may allow for a condensed timeframe for evaluation of content validity in addition to obviating the need for multiple interviews with separate patient populations.

B. Difficulties Adapting an Instrument - Concept Saturation

Due to the variation in patient type and experience in addition to the small sample size necessitated by the rarity of the condition, concept saturation may be difficult to achieve.

Suggested Solutions:

i. Identify main themes based upon a variety of sources including not just CE interviews but case reports, other published literature, clinical expert opinion.

ii. Use only very broad codes to represent general domains.

iii. Conduct CE studies using not just patients but also caregivers, clinicians, even other family members or teachers – anyone who has an intimate knowledge of the patient.

iv. Document consensus across multiple sources of information including the literature, interviews with clinical experts or caregivers, as well as patient interviews.

C. Validation Study Complications due to Multinational Study Designs

The rarity of the condition increases the likelihood that a multinational clinical trial will be needed to obtain the largest sample size possible. (It is almost impossible to attain a sufficient sample size in most RD validation studies to reach even the minimum sample sizes usually required for psychometric validation.) Conducting validation studies is more difficult under these circumstances.

Suggested Solutions:

i. When patients travel to central locations for their treatment, consider conducting test-retest reliability studies while they are on-site (if there is sufficient time to allow for retest during the visit) or you may have to consider conducting at home even if it mixes modes of data collection.

ii. Conduct cultural equivalence studies on a small number of patients/caregivers/clinicians in each country where the clinical trial will be implemented to ensure that the meaning of all items, instructions and response choices is valid across various national groups.

iii. Cultural norms will need to be examined as part of establishing content validity within each country where the COA is expected to be used. The appropriateness and relevance of a COA instrument will
need to be re-examined with separate content-validity assessment within each new culture. This assessment should focus on all relevant aspects of the instrument including the instructions, items, concepts, vocabulary, and pictorial representations. Instruments with more than one version (e.g., for different age groups or for self-report as well as observer-report) must be evaluated for cultural equivalence and validity in each country of interest.

iv. Where the clinical trial includes pediatric patients, given the possible regional variations in cultural expectations for how children behave and develop, it will be necessary to re-assess an instrument’s age-appropriateness in each country where it will be administered (Matza et al. 2013).

v. Where instrument translation is required, if possible use a single forward translation by a bilingual clinical expert in each target country, followed by a multi-disciplinary expert committee review that compares back-translated versions.

D. Difficulty Quantitatively Evaluating Psychometric Measurement Properties

Standard statistical tests for evaluating psychometric properties of measures may be underpowered due to the small sample size as well as the subtypes of the condition. For example, the heterogeneity of many RDs can confound the association between age and functional status if patient subgroups have not been completely defined.

Suggested Solutions:

i. For small sample sizes, it may be advisable to use non-parametric statistics in evaluating measurement properties. Given the sometimes skewed distributions of items in a COA instrument in RD, using nonparametric statistics is necessary. These statistics will include the typical parameters of mean and variance but, unlike parametric statistics, nonparametric statistics make no assumptions about the probability distributions of the variables being assessed. (Wasserman, 2005)

ii. Maximize chances of achieving statistically significant findings where possible by using continuous variables.

iii. Conduct sensitivity analysis to provide more information about the stability of results in different patient groups or clinical scenarios.

iv. If possible, match or stratify on important prognostic factors to increase precision of the estimates.

v. If appropriate, increase validation study sample size by including patients who have other RDs similar to the one of interest or whose clinical characteristics are similar to those evaluated in the clinical trial, or include patients with the condition of interest who do not wish to participate in the clinical trial.

E. Difficulty in Conducting Test-Retest Reliability

Estimating appropriate time intervals between tests may be difficult due to external factors such as a long duration between study visits to distant clinics or variations in availability of treatments for acute symptom exacerbations.

Suggested Solutions:

i. Recruit patients for instrument validation studies through major treatment centers (e.g., locations patients travel to receive routine care).
ii. Time assessments to patient clinic visits or consider using electronic, or telephonic data collection methods (e.g., Computer-Assisted Telephone Interviews, or CATI) that allow respondents to complete instruments offsite. Please keep in mind that it will be necessary to confirm that patients are clinically stable in order to ensure retest results are valid. It should also be noted that there may not be sufficient sample size for RD CATI validation studies.

iii. Conduct the psychometric validation study within the clinical trials. Consider conducting longitudinal validation within a Phase II or III trial. This should be discussed with regulatory authorities prior to implementing the study to ensure acceptance of instrument validation results.

iv. Evaluate test-retest reliability with non-parametric statistics where the sample size is small.

F. Difficulty Assessing Inter-Rater Reliability (IRR)

Dispersion of patients over a wide geographic area can make it difficult to obtain simultaneous observations of more than one caregiver, non-clinical observer, or clinician. Moreover, if cross-national populations are included, differences in clinical protocols for diagnosis and treatment and/or cultural differences in the interpretation of symptoms may result in between-country variations.

Suggested Solutions:

i. Consider constructing standardized clinical scenarios (e.g., written descriptions, pictures, online video clips) for observers to complete remotely; then, evaluate agreement based upon their instrument responses to these scenarios.

ii. Use results from evaluations of patient/caregiver/clinician perceptions to inform interpretation of any regional or demographic differences.

iii. If patient numbers allow, stratify IRR analysis by country or region and report within-group IRR as well as total IRR results.

G. Obstacles to Longitudinal Psychometric Analysis

Stable disease criteria can be difficult to identify in some RDs due to the lack of a well-defined natural history and an incomplete understanding of rate of progression of the condition making longitudinal psychometric analyses challenging.

Suggested Solutions:

i. Use results from evaluations recommended in Section II A. and B. to construct criteria for stable (as possible) disease and confirm these criteria with clinical experts.

ii. Collect event data at the time of COA measurement (date of latest acute event, frequency and severity of acute events, etc.) to obtain a better understanding of any changes in health status that may impact results.

CONCLUSION:

There is a growing recognition of the importance of the patient voice in understanding treatment benefit to improve clinical and medical care coverage decisions. Patients and their caregivers are increasingly demanding that these factors be taken into account. Nonetheless, obtaining the patient perspective has
been rarely attempted in RD clinical trials (Basch and Bennett 2014). While perhaps not the sole reason, certainly the lack of valid measures and the difficulties in developing or adapting measures for this purpose is a major contributor to this gap in our understanding of RD treatment efficacy.

This ISPOR Task Force report is an initial attempt to delineate the challenges encountered when measuring COAs in RD patients with potential solutions to these challenges. These emerging good practices can serve as a starting point for the development of an inventory of sound, pragmatic and creative solutions that can lead to an increase in the number of RD clinical trials with COA endpoints. Follow-on work will be needed to learn how best to interpret results given the lack of understanding of natural disease history for many RDs, as well as the small and often heterogeneous patient populations.

Other challenges to be addressed as experience increases will include working with regulatory agencies to achieve a clear set of standards that are practical for evaluating the validity of RD COA endpoints and additional work to arrive at feasible statistical tests to demonstrate measure reliability and psychometric properties in very small populations.

Furthermore, the payer perspective on the types and uses of PRO information for supporting market access decisions is not addressed in this paper. This will be important to understand given the cost and organizational budget impact of many new RD treatments and the growth in number of treatment options for RD patients.

In summary, the explosion of new treatments for rare disorders is a great opportunity for researchers, clinicians and, of course, patients, to finally have some viable options that can extend life and decrease suffering. However, the options can be bewildering to choose from without good evidence as to their actual benefit to the patient. It is crucial to provide the RD community with valid, useful information about the actual treatment experience from a real-world perspective. Despite the many conceptual and methodological challenges that remain to be solved, COAs will be an important component of decision-making moving forward. Addressing these obstacles will be both exciting and necessary for improved patient outcomes.

APPENDICES follow REFERENCES
REFERENCES


Benjamin KL. Rare Disease Research in the Modern PRO World. Webinar presented February 14, 2013.


European Organisation for Rare Diseases (EURORDIS). The Voice of 12,000 Patients: Experiences and Expectations of Rare Disease Patients on Diagnosis and Care in Europe. 2009 http://www.eurordis.org/IMG/pdf/voice_12000_patients/EURORDISCARe_FullBookr.pdf


Giraldo et al. Orphanet Journal of Rare Diseases 2012, 7:17 http://www.ojrd.com/content/7/1/17

Goodman JL. FDA's Efforts on Rare and Neglected Diseases. Statement to the Subcommittee on Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Committee on Appropriations, United States Senate June 23, 2010.


Hamburg B. Report to Congress: Improving the Prevention, Diagnosis, and Treatment of Rare and Neglected Diseases In Response to Agriculture, Rural Development, Food and Drug Administration, and Related Agencies Appropriations Act, 2010, Public Law 111-80, Section 740; March 2011.


Kole A and Faurisson F. The Voice of 12,000 Patients. Experiences and Expectation of Rare Disease Patients on Diagnosis and Care in Europe. EURORDIS; 2009.


Lesko LJ. Introduction: Rare Diseases, Orphan Drugs. Presented at the Advisory Committee for Pharmaceutical Sciences and Clinical Pharmacology March 2, 2011 Dallas, Texas.


http://www.ispor.org/TaskForces/PROChildrenAdolescents.asp


Papadopoulos E. Clinical Outcome Assessments/PROs. Presentation at the US Conference on Rare Diseases and Orphan Conditions Washington DC October 24, 2012


Ralluy-Callado M, et al. The impact of Hunter syndrome (mucopolysaccharidosis type II) on health-related quality of life. Orphanet J Rare Dis 2013;8:120

Ratcliffe, Nestler-Parr, Babela, Khan, Tesoro, Molsen, Hughes. Rare Disease Terminology and Definitions - A Systematic Global Review: Report of the ISPOR Rare Disease Special Interest Group. Value in Health (2015); 18:906-914


Shire. Rare Disease Impact Report: Insights from patients and the medical community. April 2013.


Appendix 1.

Roadmap to **PATIENT-FOCUSED OUTCOME MEASUREMENT** in Clinical Trials

1. Understanding the Disease or Condition
   - A. Natural history of the disease or condition
     - Onset/Duration/Resolution
     - Diagnosis
     - Pathophysiology
     - Range of manifestations
   - B. Patient subpopulations
     - By severity
     - By onset
     - By comorbidities
     - By phenotype
   - C. Health care environment
     - Treatment alternatives
     - Clinical care standards
     - Health care system perspective
   - D. Patient/caregiver perspectives
     - Definition of treatment benefit
     - Benefit-risk tradeoffs
     - Impact of disease

2. Conceptualizing Treatment Benefit
   - A. Identify concept(s) of interest (COI) for meaningful treatment benefit, i.e., How a patient:
     - Survives
     - Feels (e.g., symptoms)
     - Functions
   - B. Define context of use (COU) for clinical trial:
     - Disease/Condition entry criteria
     - Clinical trial design
     - Endpoint positioning
   - C. Select clinical outcome assessment (COA) type:
     - Patient-Reported Outcome (PRO)
     - Observer-Reported Outcome (ObsRO)
     - Clinician-Reported Outcome (ClinRO)
     - Performance Outcome (motor, sensory, cognition)

3. Selecting/Developing the Outcome Measure
   - A. Search for existing COA measuring COI in COU:
     - Measure exists
     - Measure exists but needs to be modified
     - No measure exists
     - Measure under development
   - B. Begin COA development:
     - Document content validity (qualitative or mixed methods research)
     - Evaluate cross-sectional measurement properties (reliability and construct validity)
     - Create user manual
     - Consider submitting to FDA for COA qualification as exploratory endpoint
   - C. Complete COA development:
     - Document longitudinal measurement properties (construct validity, ability to detect change)
     - Document guidelines for interpretation of treatment benefit and relationship to claim
     - Update user manual
     - Submit to FDA for COA qualification as effectiveness endpoint to support claims

---

## Appendix 2

### Example of RD Heterogeneity: Amyloidosis

<table>
<thead>
<tr>
<th></th>
<th>AL (Primary) Amyloidosis</th>
<th>AA (secondary) Amyloidosis</th>
<th>AF (Familial) Amyloidosis</th>
<th>Localized Amyloidosis</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Etiology</strong></td>
<td>Unknown; some forms co-occur with multiple myeloma</td>
<td>Chronic inflammatory conditions (e.g., arthritis, lupus, IBD)</td>
<td>Hereditary; caused by an abnormal amyloid transthyretin (TTR) protein</td>
<td>Local production and deposition of an amyloidogenic protein; often co-occurs with age-related conditions especially diabetes, Alzheimer’s</td>
</tr>
<tr>
<td><strong>Organ Involvement</strong></td>
<td>Multiple; Kidney, Heart, Liver, Intestines, Neurological system, Lungs</td>
<td>Multiple; Spleen, Kidneys, Liver, Adrenal glands, lymph nodes</td>
<td>Multiple; Liver, heart, Neurological system (depends on subtype)</td>
<td>Single organ; CNS, skin, bladder, upper and lower respiratory system, GI</td>
</tr>
<tr>
<td><strong>Average Age of Onset, Gender</strong></td>
<td>65</td>
<td>Depends on onset of underlying disorder; can be as early as 15 years for juvenile arthritis. More common in males</td>
<td>Varies; Portugal and Japanese subtypes = 33; Swedish subtype = 56; other regions = 60+</td>
<td>Varies; middle age most common; More common in males</td>
</tr>
<tr>
<td><strong>Ethnicity/Region</strong></td>
<td>N/A</td>
<td>N/A</td>
<td>Varies: TTR-FAP subtype includes Portugal, Japan, Sweden, African Americans</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Treatment</strong></td>
<td>Chemotherapy, SCT</td>
<td>Treatment of underlying disorder, TNF inhibitors, dialysis</td>
<td>Liver transplant, chemotherapy, some medications</td>
<td>Organ-specific symptom palliation, surgery where possible</td>
</tr>
<tr>
<td><strong>Progression</strong></td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
<td>N/A</td>
</tr>
<tr>
<td><strong>Prognosis</strong></td>
<td>Depends on time to diagnosis and cardiac involvement; median survival if diagnosed early and no cardiac involvement = 5 years</td>
<td>Depends on control of underlying condition and degree of renal involvement; average with renal involvement = 2 – 5 years</td>
<td>Depends on regional subtype; 10 – 20 years</td>
<td>Depends on type of organ involvement</td>
</tr>
<tr>
<td><strong>Presenting Symptoms</strong></td>
<td>Fatigue, Weight loss, Paresthesias, Hoarseness, Edema, Carpal tunnel syndrome, Macroglossia, Macrococutaneous lesions, Hepatomegaly, Edema</td>
<td>Most common symptoms: Nephrotic syndrome, Weakness, Weight loss, Peripheral edema</td>
<td>Peripheral neuropathy, Autonomic neuropathy, Cardiomyopathy, Nephropathy, Ocular deposition, Anhidrosis, Sexual impotence, Diarrhea, Constipation, Nausea, Vomiting, Orthostatic hypotension, Neurogenic bladder, Anemia</td>
<td>Depends on type of organ involvement</td>
</tr>
</tbody>
</table>
Appendix 3

EXAMPLE

TWO APPROACHES TO DEVELOPING A MEASURE OF HRQL FOR PEDIATRIC HAEMOPHILIA – A PATIENTS

Condition: Haemophilia - A is a congenital disorder caused by a deficiency in plasma coagulation Factor VIII. It is a rare disease affecting males almost exclusively, affecting approximately 1 in every 5,000 live male births worldwide. About 40% of all patients have the severe form of the disorder (<1% Factor VIII). These patients experience spontaneous bleeding unrelated to trauma, usually in their joints, muscles and soft tissue. The bleeds can result in short-term pain and disability, and over time can cause permanent joint damage with significant impairment of physical function. Due to the unpredictable nature of bleeding episodes there may also be significant psychological and social impairment.

Treatment: There is no available cure for haemophilia. Treatment consists of replacing Factor VIII to induce coagulation. There are three strategies for Factor VIII replacement. The oldest method is to stop a bleed by providing Factor VIII “on demand,” at the time of a bleed. Two prophylactic strategies are also available; a primary prevention method whereby patients receive Factor VIII on a regular basis 2-3 times weekly before they experience any bleeds, and a secondary prevention method whereby patients receive Factor VIII regularly to prevent future episodes after they have experienced bleeds. Factor VIII must be administered by infusion, either in a clinic or a home setting.

Main treatment and condition-related concerns: While prophylactic Factor VIII replacement has been shown to prevent bleeds and consequent joint damage and complications, resulting in significantly less impairment, these strategies are expensive and inconvenient. Access is thus restricted in some countries or regions due to the expense. There is also some controversy regarding the effectiveness of secondary prophylaxis in terms of reducing or preventing joint damage and disability in patients who already have some impairment. The demands and inconvenience of prophylactic regimens have been demonstrated to lead to nonadherence, especially in adolescent and young adult populations. There is also some evidence that the procedure itself can result in emotional and social problems, especially in young children. Thus, there is a need for valid disease-specific measures of patient function and HRQL to understand the impact of haemophilia –A and its treatment on patients’ lifestyle, self-concept, relationships, plans and activities. In children, there is a need to understand the impact of haemophilia A on developmental tasks.

COA Instrument development: Two PRO measures aimed at pediatric patients were developed in the early 2000’s and have been validated for a variety of contexts of use. The Canadian Haemophilia Outcomes – Kids Life Assessment (CHO-KLAT) was developed for use in Canadian patient populations to ensure that the multiple Canadian cultures and Canada-specific treatment patterns were adequately accounted for (Young et al. 2004). The HaemoQOL was developed in Europe using patient populations in six different countries to ensure validity across multiple national groups (von Mackensen et al., 2004). While the objectives of these measures were the same (to develop a tool that could be used to measure patient subjective experiences for assessment of clinically important changes in patient functioning and HRQL for clinical, research and reimbursement decision-making purposes), the methods of instrument development were different (see Table). However, both instruments were designed taking into account several limitations of studying patients with rare diseases.
The small population size necessitated developing a measure valid for use in multiple age groups, including the need for a parent proxy measure to be administered if necessary, especially in very young children. The instruments have been translated and validated for use in multiple countries. Despite their similarities, these measures were developed using very different methods. The CHO-KLAT was developed using patient (and parent) input in addition to clinical and literature sources, with the greatest weight given to the child’s perspective.

The HaemoQOL, while psychometrically tested on patient samples, was not developed with patient input. The CHO-KLAT comprises a single self-report instrument that has been validated for all age groups (ages 4 – 18) to facilitate comparisons across age groups and to ensure that there were no systematic differences in item content. The developers included a “Not applicable” response choice for all questions to allow for developmental differences. This flexibility also has the advantage of allowing for variations in treatment patterns. The Haemo-QOL was developed with different versions for different age groups, including an interviewer-administered version for very young children. These versions include different numbers of items and different content, with adolescent patients completing questions addressing two additional domains to capture their views of their future and possibility of relationships such as having a romantic partner. This instrument was later adapted and validated for use in adult populations (Rentz et al, 2008).

Subsequent research (Bradley et al. 2006) confirmed the adequacy of the psychometric properties of both instruments. However, the CHO-KLAT was found to have better concordance between children and parent versions, and also demonstrated significant differences between outcomes captured in generic pediatric QOL instruments (the PedsQL and a VAS rating scale) and CHO-KLAT outcomes, indicating that this measure is collecting disease-specific information. The Haemo-QOL outcomes were significantly correlated with the generic measure results and the correlations between the child and parent measures were not significant. However, the Eurocentric focus of the Haemo-QOL may make it a more appropriate measure of outcomes in European patient populations or where both adult and pediatric patients are included in the study population.