



Disclaimer:

Presentation slides from the Rare Disease Workshop Series are posted by the Kakkis EveryLife Foundation, for educational purposes only. They are for use by drug development professionals and statisticians, and are not to be used to guide the prescribing or use of any of the drugs mentioned in the slides. To obtain information on a particular drug, refer to the drug labeling. Do not reproduce or distribute the slides (full set or any portion of) without the permission of the author.



RARE DISEASE WORKSHOP SERIES
Improving the *Clinical Development Process*

Workshop #2

Clinical Evaluation of Rare Disease Treatments

June 14-15, 2011

Summary of Workshop #2

Emil D. Kakkis, M.D., Ph.D.
President, Kakkis EveryLife Foundation



The Rare Disease Workshop Series: Why?

- Rare disease treatment development provides difficult challenges
- The typical approaches to clinical development may not work well
- Can we define better ways of approaching the development of treatment information
- *We can do better than we are doing now*



Workshop Series Topics

- Workshop #1 Statistical analyses of rare disease studies
- Workshop #2 Clinical evaluation of rare disease treatments
- Workshop #3 Surrogate endpoints & accelerated approval
- Workshop #4 *Perfect* policy solutions: all things for all people



Workshop #2: Topic Areas for the Workshop

Morning Day 1

- Selecting an endpoint

Morning Day 1

- Establishing an MID/responder definition

Afternoon Day 1

- Developing a PRO for rare diseases

Morning Day 2

- Clinical evaluation tools for rare diseases

Afternoon Day 2

- Multi-domain assessment of rare diseases



RARE DISEASE WORKSHOP SERIES
Improving the *Clinical Development Process*

Selecting an Endpoint for Rare Disease Studies

Morning Day 1



Workshop #2 Agenda: Day 1

Selecting an endpoint

Day 1 – Morning

8:00 - 8:30 am Registration and Breakfast

Clinical evaluation of rare disease: lessons learned in common diseases relevant to rare diseases

8:30 am Welcome

Emil Kakkis, M.D., Ph.D., President, Kakkis EveryLife Foundation

8:40 am Introduction to the major questions and issues for evaluation of disease

Tom Fleming, Ph.D., Prof. of Biostatistics and Statistics, University of Washington

9:00 am Clinical endpoint selection and interpretation

John Powers, III, M.D., Senior Medical Scientist, NIH

9:20 am Selection of meaningful endpoints

Anne Pariser, M.D., Associate Director for Rare Diseases, CDER, FDA



Summary of Speakers: Tom Fleming

- Provided general issues around choices
 - Measurable, sensitive, clinically relevant
- Clinical endpoint is one that measures how patient feels, functions or survives
- Cautions on biomarkers and fallibility
- Emphasizes the role of clinical research
 - Not just getting a p value alone

“To obtain a statistically reliable evaluation regarding whether the experimental intervention is safe and provides clinically meaningful benefit.”

** Fleming, Statistics in Medicine, 2006*



Summary of Speakers: John Powers

- Review regulatory basis of approval
- Endpoints and measurement
 - Content validity: the accurate measurement of something important to how the patient feels, functions or survives
 - Construct validity: the methodology of measurement is both reliable and responsive, precise and accurate. Reliable but inaccurate may be reliably wrong.
 - Must limit bias through randomizing/blinding
- Risks of clinician evaluation



Summary of Speakers: Anne Pariser

- Review Approval Pathways
 - Standard and Accelerated approval
 - Standards for adequate and well controlled trials
- Endpoint choices
 - Prominent and important effect of the disease but not necessarily the worst
 - Need information on natural history to judge and choose optimally
 - Applying endpoints previously used is often done and may be easier
 - Must be relevant to the disease state



Workshop #2 Agenda

Selecting endpoints: How to do it?

9:40 am Discussion:

- What is the **conceptual framework** for understanding the disease as a whole, then selecting the most worthwhile and feasible of these endpoints to be measured in a clinical trial?
- When should a **previously used endpoint** versus an **untested endpoint** be used?
- How do we go about **qualifying a new endpoint** that has not been tested before?

10:25 Midmorning Break



Endpoint choice in rare diseases

- More effort is needed in evaluating natural history to obtain better information to guide choice of endpoints
- Clinical endpoints previously used in other programs have been more commonly used in rare diseases
- Challenges still exist in using untested endpoints or measures that may be valid for a rare disease but for which there is little regulatory experience or natural history information to support
- Qualifying an new endpoint could require “exhaustive discussion”



Key findings relevant to rare diseases

- Endpoint choice
 - Know the disease well for content validity
 - Natural history, survey studies
 - Establish a framework for the diseases
 - Provide possible endpoint plan
 - Talk to FDA frequently
- Choosing known endpoints is easier than new endpoints
- New endpoints require substantial support



RARE DISEASE WORKSHOP SERIES
Improving the Clinical Development Process

Minimally Important Differences and Responder Criteria

Morning Day 1



Workshop #2 Agenda: Day 1

Assessing MID/Responder Definition

Assessing minimum important difference/responder definition in clinical evaluation of rare diseases

- Strategies for developing and using MID in clinical efficacy evaluation
 - 10:40 am **Gordon Guyatt, M.D.**, Prof. of Clinical Epidemiology and Biostatistics, McMaster University
 - 11:10 am Discussion
 - 11:20 am **John Powers, III, M.D.**, Senior Medical Scientist, NIH
 - 11:50 am **Marc Walton, M.D., Ph.D.**, Associate Director, Office of Translational Sciences, OND, CDER, FDA
- Assessing the responder definition in rare disease situations
 - 12:20 pm **Gerry Cox, M.D., Ph.D.**, Vice President of Clinical Research, Genzyme Corporation
 - 12:40 pm **Kathy Wyrwich, Ph.D.**, Senior Research Scientist, Center for Health Outcomes Research, United BioSource Corporation
 - 1:00 pm Lunch



Minimally important differences: Guyatt

- Introduced concept of minimally important difference (MID)
 - The minimal difference which a patient would consider important
 - Not just a minimal detectable difference
- Evaluation of clinical data
 - MID on the mean difference: How big is the change, using the MID as ruler
 - MID on each patient, “responder analysis”: how many patients had the important difference
 - Both are important and provide different information
- Anchoring assessment using the global ratings of change: 15 point scale



Minimally important differences: Powers

- Reviewed definitions
- Relative value of doing mean versus responder analyses
- Must focus on individual endpoints and not composites for clarity
- Anchor MID may require “triangulation” with multiple measures
- Can anchor with global ratings of change
 - Why not use as an endpoint: Not as well defined as the outcome measures
- Need to use early phase studies to develop endpoints, measures and analyses



Marc Walton

Office of Translational Sciences FDA

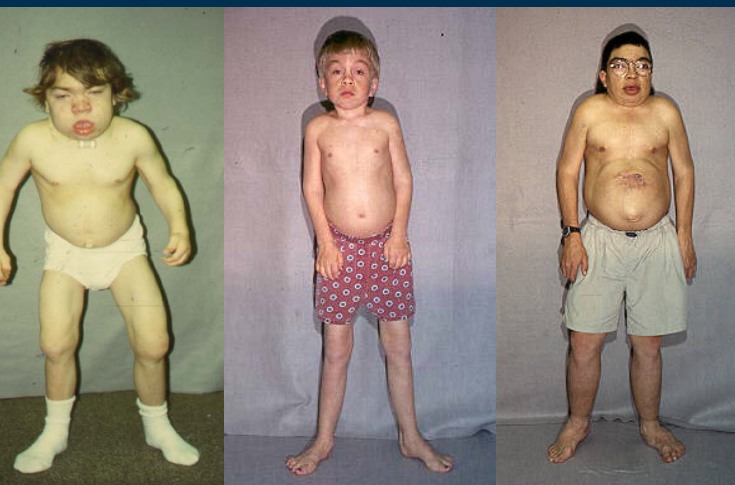
- Reviewed the fundamentals of choosing a Clinical Outcome Assessment (COA)
- Choosing existing versus new COA's
- MID is one context to put for a COA result
 - Larger effects may be important
 - MID not important if treatment effect is clear without knowing it
- The MID might not be incorporated into the primary endpoint but may provide better interpretability of the result
- Risks may exist if program has positive statistical result but unclear clinical interpretability



RARE DISEASE WORKSHOP SERIES

Improving the Clinical Development Process

Gerry Cox discussed the application of MID's to the analysis of Phase 3 Aldurazyme data using a multidomain assessment of outcome rather than a single domain outcome



Assessing Responders for Aldurazyme in MPS I

Gerry Cox, MD, PhD

Vice President, Clinical Development
Genzyme Corporation



Phase 3 Study Design

- 45 patients, multi-center, multi-national, randomized, double-blind, placebo-controlled, 26-week study followed by 3.5 year extension
- **Co-Primary Endpoints:** Changes in Percent Predicted FVC (Absolute), 6-Minute Walk Test (meters)
- **Secondary Endpoints:** Changes in Liver Volume (%), AHI (Events/Hr), Shoulder Flexion (Degrees), uGAG (%)
- **Tertiary Endpoints:** Changes in Percent Predicted FEV₁, TLC, Shoulder Extension, Knee Flexion and Extension, Growth, Acuity, ADL, QoL



Composite Endpoint Approach

- Accommodates patient heterogeneity
- More comprehensive view of patient response
- More sensitive than single organ endpoint
- Define domains with thresholds of clinically significant change (+1 improve, 0 unchanged, -1 decline)
- Endpoints
 - Responders
 - Proportion of patients with net improvement
 - Net Change
 - Improvements minus declines per patient



Phase 3 Study: Composite Endpoint

Domains

Clinically Significant Thresholds

-
- FVC $\pm 11\%$
 - 6MWT ± 54 meters
 - Apnea-Hypopnea Index ± 10 events/hour
 - Shoulder Flexion ± 20 degrees
 - Visual Acuity ± 2 lines on eye chart



Phase 3 Study: Composite Endpoint

Placebo

Patient	FVC	6MWT	SHFLEX	AHI	ACUITY
	11%	54m	20 deg	10 ev/hr	2-lines
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					
20					
21					
22					
23					

Aldurazyme

Patient	FVC	6MWT	SHFLEX	AHI	ACUITY
	11%	54m	20 deg	10 ev/hr	2-lines
24					
25					
26					
27					
28					
29					
30					
31					
32					
33					
34					
35					
36					
37					
38					
39					
40					
41					
42					
43					
44					
45					

Clinically Significant Changes



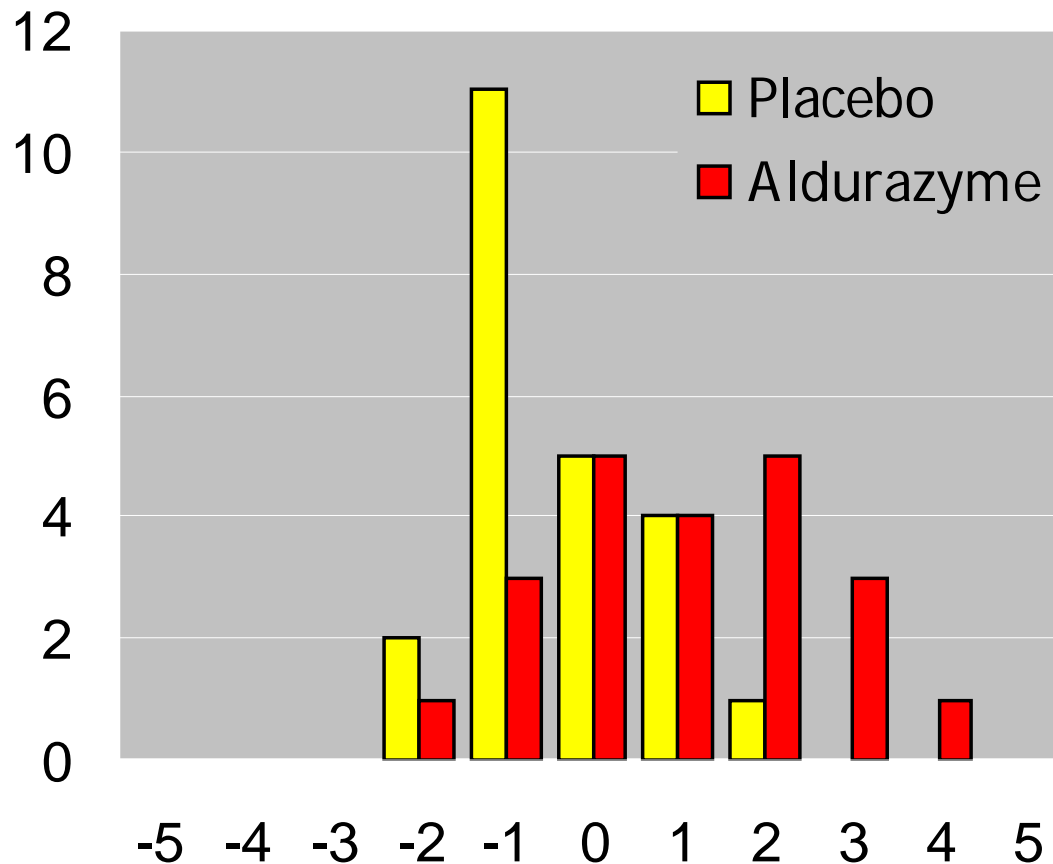
Improve
No Change



Decline
Not Available



Phase 3 Study: Responders and Net Change



Responders

59% Aldurazyme
22% Placebo
($p=0.016$)

Mean Net Change

1.0 Aldurazyme
-0.4 Placebo
($p=0.003$)



Summary

- Adopting MCD/MCID's from other disease can be useful strategy for rare diseases, but ideally each disease should have its own MCID/MID's determined
- Responder index, along with natural history data and precedents for change in other diseases, helps to provide clinical meaningfulness
- Composite endpoints accommodate patient heterogeneity and provide a measure of "overall benefit" and "responder" for clinical treatment trials of multisystem diseases, especially those that are rare and clinically heterogeneous
- Need buy-in from regulatory agencies and validation by experts
 - Prospective agreement upon domains, weighting, clinically significant change thresholds, scoring system, definition of "responder", and statistical analysis



Interpretation of Health Outcomes Data Within Rare Disease Clinical Trials

Kathy Wyrwich, PhD
United BioSource Corporation

Rare Disease Workshop 2
June 14, 2011



Guidance for Industry Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims

Additional copies are available from:

*Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 51, rm 2201
Silver Spring, MD 20993-0002*

*Tel: 301-796-3400; Fax: 301-847-3714; E-mail: druginfo@fda.hhs.gov
<http://www.fda.gov/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/default.htm>*

or

*Office of Communication, Outreach, and Development, HFM-40
Center for Biologics Evaluation and Research
Food and Drug Administration*

*1401 Rockville Pike, Suite 200N, Rockville, MD 20852-1448
Tel: 800-835-4709 or 301-827-1800; E-mail: ocod@fda.hhs.gov
<http://www.fda.gov/BiologicsBloodVaccines/GuidanceComplianceRegulatoryInformation/default.htm>*

or

*Office of Communication, Education, and Radiation Programs
Division of Small Manufacturers, International, and Consumer Assistance, HFZ-220
Center for Devices and Radiological Health
Food and Drug Administration*

*1350 Piccard Drive, Rockville, MD 20850-4307
DSMICA E-mail: dsmica@cdrh.fda.gov
DSMICA Fax: 301-443-8818
(Tel) Manufacturers Assistance: 800-638-2041 or 301-443-6597
(Tel) International Staff: 301-827-3993
<http://www.fda.gov/MedicalDevices/DeviceRegulationandGuidance/GuidanceDocuments/default.htm>*

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)**

**December 2009
Clinical/Medical**



Interpretation is More Than Achieving Statistical Significance

- Need to achieve statistically significant differences between the active treatment and placebo arms for clinical trials, but it's just not enough (but *is* the way to properly power most trials with a continuous endpoint measure!)
- Need a way to interpret if statistically significant differences are meaningful and important to clinical trial participants
- Can't rely on the statistical significance to demonstrate an interpretable difference
 - Many PRO scales and other health outcomes are new to label readers and familiarity with what types of changes are important requires experience over time



Anchor-Based Methods

to establish the MID for a measure using a global score to calibrate the endpoint change

- The related response between targeted PRO instrument and the closely-related concept can provide meaning or interpretation of change in a PRO measure
- Anchor selection should have intuitive meaning



Example Types of Anchors

- Clinical measure
 - *a 50% reduction in incontinence episodes* might be proposed as the anchor for defining a responder
- Patient global ratings



HAE

- Hereditary angioedema (HAE) is a rare autosomal dominant disorder affecting approximately 1 in 10,000 to 50,000 persons worldwide
- Symptoms of HAE are characterized by recurrent, highly variable, acute episodes (attacks) of swelling of any part of the body, including the face, larynx, gastrointestinal tract, extremities, trunk, and/or genitals.



2 New HAE PROs Developed for Clinical Trials

- Treatment Outcome Score (TOS)
- Mean Symptom Complex Severity (MSCS)
- Secondary analysis of DX-88/14 EDEMA3[®] clinical trial data pooled across treatment and placebo groups
- N = 71 with baseline and 4 hour responses



Anchor-Based Results at 4 Hours

	Significant Improvement (<i>N</i> = 30)	Improvement (<i>N</i> = 21)	Same (<i>N</i> = 18)	Worsening (<i>N</i> = 1)	Significant worsening (<i>N</i> = 1)
TOS at 4 h	93.7 (3.5)	44.7 (4.2)	3.6 (4.5)	-35.9	-97.2
MSCS at 4 h	-1.4 (0.1)	-0.9 (0.1)	-0.3 (0.1)	-0.8	0.3



Minimum Important Difference (MID)

- Defined in *Draft PRO Guidance (2006)* as:
 - The smallest difference between clinical trial treatment arm mean change from baseline (point estimates) that will be interpreted as important
- MID represents the between groups criterion that needs to be met or exceeded in order for study results to be considered clinically meaningful



MID

- Why is MID not included in Final PRO Guidance?
 - Term is interpreted inconsistently (intra-patient change vs. inter-group difference of mean change from baseline)
 - Point estimates of the difference in means between two groups may mask important changes for individual patients or types of patients in each group
 - Responder definitions offer a direct approach to intra-patient change and treatment differences across a range of clinical anchors that can be presented in a cumulative distribution function



Cumulative Distribution Function

- An alternative or supplement to responder analysis
- Mentioned prominently in the 2009 FDA Guidance on PRO label and promotional claims

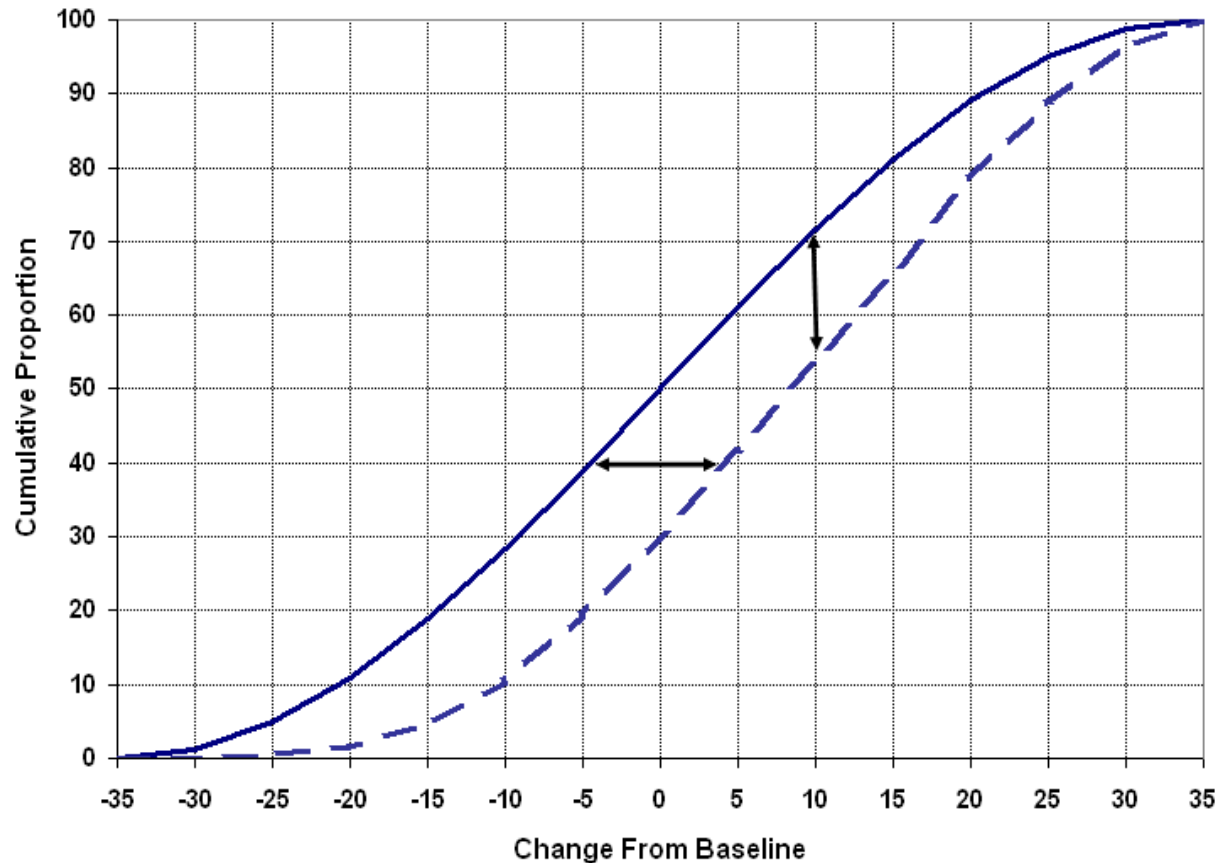


Cumulative Distribution Function

- Display a continuous plot of the percent change (or numeric change) from baseline on the horizontal axis and the cumulative percent of patients experiencing up to that change on the vertical axis
- Such a cumulative distribution of response curve – one for each treatment group – would allow a variety of response thresholds to be examined simultaneously and collectively, encompassing all available data



Illustrative Cumulative Distribution Function: Experimental Treatment (solid line) better than Control Treatment (dash line) -- Negative changes indicate improvement





Aricept® Label from 10/13/2006

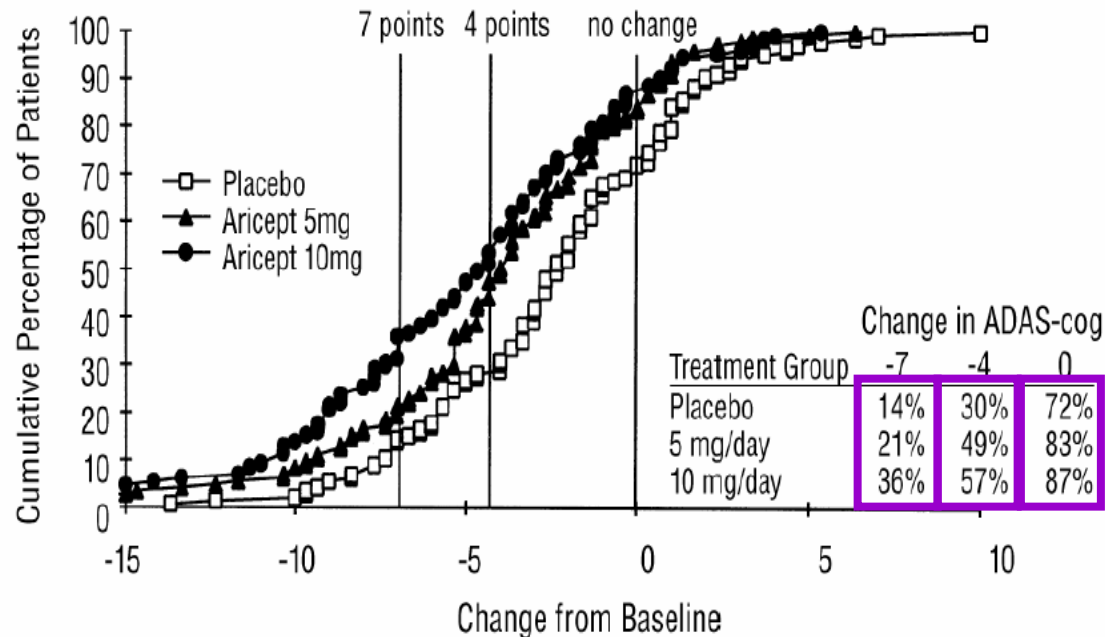


Figure 5. Cumulative Percentage of Patients with Specified Changes from Baseline ADAS-cog Scores. The Percentages of Randomized Patients Within Each Treatment Group Who Completed the Study Were: Placebo 93%, 5 mg/day 90% and 10 mg/day 82%.



Workshop #2 Agenda

MID/Responder Definition: How to do it?

1:20 pm Discussion:

How should MID/responder definition be determined in rare disease situations: translation from other diseases or empirical testing in natural history or clinical data?

How should MID/responder definition be analyzed: As a method to interpret clinical meaningfulness of data after obtaining a positive statistical result, or built into the primary analysis?



RARE DISEASE WORKSHOP SERIES
Improving the Clinical Development Process

Patient Reported Outcomes in Rare Disease Studies

Afternoon Day 1



Workshop #2 Agenda: Day 1

Patient Reported Outcomes

Day 1 – Afternoon

- Exploring patient reported outcomes to support a label claim or to support the results of clinical outcomes
- 2:00 pm **Emil Kakkis, M.D., Ph.D.**, President, Kakkis EveryLife Foundation — Introduction to PRO interpretation challenges for rare diseases: Aldurazyme response shift case example
- 2:05 pm **Laurie Burke, R.Ph., MPH**, Associate Director, OND, CDER, and Director, Study Endpoints and Labeling Development, FDA — The latest FDA PRO Guidance translated to rare diseases
- 2:50 pm **James Witter, M.D., Ph.D.**, Chief Science Officer, PROMIS & Medical Officer, NIAMS,NIH — Customizing a patient reported outcome for rare diseases and utilizing PROMIS tools
- 3:20 pm **Nancy Kline Leidy, Ph.D.**, Senior Vice President, Scientific Affairs and Senior Research Leader, United BioSource Corporation – PROs



PROs in Rare Diseases: Kakkis questions

- Challenges using PRO's in chronic progressive diseases
- Unique and different than acute onset disease or short term changes in function
- Question of “response shift” raised
 - Patients with chronic disease adapt to their condition and rate themselves as “fine”
 - When better, their disease changes but their calibration of their response changes
- Measuring change in the context of chronic disease can be challenging
 - Trade off of recall efficiency versus value of integrated assessment: I am better or worse



Exploring Clinical Outcome Assessments in Rare Disease Trials

Laurie B Burke

Associate Director for Study
Endpoints and Labeling

Office of New Drugs, CDER, FDA



Guidance for Industry
Patient-Reported Outcome Measures:
Use in Medical Product Development
to Support Labeling Claims

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Center for Devices and Radiological Health (CDRH)

December 2009
Clinical/Medical

PRO Measures: Standard of Evidence

Feb 2006: Draft

Dec 2009: Final

<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM193282.pdf>

FDA PRO Guidance

- Explains how FDA evaluates PRO instruments for their usefulness in measuring and characterizing treatment benefit as perceived by the patient.
- Explains how FDA reviews and interprets evidence that a PRO instrument measures the concept represented by a treatment benefit claim.

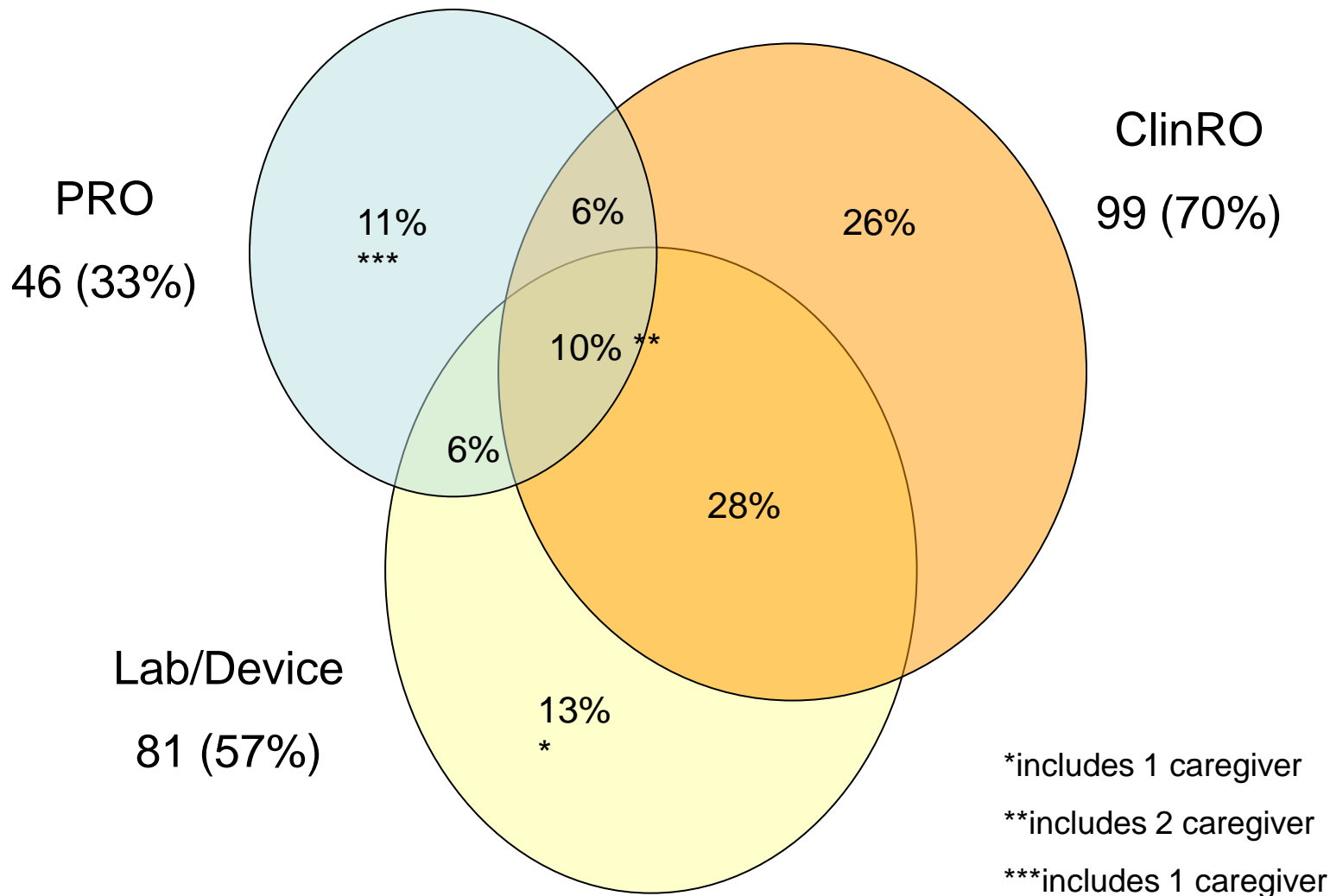
Excellent reference on the PRO Guidance

Patrick et al 2007 Patient Reported Outcomes to Support Medical Product Labeling Claims: FDA Perspective. Value in Health Vol 10: S(2); S125

Types of Outcome Assessments in Clinical Trials

- Biomarkers (non-clinical)
- Clinical Outcome Assessments
 - Patient Reported (PRO)
 - Observer Reported (ObserverRO)
 - Clinician Reported (ClinRO)

Endpoints in Labeling for 141 NMEs 2003-2008



NMEs Approved 2003-2008 Based Only on a PRO

- Relistor (constipation)
- Vimpat (seizures)
- Toviaz (OAB)
- Amitiza (constipation)
- Omnaris (rhinitis)
- Sanctura (OAB)
- Vesicare (OAB)
- Enablex (OAB)
- Prialt (pain)
- Emend (N & V)
- Aloxi (N & V)
- Levitra (ED)
- Elestat (ocular itching)
- Cialis (ED)

Content Validity

- Evidence that:
 - The score represents the intended concept in the context of use studied
 - The items in the assessment adequately cover the concept

Content Validity

- Content validity is established for the intended purpose
 - Measurement concept matches targeted claim
 - Item content development includes target population input (qualitative research)
 - Item content captures the intended concept in the intended treatment population
 - Measurement concept conforms with the proposed clinical trial objectives

Importance of Content Validity in Rare Diseases

- Subject variability identified
- Contributors to measurement error identified
- Measurement mistakes avoided
- Contributors to experiment error avoided

↳ **Variability of a COA minimized** ⇒
Increased assay sensitivity ⇒ **Better clinical study efficiency**



Content Validity: Establishing and Reporting the Evidence in Newly-Developed Patient-Reported Outcome (PRO) Instruments for Medical Product Evaluation - Good Research Practices

Donald L. Patrick PhD, MSPH, Laurie B. Burke RPh, MPH,

Chad J. Gwaltney PhD, Nancy Kline Leidy PhD, Mona L. Martin RN, MPA, Lena Ring PhD

Elizabeth Molsen RN

Liaison to ISPOR PRO Task Force Initiatives



Part 1: Eliciting Concepts

Five Steps

1. Determine the context of use.
2. Develop the research protocol for qualitative concept elicitation and analysis.
3. Conduct the concept elicitation interviews and focus groups.
4. Analyze the qualitative data.
5. Document concept development and elicitation methodology and results.



Figure 2: Example Item Tracking Matrix

Concept Name	Pain	Sleep Impact	Emotional Impact
Item Number	Item #	Item #	Item #
Concept Definition	Pain related to (condition)	Disturbance to sleep quality caused by condition-related pain	Emotional difficulties caused by condition related pain
Original Item	Since you woke up this morning, how severe was your pain?	How many times did you wake up in the night because of your pain?	How worried have you been because of your pain?
Original Item Response Options	0-10 scale (0= not severe at all, 10=as severe as I can imagine)	Enter number: _____	0-10 scale (0= not worried at all, 10=as worried as I can imagine)
Attribute to measure	Severity	Frequency	Magnitude (of worry)
Change from 1st group of cognitive interviews	Since you woke up this morning, how severe was your pain at its worst?	How many times did you wake up last night because of your pain?	No changes in first group of cognitive interviews.
Rationale for Change	Patients were not sure if they should think about their overall or most intense experience.	Patients reported seeing “in the night” as general and could mean “any night” as opposed to specifically “last night.”	
Examples of Patient Quotations	<p><i>“I had pain several times today, some I would rate low because it didn’t bother me so much, but one pain was really bad..”</i></p> <p><i>“I’m not sure if I should think about all pain in the day and average it, or just pick one I remember best to answer about...”</i></p>	<p><i>“I was thinking in an average night, how many times do I usually wake up”....</i></p> <p><i>“most nights I only wake up once or twice”</i></p>	



Guidance for Industry

Qualification Process for Drug Development Tools

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit comments to the Division of Dockets Management (HFA-305), Food and Drug Administration, 5630 Fishers Lane, rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document contact (CDER) Shaniece Gathers, 301-796-2600.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)

October 2010
Clinical/Medical

DDT Qualification Process

Oct 2010: Draft

<http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM230597.pdf>

Drug Development Tool (DDT) Qualification Process

- DDT developer provides background information so FDA can decide whether or not to participate in the qualification process for the DDT
 - Applicant may submit a **Feasibility document** (letter of intent) for
 - Targeted endpoint model
 - Targeted claim
 - Hypothesized conceptual framework
 - Rationale for DDT qualification
 - Methodology to collect evidence to support DDT qualification
 - FDA reviews Scoping document before agreeing to start the qualification process

- *Qualification Stage 1: Consultation and advice*

- **Qualitative research summary** { e.g., qualitative research protocols (concept elicitation, cognitive debriefing), transcripts, reports, draft of the instruments, revised CF, translation methodology, item map
- **Instrument finalization summary** { e.g., protocols for item reduction, analyses plan, scoring algorithm, draft of the instruments, item map
- **Quantitative research summary** { e.g., protocols, analyses plan, report, final instrument, item map, interpretation,

- Once DDT is ready for qualification, the Agency will conduct the final review of the **Qualification Package**
- Public notice of DDT qualification
- Public availability of DDT tool

Benefits of Collaboration

- Allows FDA to work with multiple partners to leverage expertise and resources toward PRO instrument development
 - Improved/transparent FDA internal (review) processes
 - Pooling of industry know-how and resources
 - Provide a basis for eventual comparison of labeling claims by physicians
 - Issuance of best practices and guidances toward future instruments and product development

Summary

- Terminology is becoming standardized
 - Concept
 - Treatment benefit
 - Context of use
 - Content validity and other measurement properties
 - Clinical Outcome Assessments (COAs)
- Good measurement begins with content validity and principles apply to all COAs--PROs, ObserverROs and ClinROs
- The regulatory qualification process offers the potential for increased efficiency in the development of “well-defined and reliable” outcome assessments for use in clinical trials



PROMIS[®]

Patient-Reported Outcomes Measurement Information System[®]

KAKKIS Foundation Clinical Evaluation of Rare Disease Treatments

June 14-15, 2011
Washington DC

James Witter MD, PhD
Chief Science Officer
NIH/NIAMS

www.nihpromis.org

Dynamic Tools to Measure Health Outcomes from the Patient Perspective



NIH Assessment Tools

- **NIH Toolbox** www.nihtoolbox.org/default.aspx
 - *Assessment of Neurological and Behavioral Function*
- **Neuro-QOL** www.neuroqol.org/default.aspx
 - *Quality of Life in Neurological Disorders*
- **The EXAMINER** <http://examiner.ucsf.edu>
 - *Neurobehavioral Evaluation and Research*
- **Phen-X** www.phenx.org
 - *Consensus Measures for Phenotypes and Exposures*
- **PROMIS[®]** www.nihpromis.org
 - *Patient-Reported Outcomes Measurement Information System*

PROMIS®

Dynamic Tools to Measure Health Outcomes from the Patient Perspective



[About PROMIS®](#)

[Measures](#)

[Science](#)

[Software](#)

[What's New](#)

[Related Resources](#)



[PROMIS® For You](#)

[Search](#)

Patient Reported Outcomes Measurement Information System (PROMIS), funded by the National Institutes of Health (NIH), is a system of highly reliable, valid, flexible, precise, and responsive assessment tools that measure patient-reported health status.

Researchers

Provides efficient, reliable, and valid assessments of adult and child (pediatric) self-reported health

- ▶ [PROMIS Instruments Selected References](#)
- ▶ [PROMIS In Research](#)



Clinicians

Provides data about the effect of therapy that cannot be found in traditional clinical measures

- ▶ [PROMIS for Clinicians](#)
- ▶ [Select Publications](#)
- ▶ [Computer Adaptive Test \(CAT\) Demonstration](#)



Patients

Measures what you are able to do and how you feel

- ▶ [More on PROMIS](#)
- ▶ [What Patient Reported Outcomes \(PROs\) are](#)
- ▶ [PROMIS Measures](#)



USA.gov

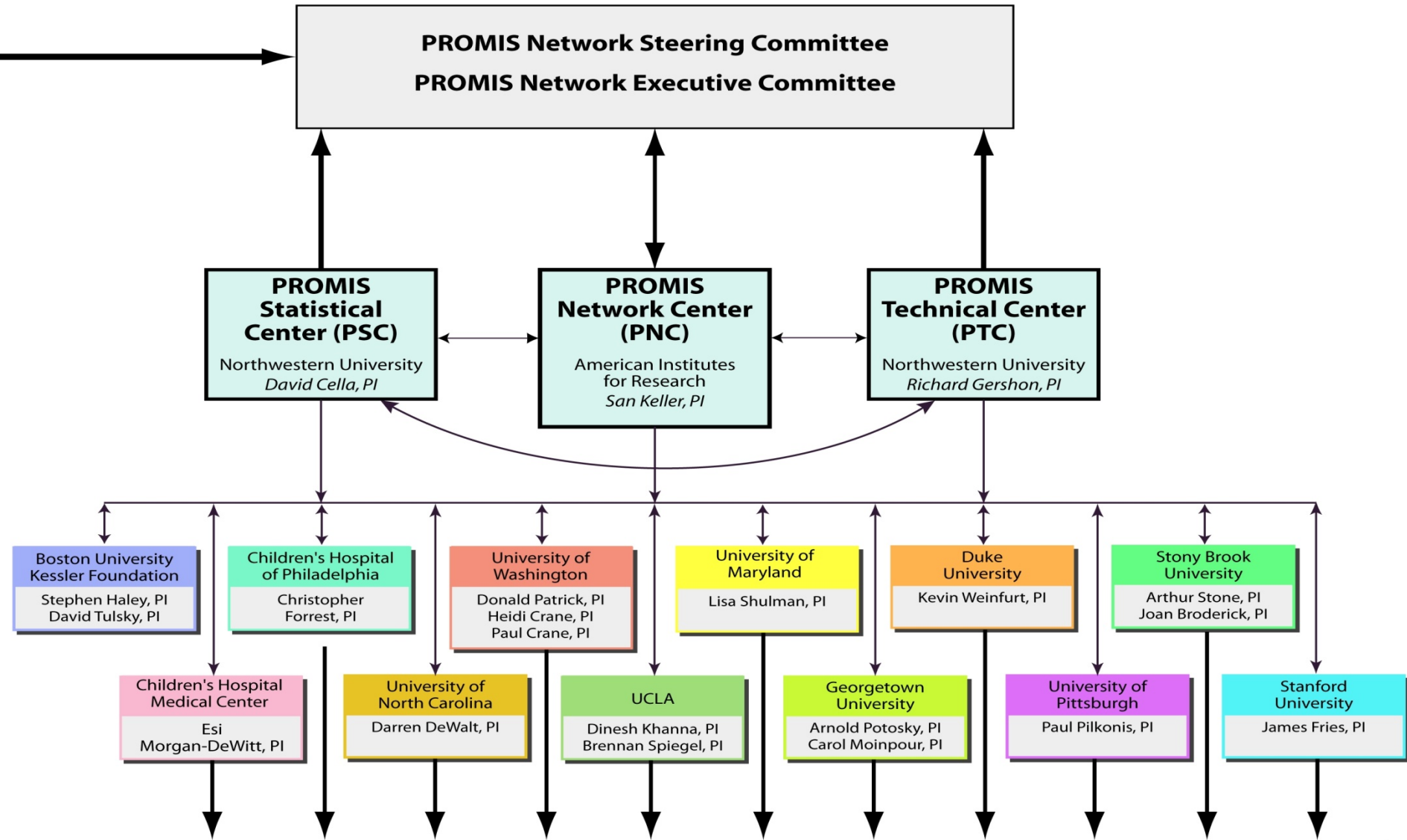
[Contact PROMIS](#) | [Policies](#) | [Accessibility](#) | [FAQs](#) | [Glossary](#) | [File Formats](#) | [Site Map](#) | [PROMIS Researcher Login](#)



PROMIS: Present



PROMIS II Network Diagram



FDA/C-PATH PRO Process

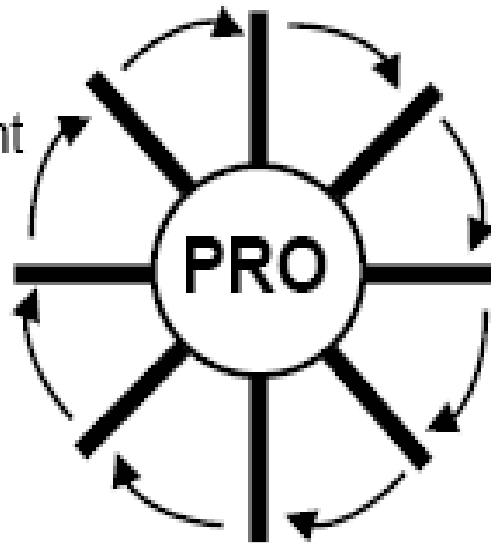
FDA/C-Path meeting: Dec. 4, 2008

i. Identify Concepts and Develop Conceptual Framework

Identify concepts and domains that are important to patients.
Determine intended population and research application.
Hypothesize expected relationships among concepts.

iv. Modify Instrument

Change concepts measured, populations studied, research application, instrumentation, or method of administration.

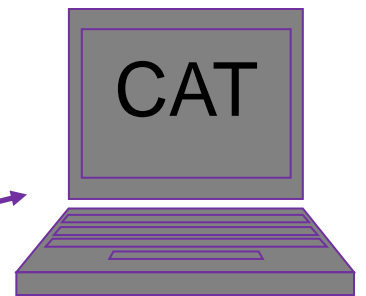
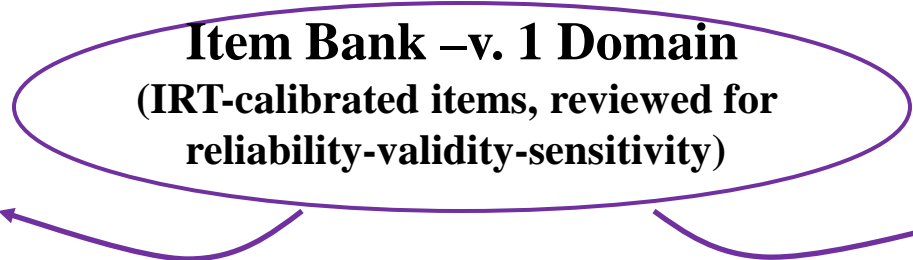
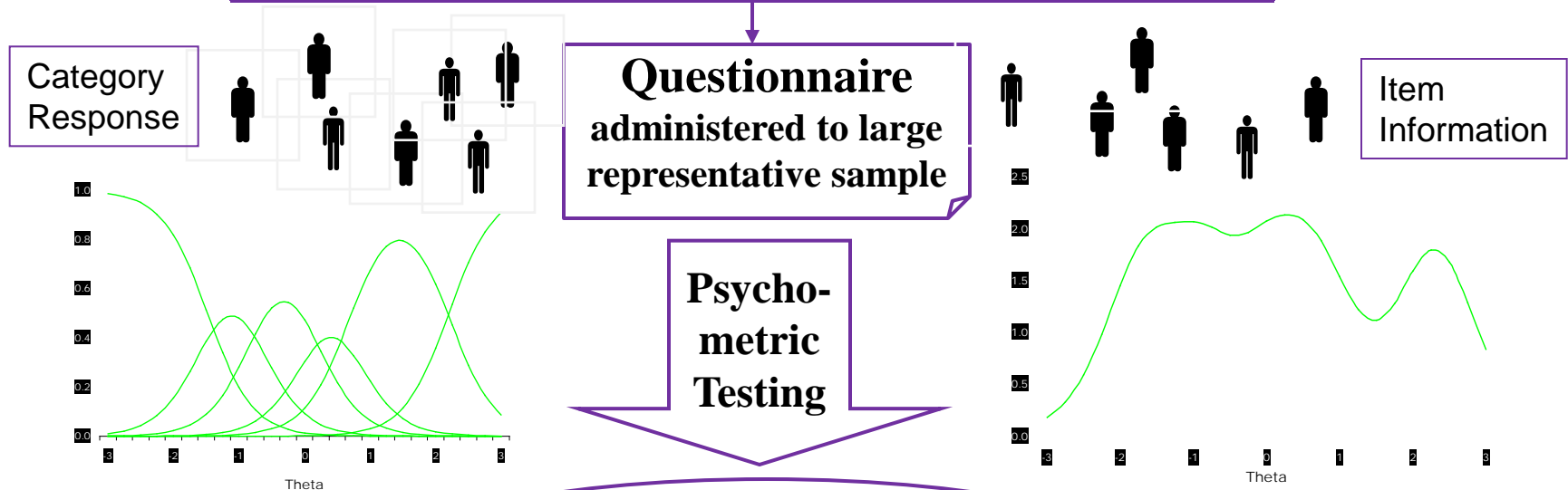
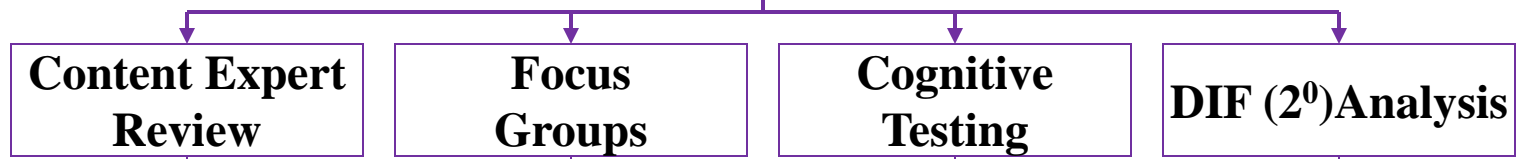
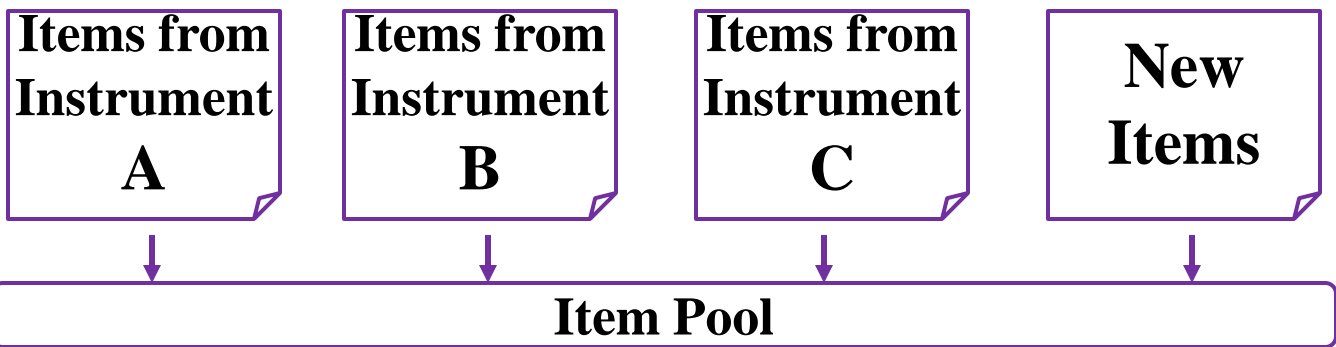


ii. Create Instrument

Generate Items.
Choose administration method, recall period, and response scales.
Draft Instructions.
Format Instrument.
Draft procedures for scoring and administration. Pilot test draft instrument. Refine instrument and procedures.

iii. Assess Measurement Properties

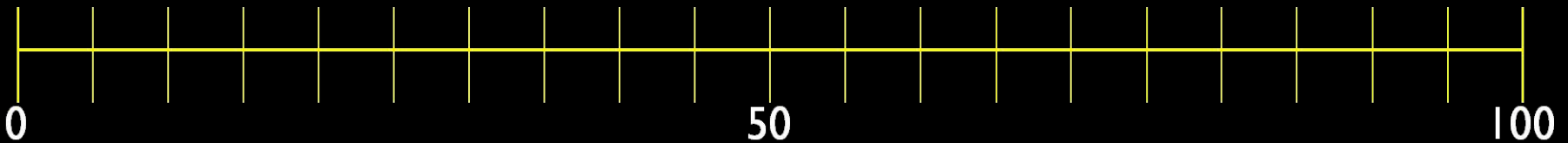
Assess score reliability, validity, and ability to detect change.
Evaluate administrative and respondent burden. Add, delete, or revise items.
Identify meaningful differences in scores. Finalize instrument formats, scoring, procedures, and training materials.



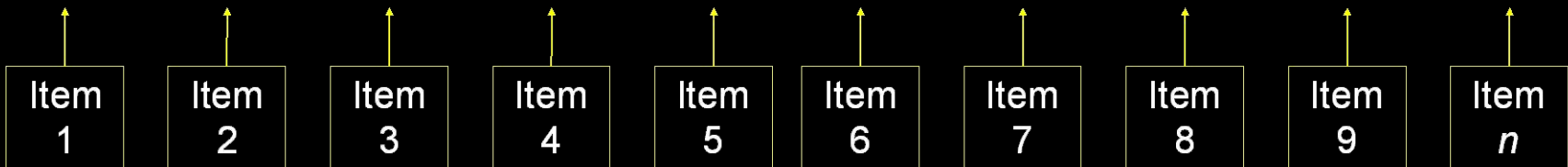
- Items are questions that have:
 - Context, stems, responses
 - Parameters (Difficulty, Discrimination)
 - Important for analyses, calibration
- Banks
 - Collection of items that have been “calibrated” on a common metric
 - difficulty and discrimination have been estimated
 - Define common concept-domain
 - Allow computerized adaptive testing (CAT)
- Domains
 - Define latent traits/abilities

*An **item bank** is a large collection of items measuring a single domain.*

Any and all items can be used to provide a score for that domain.



Physical Functioning Item Bank



Are you able to get in and out of bed?

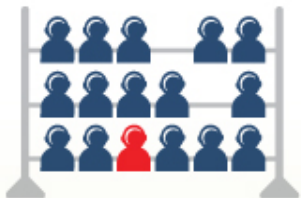
Are you able to stand without losing your balance for 1 minute?

Are you able to walk from one room to another?

Are you able to walk a block on flat ground?

Are you able to run or jog for two miles?

Are you able to run five miles?



RARE DISEASE WORKSHOP SERIES

Improving the Clinical Development Process

Addressing Content Validity of PRO Measures: The Unique Case of Rare Diseases

Nancy Kline Leidy PhD
Sr. Vice President, Scientific Affairs
United BioSource Corporation
Bethesda MD



Overview

- I. The Concept & Content Validity
- II. Measurement Options
- III. The Unique Case of Rare Diseases
Challenges & Opportunities
- IV. Conclusions



Concept Clarification

- What is the target patient population?
 - Diagnosis, stage of disease, age group
- What is the clinical problem?
 - Biologic, physiologic, symptomatic, functional
- How is this problem to be addressed?
 - Intervention - Product or product line
- What is the intended outcome/concept/claim?
 - Improve? Stabilize? Prevent?
- How is this outcome currently defined?
 - Empirically? Clinically?
- How is this outcome (concept) currently measured?
 - Instrument? Other?
- Is this approach or instrument appropriate?
 - Fit for purpose? Sufficiently sensitive?



Content Validity

- The extent to which scores produced by a research instrument represent the target concept(s).
 - contains the relevant & important aspects of the concept.
 - contains a sufficient sampling of content to represent the concept



Addressing Content Validity of an Existing Instrument

- Rothman, M, Burke, L, Erickson, P, Leidy NK, Patrick D, Petrie CD. (2009). Use of existing patient-reported outcome (PRO) instruments and their modification: the ISPOR good research practices for evaluating and documenting content validity for the use of existing instruments and their modification PRO task force report. Value in Health, 12 (8): 1075-83; Epub 2008 Sept 25.



Threats to Validity of Existing Instruments (Rothman et al, 2009)

- Absent or unclear conceptual match between the instrument & claim
- Lack of direct patient input into item content from the target population.
- Lack of evidence regarding saturation – no evidence that the most relevant and important item content is contained in the instrument
- Modification (Adaptation) of an instrument



Adapting Instruments

- “Adaptation” includes:
 - Content – Item stems or response options
 - Change, add, or delete
 - Instructions
 - Recall
 - Mode of administration



Qualitative Methods: Evaluating Existing or Adapting Instruments

Concept Elicitation – Content Mapping

- Focus Groups and/or Interviews
 - Sample size varies based on concept and a priori knowledge
 - To saturation
 - Possible Range: 15-30 patients, more or less

Cognitive Interviewing - Evaluation & Understanding

- Interviews
 - Sample size varies – to assure comprehensiveness & clarity
 - Possible Range: 10 to 20 patients, more or less



Addressing Content Validity of a New Instrument

- Patrick D, Burke L, Gwaltney C, Leidy NK, Martin M, Ring L. Establishing and reporting evidence of the content validity of newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: Good research practices, Part 1 – Eliciting concepts for a new PRO instrument. ISPOR Task Force Report under submission, Value in Health.
- Patrick D, Burke L, Gwaltney C, Leidy NK, Martin M, Ring L. Establishing and reporting evidence of the content validity of newly-developed patient-reported outcomes (PRO) instruments for medical product evaluation: Good research practices, Part 2 – Assessing respondent understanding. Task Force Report under submission, Value in Health.



Qualitative Methods: Developing a New Instrument

Concept Elicitation

- Focus Groups:
 - Generally, 4 to 8 groups of 5 to 8 people
- Interviews
 - Often 15 to 40 people
- Broad concepts require more participants

Evaluation & Understanding

- Interviews
 - Generally 5 to 20 people



Select Focused Outcomes

- Example
 - “function” versus range of motion or muscle strength
 - “fatigue” versus muscle strength/weakness
 - “health-related quality of life” versus pain
- Advantages
 - Easier to understand and communicate
 - Less qualitative data required to achieve saturation
 - Likely to be less variable – within and between patients



Use or Adapt Existing Instruments

- Develop a disease model (Patrick et al, *Value Health*)
- Match content (Rothman et al., *Value Health*, 2009)
- Select/decide carefully
 - Existing \neq Good
- If match, document content validity
 - Cognitive interviewing
 - Elicitation & cognitive interviewing



Consider Alternative Methods

Be creative and scientific

- “Modes” of data collection
 - Telephone Interview
 - Virtual focus groups – conference call, web camera
- Sample/respondents
 - Excellent informants describing patient experiences
- Existing resources
 - Registries, Patient advocacy group
 - Exit interviews in clinical trials
- Validation studies
 - Phase II, patient registries?
- Option B
 - Post-approval PRO labeling?



Patient-Reported Outcomes

- Represent the patient's voice in research.
- Content validity = the accuracy of the voice
 - Scores represent the content (voice)
- Patients with rare disease deserve an accurate voice.
- Optimizing content validity is challenging.
- There are opportunities to address these challenges.



Workshop #2 Agenda: Day 1

Patient Reported Outcomes

Adapting existing versus custom solutions

- **3:50 pm Discussion**
 - For rare diseases, when should developers adopt existing common instruments as their primary strategy for using PROs, and when should they develop new, disease-specific instruments?
 - How should a developer meet the guidelines set forth in the FDA PRO Guidance to qualify an endpoint while managing the challenges posed by rare disease drug development, both in adopting an existing instrument and developing one from scratch?
 - Can PROs be used to support the clinical meaningfulness of clinical outcomes?



Key conclusions from the PRO discussion relevant to rare diseases

- Developing new instruments is difficult to meet the FDA guidance in rare diseases
- Best choice: Adapt a high quality instrument (see Leidy's talk)
 - Pick a well developed, validated instrument
 - Pick a domain-focused instrument not a broad: Easier to adapt and verify content validity
 - Assess content validity
 - Adapt and retest cognitive testing (see Leidy and Burke) along with retesting if needed



RARE DISEASE WORKSHOP SERIES
Improving the *Clinical Development Process*

Clinical Evaluation Tools for Rare Diseases

Workshop Day #2 Morning



Workshop #2 Agenda: Day #2

Rare Disease Experiences

Day 2 – Morning

8:00 - 8:30 am Breakfast

8:30 am Brief previous day review

Emil Kakkis, M.D., Ph.D., President, Kakkis EveryLife Foundation

- Examples of development of clinical evaluation tools to assess degree of clinical benefit in neurological and rare diseases

8:50 am **Ron Crystal, M.D.**, Prof. of Genetic Medicine, Weill Cornell — Natural history data, development of LINCL clinical scale, interpretation of brain imaging scans

9:20 am **Ali Skrinar, M.A., MPH**, Senior Director, Clinical Research & Regulatory Affairs, Enobia Pharma — Development of rare disease evaluation tools

9:50 am Midmorning Break

10:10 am **P.K. Tandon, Ph.D.**, Senior Vice President, Global Biomedical Data Sciences and Informatics, Genzyme Corporation — Use of registry information to inform on clinical evaluation of rare diseases

10:40 am **Andy Blight, Ph.D.**, Chief Science Officer, Acorda Therapeutics — Successful application of patient reported outcomes



Rare Disease Experiences: See original slides for optimal details

- Ron Crystal on neuronal ceroid lipofuscinosis and neurological assessment (see original slides)
- Ali Skrinar on developing instruments for rare diseases and examples (summarized here)
- PK Tandon on the use of registries (see original slides)
- Andy Blight on the use of a PRO to support a performance measure (see original slides)



RARE DISEASE WORKSHOP SERIES
Improving the *Clinical Development Process*

Development of Rare Disease Evaluation Tools

Alison Skrinar

Senior Director, Clinical Research &
Regulatory Affairs

Enobia Pharma



Presentation Overview

- Define burden of illness
- Outline potential uses for this data
- Provide rationale for instrument development in rare diseases
- Review development/validation process for PROs
- Present challenges to development, validation and interpretation of data
- Provide examples of instruments developed for rare disease
- Discuss path forward



How Do You Establish Burden of Illness?

- Quality of life
 - Physical health scores may show significant limitations
 - Mental health scores may not be abnormal if clinical course is slowly progressive
- Resource utilization
 - Many disease-related complications require surgical intervention and the use of medical devices
 - Overall consumption may be low when progression is slow, no treatment options exist and/or care is palliative
- Disability
 - Impact of disease on FUNCTION as described by the PATIENT



What Type of Instrument Do You Need?

Patient-Reported Outcomes

- Advantages:
 - Most reliable source of information about the clinical symptoms of a rare disease and the impact of these symptoms on daily function
 - Best means of putting a “face to a name”
- Disadvantages:
 - Scores affected by compensatory behaviors
 - May not be sensitive enough to detect small changes in function in a clinical trial setting
 - Long and arduous development and validation process



What Type of Instrument Do You Need?

Performance Measure

– Advantages:

- Better reflection of the physical capabilities of patient
- Increased ability to detect small changes in function in a clinical trial setting
- Easier to interpret findings and put into context for various audiences

– Disadvantages:

- Limits ability to identify the source of the change
- Limited ability to characterize the value of a new therapy for rare disease



Why Not Use Existing Instruments?

- Rare diseases are not well characterized in the literature from a functional perspective
- Clinical manifestations of rare diseases are often unique and multi-systemic
- Substantial inter-relationships among clinical symptoms that can't be captured with a single instrument
 - What CAN'T the patient do?
 - Why CAN'T they do it?
 - Are the reasons the same for different diseases?



How Do You Develop a New Instrument?

- START EARLY AND START WITH THE PATIENTS!
 - Natural history studies are critical to establishing the burden of disease
 - Best source of information on diagnosis, clinical presentation, disease progression and disability
 - Heart of clinical trial design
- FDA Guidance for Industry: PROs
 - Endpoint model
 - Choice of PRO instrument
 - Conceptual framework
 - Content validity
 - Reliability, other validity and sensitivity to change
 - Instrument modification



Why Aren't More PRO Instruments and Performance Measures used in the Rare Disease Drug Development Process?

- Natural history studies and instrument development are started too late
- FDA Industry Guidance for PROs is difficult to implement in rare disease
- Small, multinational, clinically heterogeneous patient populations complicate validation process
- Need for statistical power in clinical trials limits patient selection process
- Is there a way to adjust this process to accommodate rare diseases?



RARE DISEASE WORKSHOP SERIES
Improving the Clinical Development Process

**Multi-domain assessment:
Strategy and Evaluation using
Phase 3 MPS ERT Data**

Workshop Day #2 Afternoon



Workshop #2 Agenda: Day 2

Multi-domain assessment of rare disease

Day 2 - Afternoon

- Evaluation of clinical benefit across multiple physiological domains
 - 12:55 pm **Emil Kakkis, M.D., Ph.D.**, President, Kakkis EveryLife Foundation — The necessity for improving the accessibility to multi-domain assessments in rare disease treatment development: A proposal for moving forward
 - 1:25 pm **L.J. Wei, Ph.D.**, Professor of Biostatistics, Harvard School of Public Health — Approaches and challenges to multi-domain analysis and study design
 - 1:45 pm **James Signorovitch, Ph.D.**, Manager, Analysis Group — Evaluation of various statistical approaches for multiple domains using data from three MPS studies



RARE DISEASE WORKSHOP SERIES

Improving the Clinical Development Process

The necessity for improving the accessibility to multi-domain assessments in rare disease treatments

Emil D. Kakkis, M.D., Ph.D.
President, EveryLife Foundation



Evaluation of Multi-system Diseases: An area in need of innovation

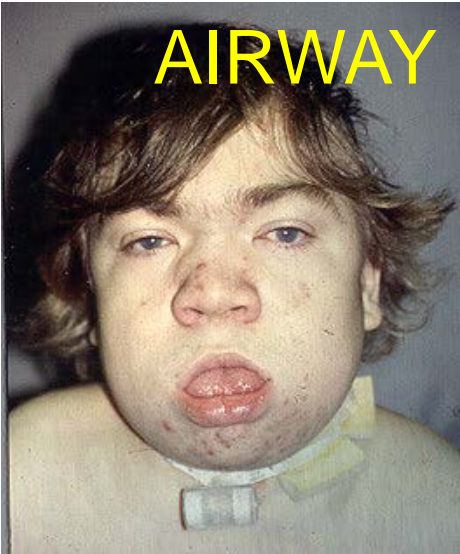
- Original targets of drug development
 - Single clear indication (HTN, Asthma, Infection)
 - Driven by single problems/primary endpoints
- Next generation of diseases targets
 - Multisystem complex, heterogeneous diseases
- Does the current paradigm for clinical evaluation make sense for multisystem diseases?



Evaluation of complex heterogeneous diseases will not be simple

- Heterogeneity in disease domain severity
- Hard to know how interplay between clinical domains affects outcomes
- Irreversibility/progression
- Few patients to explore issues

MPS I: Devastating multi-system disorder



FOUNDATION

No Disease Is Too Rare to Deserve Treatment



How heterogeneous is disease?

- 6 min walk test $\sim 1/3$ rd normal
- Forced vital capacity $\sim 1/3^{\text{rd}}$ normal
- Shoulder ROM $1/2 > 90$ degrees
- Sleep apnea AHI $1/2 < 10$ events/hr
- Visual acuity $2/3^{\text{rd}}$ near normal

Very hard to conduct studies focused on one endpoint and still get useful data on other domains



Conducting a clinical study

- Pick the endpoint
- Select “right” affected patients
- Study the endpoint
- File for approval
- Get a single claim



What does it really mean?

- Do we understand what the drug does for all aspects of each patient?
- Do we understand the complexity of who benefits?
- Do we understand how disease heterogeneity varies response?
- Do we understand safety in all types?



Proposal

New study in a new disease

- **Enroll all patients** except in whom harm or benefit is not justified
- Treat all patients
- Evaluate 5-6 major domains of disease (individual endpoints)
- Analyze data by continuous variable methods for significance



Analysis of this new design

- Each patient evaluated only for each disease domain affected
(a priori inclusion criteria for subsets)
- Each domain evaluation managed for clinical meaningfulness in evaluation
- Multi-domain continuous variable analysis completed



Interpretation if positive study

- Each domain evaluated independently using proper affected subset
 - Statistical significance
 - Responder/clinical significance
- Label constructed from positive primary and individual secondaries



Benefits

- Safety in a wide variety of patients
 - No selection based on 1^o Endpoint
- Learn about more domains
- Avoids hazard of picking wrong on a primary (manage multiplicity)
- Capture benefit across multiple domains in variable patients



Discussion of approaches to multi-domain assessments: LJ Wei

- A number of possible methods (see his slides)
- Complexities in application and interpretation
- Different methods for different type of data
 - See slides for details
- Assessing multi-endpoints within an individual patient might be more relevant than combining endpoints across patients



RARE DISEASE WORKSHOP SERIES

Improving the Clinical Development Process

Statistical Analyses for Multi-Domain Outcomes

James Signorovitch, PhD,
Analysis Group Inc, Boston MA



Objectives

- Illustrate the use of multi-domain methods using MPS trial data
- Assess tradeoffs between interpretability and sensitivity vs. single domain analyses

Interpretability	Sensitivity
<ul style="list-style-type: none">-How can we assess whether there is substantial benefit?-How can we interpret the benefit?	<ul style="list-style-type: none">- When is a multi-domain approach more sensitive than a single domain?



Wilcoxon and O'Brien Test Results in Individual Trials

- In these 3 MPS trials, the 3-domain O'Brien test would have been more sensitive than Wilcoxon tests based on any single domain

Test / Outcome	#Trials with a Statistically Significant Treatment Effect
Wilcoxon Tests	
Outcome 1	1* out of 3
Outcome 2	0 out of 3
Outcome 3	1* out of 3
O'Brien test combining all three outcomes	2** out of 3

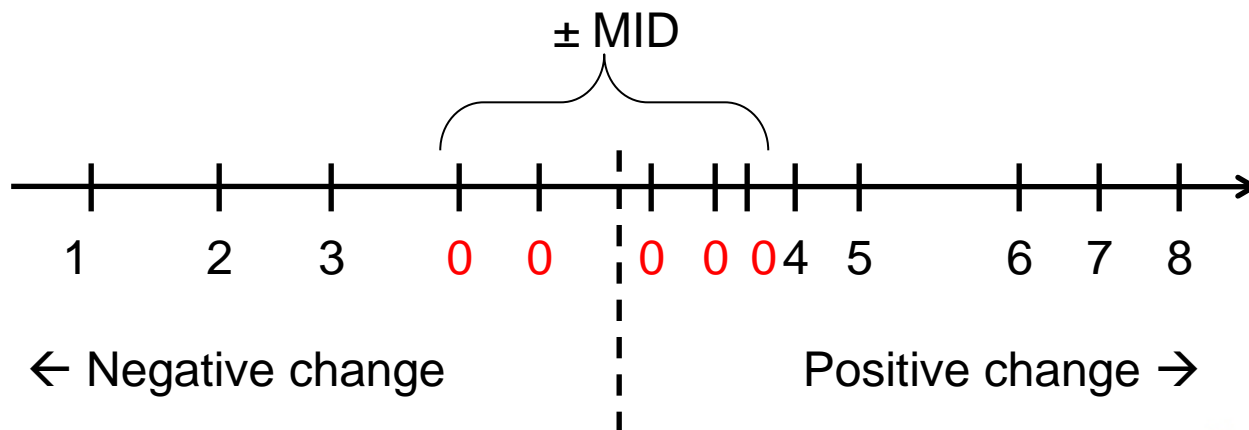
*Different trials

** Significance in all 3 trials if using absolute vs. %-predicted FVC per protocol



1. Is this change clinically meaningful?

- Partial answer: sensitivity analysis:
 1. Set to 0 all changes with magnitude $<$ the MID and re-rank
 2. Re-run the O'Brien test with the new ranking





Wilcoxon and O'Brien Test Results in Pooled Trial Data: Sensitivity Analysis

- The treatment effect remained after ignoring changes in individual domains that did not exceed the MID
- Difference is not due only to many changes $<$ MID
- Still need to assess whether the treatment difference is meaningful

Test	Mean Rank Sum Difference	SE	P-Value
Wilcoxon tests			
6MWT	20.0	5.8	<0.001
FVC	20.5	5.9	0.007
Shoulder	4.9	5.2	0.341
O'Brien (6MWT, FVC, Shoulder)	15.1	3.4	<0.001



Correlations with CHAQ/HAQ

The patients perception of health change correlated better with the 3 endpoints together (O'Brien) than each individually

Objective assessment	Pearson Correlation with CHAQ/HAQ*	P-Value
Continuous scores		
6MWT	-0.24	0.032
FVC	-0.19	0.091
Shoulder	-0.23	0.051
O'Brien rank score (6MWT, FVC, Shoulder)	-0.50	<0.001

*Patient-assessments pooled from treatment and placebo arms (n=76); Higher CHAQ/HAQ scores indicate greater impairment .



Workshop #2 Agenda: Day 2

Multiple Domain Assessment in Rare Diseases: How do we develop a multi-domain strategy?

- **2:15 pm Discussion**

- How do we capture and interpret the real effect of a treatment on patients using multi-domain endpoints?
- How would we then interpret clinical data from multi-domain results and accurately make labeling claims based on the result of a multi-domain analysis?



RARE DISEASE WORKSHOP SERIES
Improving the Clinical Development Process

Key Findings for Workshop #2

Emil Kakkis, M.D., Ph.D.



Endpoints

Key findings relevant to rare diseases

- Selecting Endpoints
 - Know the disease well for content validity
 - Natural history, survey studies
 - Establish a framework for the diseases
 - Provide possible endpoint plan
 - Talk to FDA frequently



MID/Responder Definitions

Key findings relevant to rare diseases

- Minimally important differences
 - Establish via a global ratings anchor better than a distributional method
- MID's are difficult to establish in rare diseases without some treatment data or natural history data
- Use of statistically based MID's is not recommended
 - MID as one std dev. of the mean for example
- Preferable to establish significance for continuous variables (no threshold effect) with continuous analysis first
 - Use MID/responder for clinical interpretation afterward



PRO for Rare Diseases

- Full development is multistep and complex to do well: plausible for rare but perhaps not for ultra-rare
- For a new instrument, start early with development and get natural history data; consult guidance and key publications; expert consultant recommended
- For adaptation of PRO tool
 - Pick more focused tool to a domain or function (content valid/PROMIS)
 - Perform content validity testing and cognitive to test tool in specific population
 - Interviews plausible to do 10-20 plus 5-10 for cognitive testing



Multi-domain analysis does show promise in these MPS study analyses

- Increased power to detect change
- Works even if we selection criteria removing change $< \text{MID}$
- Correlation between the multi-domain assessment was twice as large with patient reported health assessment
- Need new standards for using multi-domain analyses without creating mathematical composites



Output and actions from the workshop

- Presentations to be posted at www.kakkis.org
- Will consider writing review articles with multiple authors if discussion warrants
- Consider incorporating ideas in policy recommendations at Workshop #4
- Provide information to key decision makers if relevant to policy decisions