



## 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.

# Assessments and Endpoints

## *Demonstrating Effectiveness*

Clinical Evaluation in Rare Diseases Workshop  
Washington, DC  
June 2011

Marc K Walton MD, PhD  
Associate Director for Translational Medicine  
Office of Translational Sciences  
CDER-FDA

*The views expressed are those of the author, and do not necessarily represent an official FDA position*

# Clinical Evaluation of a Treatment

- MID
  - What is the MID?
    - ❖ Not the primary question, nor
      - How do we determine the MID?
      - Is the MID good enough?
  - The MID of What?
  - Initial question is
    - ❖ What assessments will be useful?
    - ❖ What is the endpoint that should be used?
  - Evaluation of MID
    - ❖ Suited to the type of assessment
    - ❖ May follow reasoning that led to selection of assessment

# What is an Endpoint?

- Study endpoint consists of
  - A specific clinical outcome assessment
    - ❖ Evaluated at specific time(s) & circumstances
  - Analyzed in a specified manner
  - Both aspects effect the utility of the endpoint
- Careful selection of the study endpoint is important to an efficient and successful clinical development program
  - Endpoint is an aspect of the program; not just the study
  - Previously emphasized that many features of the disease and patient population, and early attention to these, are important to understand

# Endpoint Utility

- Utility of a study endpoint derives from
  - Ability to convincingly show a treatment effect in a clinical study
  - Ability to interpret that effect as *effectiveness*
    - ❖ A tangible benefit to the patient
- Marketing approval for a drug based on demonstrating that:
  - The drug is effective
  - The benefits outweigh the risks

# Selecting a Clinical Outcome Assessment (COA) <sup>(1)</sup>

- Comprehensive consideration of all the different effects of the disease on the patients
  - Activities affected
  - Body elements impaired producing the disability
  - Prevalence of each aspect
  - Concordance of impairment patterns across patients
  - Range of severity
  - Rate of worsening
  - Within patient variability (day to day, wk to wk)
  - Identifiable phenotype categories
- Select the aspect(s) of patient clinical status best suited to be the objective for measurement and evaluation of treatment effects

# Selecting a COA (2)

- Are there any existing well described COAs that measure or relate to that clinical aspect?
  - What are the measurement characteristics of those tools?
    - ❖ Coarse vs fine gradations
    - ❖ Intra-patient, inter-evaluator reliability
    - ❖ Burden to obtain measurement
    - ❖ Range of measurement relative to this patient population (floor, ceiling limitations)
  - Are any well-suited?
- Alternative is create a new COA
  - Designed to suit this disorder

# Types of COAs

- Direct observation, recording of patient's typical daily functioning
  - Self or observer or interview with clinician
- Report of activities or events in usual daily life that are thought to be due to the selected impaired functions but not directly meaningful
  - E.g., record of as-needed pain medication use
- Measurement of an activity not a part of usual daily life
  - E.g., *Artificial procedure* performed in clinic
  - Thought to be evaluating impaired abilities that are used to perform daily life activities
  - Clear articulation of related usual daily life activities that are intended; Often are not self-evident

# Artificial Procedure COA (1)

- Can be uniformly applicable to all study patients
- Can be administered in a consistent manner.
- May be structured to stress-test the isolated basic actions
  - The normal daily life activity might not push the body-function to maximal functional ability
  - May be very sensitive to changes in functional ability of a patient
  - Over-stressing can introduce non-meaningful variability (noise)

## Artificial Procedure COAs (2)

- They are indirect measures of a meaningful aspect of the disease's effects on the patient
  - The patient does not usually perform these procedures in daily life
    - ❖ E.g., 6-minute walk, ETDRS visual acuity tests
  - Are meant to imply some functional ability of the patient in daily life
  - Measurements cannot be intrinsically interpreted as to clinical meaning
- Prospective planned efforts enable linkage to 'real' daily activities, interpretability
- Development of new tests, qualification of existing tests in new a patient group initiated in advance of A&WC study

# Selecting a COA <sup>(3)</sup>

- Construction of new COA with consideration of natural history (slide 3) may be needed
  - Existing, but ill-suited, COAs for other disorders may impair sensitivity to treatment effects
- Disease expression between patients important
  - Uniform vs. variable?
  - If variable a single feature-focused assessment may not detect benefit to patients where the selected feature is not (presently) prominently affected

# Selecting a COA (4)

- Consider using multiple assessments when expression variability present
  - Can ensure all patients have at least one substantially affected ability included among the assessments
    - ❖ Enables detecting treatment's benefit in each patient
  - Combine the multiple assessments in endpoint
    - ❖ Multiple analytic methods available to combine
    - ❖ Interpretability of endpoint can be differently affected by different analysis methods
  - Some multi-domain PROs may be intended, in part, to employ this approach

# Analysis and Interpretability (1)

- Analysis method of COA impacts the intrinsic interpretability of the endpoint's observed treatment difference
- There is a tension between sensitivity of the endpoint and interpretability
  - Finely gradated continuous scale COAs analyzed in that form may be sensitive to small differences
  - Often easier to judge meaning of larger differences in measured values
  - Consider when specifying endpoint
    - ❖ Endpoint = COA + analysis

# Analysis and Interpretability (2)

- Some COAs have natural analysis method
  - Clinical Event endpoints
    - ❖ Difference or ratio of rates
    - ❖ Time to event
- Continuous scale COA
  - Analyze in continuous form
  - Analyze after changing into some categorized form
    - ❖ Categorization by outcome measurement
    - ❖ Categorization by change from baseline
    - ❖ Less sensitivity traded for greater interpretability

# Interpretation

- General clinical meaningfulness of a COA
  - Does not establish clinical meaningfulness of any particular observed treatment difference
- Understanding the meaning of an observed treatment difference essential
  - Value of observed treatment effect is treatment's benefit
- Approval based on judgment that benefits outweigh risks

# Interpretation

- MID: Minimum effect size with value to the patient
  - Effect size with minimal value to the patient
- Relative value of larger effect sizes also useful to understand
  - Differences with minimum, moderate, large importance
    - ❖ Especially when risks are not minimal
  - MID not important if observed treatment effect can be interpreted without precisely knowing MID
  - Smallest confidently-affirmed important difference
    - ❖ May be larger than the unknown MID
    - ❖ Difficulty of achieving high precision in interpretability or
    - ❖ Difficulty from reliability of the instrument

# Interpretation

- Clinical meaning is not a purely statistical evaluation of the COA
  - E.g., 2 sd of intra-patient variability may identify a reliably detectable difference
    - ❖ Does not establish that it is a meaningful difference
  - 'True' MID may be greater or less than reliably determined difference
- For continuous scale or interval COAs, the value to the patient of small changes may not be the same at different locations in the scale
  - E.g. 5 pt difference from baseline 42 may not mean the same as from baseline 87 or 23

# Achieving Interpretability

- Different categories of COAs will have different approaches
  - Types of information and relative weights
- Naturalistic COAs
  - May be able to rely heavily on face validity for assurance of broad meaningfulness
    - ❖ E.g., pain VAS, individual sub-elements of a questionnaire
  - Psychometric properties when combined into complex tool
  - Contains directly meaningful items that aid establishing interpretation of indirectly-meaningful reports

# Achieving Interpretability

- Artificial procedure measurement COA
  - More complex to establish meaningfulness
  - Specify the normal life aspect it is intended to represent
  - How are COA measurements related to reports of those activities
- FDA Guidance on PRO tools
  - Specifically for creation and qualification of PROs
  - Conceptual elements of guidance also relevant for other tools
  - Will be discussed later talks in this conference

# Achieving Interpretability

- Information basis for determining meaningfulness of a COA / endpoint
  - Relevant to specific type of patients in the intended clinical trial
  - Might not be all patients with any form / stage of the disease
  - Same disease or, possibly, an adequately similar different disease

# Endpoint Effect and Interpretation

- Does clinical meaning of the effect need to be an intrinsic part of the endpoint?
  - Not necessarily
- Interpretation easier if meaningfulness is an intrinsic part of the endpoint
  - E.g., difference in percentage of patients experiencing a change that is confidently meaningful (responder type of endpoint)
  - Statistically significant difference is automatically clinically significant

# Endpoint Effect and Interpretation

- Not doing so has program risks
  - Study primary endpoint result may leave uncertainty whether there is tangible benefit
  - Benefit-risk comparison may be uncertain
    - ❖ Potential for critical differences in judgment of favorable vs unfavorable
    - ❖ May prevent or delay marketing approval
  - Tension between high sensitivity vs clarity of meaning should be considered
- Separate conclusions feasible
  - Well-planned, prospective approach as to how meaningfulness of observed effect is established
  - Study, other data, planned to be persuasive