IMPLICATIONS OF FDA PATIENT-REPORTED OUTCOMES GUIDANCE FOR ASSESSING PAIN INTENSITY

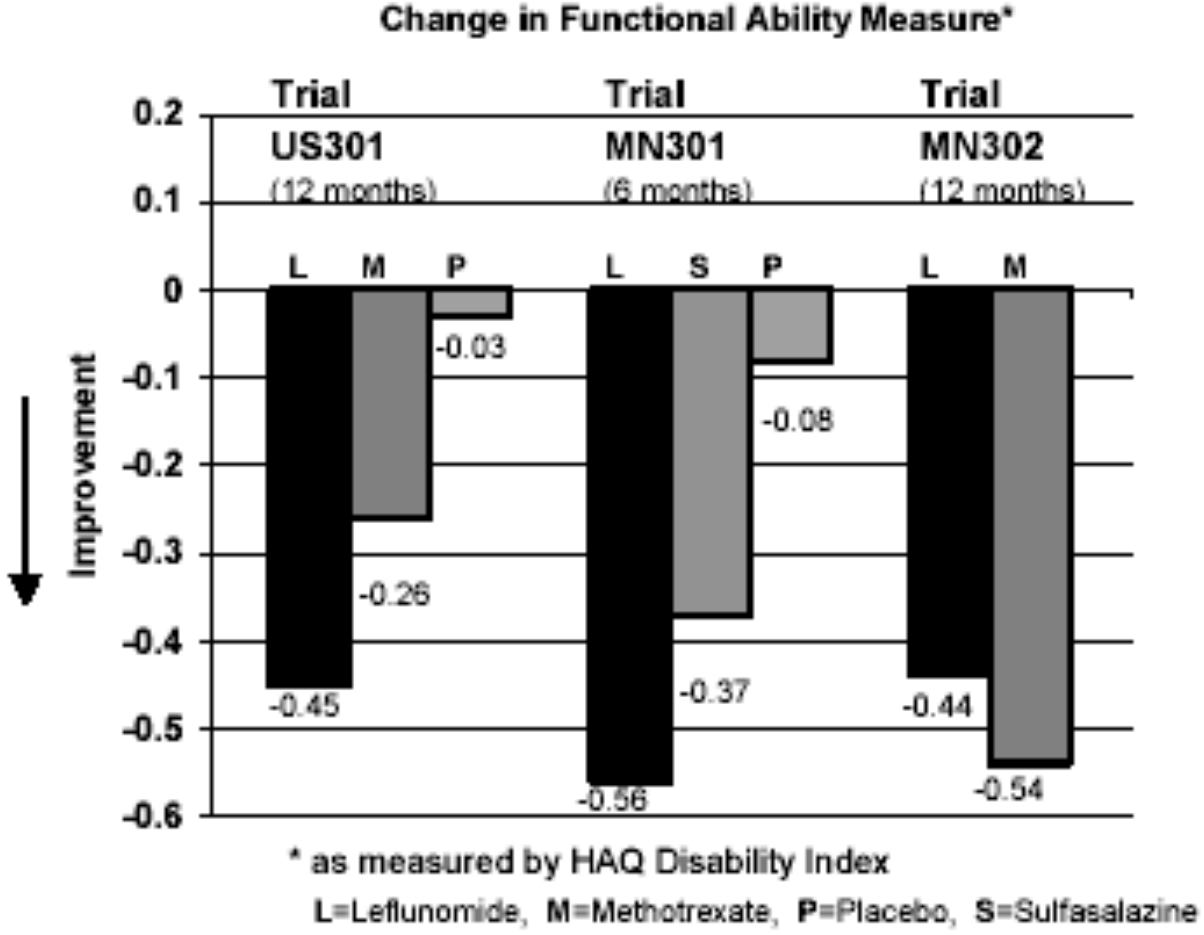
Laurie B. Burke
Associate Director for
Study Endpoints and Labeling Development
Office of New Drugs, CDER, FDA

IMMPACT XIII
June 24-25, 2010

Agenda

- Measurement in the news
  - Comparative effectiveness
  - Drug development “in crisis”
- Addressing assay sensitivity
  - Analgesic trial design issues
  - PRO measurement issues

Leflunomide Improves Physical Function???



Assay Sensitivity

"Whether the purpose of the trial is to show efficacy of the new treatment or to compare two treatments, the question of whether the trial would be capable of distinguishing effective from less effective or ineffective treatments is critical."

…ICH-E10: *Choice of Control Group and Related Issues in Clinical Trials* [www.fda.gov/cder/guidance/4155fnl.htm]

Problems of Active Control Trials

If, for one reason or another, they are poorly run, they may lack assay sensitivity. In that case the study fails. Another study would be needed. When this happens in the NI study, the study succeeds, i.e., it seems to "show" effectiveness.

That is the regulator's problem; we very much do not want to approve an ineffective drug.

Comparative Effectiveness in the News

FDA expects more and more data on the comparative effectiveness of drugs, driven by interest among health systems, payers and patients, and the new funding and research initiative provided under the American Recovery and Reinvestment Act and the health reform law.

Drug Development is "in Crisis"

- Societal expectations for more data (real world)
- Who in the U.S. is responsible for generating this additional evidence?
- What evidentiary standards will be applied?
- The public wants greater certainty about product outcomes
- The public also wants access to innovative treatments
- This is the perfect time to discuss efficiency in drug development and in measurement, in particular

Dr. Woodcock's Ideas for the Future: Enabling Efficient and Effective Drug Development

- Clinical trial infrastructure
- New science for "personalized" medicine
- New science of safety

Clinical Trial Infrastructure

- In the US, biomedical laboratory research extensively supported from 1970s on, while clinical research became more the province of the private sector
- Thus, the US lacks a stable clinical research infrastructure
  - Clinical research personnel and research sites
  - Academic research programs
  - Funded IRB and other human research structures

Problems Caused by Lack of Clinical Research Infrastructure

- Problems with data completeness and integrity
- Lack of standardization of basic procedures
- Difficulties in accomplishing patient follow up resulting in missing data

A US Clinical Research Infrastructure is Needed (cont'd)

- What demands does place on measurements?
  - "Standardized" across trials
  - Applicable in diverse populations, disease severity
  - Reliable (minimal variability)
  - Sensitive to change
  - Readily interpretable
  - Familiar to clinicians and patients

Assay Sensitivity: Study Population Considerations

Study Enrollment Criteria

Pain Palliation

- Evidence of stable baseline pain intensity in order to establish a treatment effect

- Avoid extremes in baseline pain intensity
  - Patients with no or minimum baseline pain, evidence of treatment response not possible;
  - Patients with significant baseline pain require pain management prior to study entry

Study Enrollment Criteria (cont'd)

Pain Progression

- Evidence of no or minimum stable baseline pain intensity in order to establish a treatment effect
- Avoid enrolling patients with significant baseline pain

Baseline analgesic use should be included in determining a patient's baseline pain both in pain palliation and pain progression trials

Special Study Population Concerns of PRO measures

- **Cultural or language subgroups**
  - Adequate translation and cultural adaptation
- **Children and adolescents**
  - Age-appropriate vocabulary and recall period
- **Patients cognitively impaired or unable to communicate**
  - Observer reported outcome preferred over proxy-reported outcome

Observer vs. Proxy Reported Outcome Measures

Observer Reported Outcome Measure

- **Based on observations but not interpretations** of the patient's health condition from the caregiver
- Useful when the patient is unable to convey their subjective state
- Recommended when PRO is not possible

  Example: Parent assessment of frequency of baby crying

Observer vs. Proxy Reported Outcome Measures (cont'd)

Proxy-Reported Outcome Measure

- A report by someone other than the patient reporting as if they were the patient on a symptom known only to the patient
- **Based on observations and interpretations** of the patient's health condition from the caregiver
- Purports to represent the patient's perception but impossible to know whether it actually correlates with patient's perception
- Discouraged by FDA in defining a treatment outcome
  - Example: Asking parents to rate the severity of their infant's pain

Assay Sensitivity:
Study Design Issues
Including frequency of Measurements

Frequency and Timing of Efficacy Assessments

- Appropriate timing of pain/analgesic assessments: (i.e., in relation to analgesia assessment, analgesic administration, time of day, or clinic visit)

Hierarchy of Endpoints

Symptoms vs. Impacts vs. HRQL

Symptoms
(Pain)

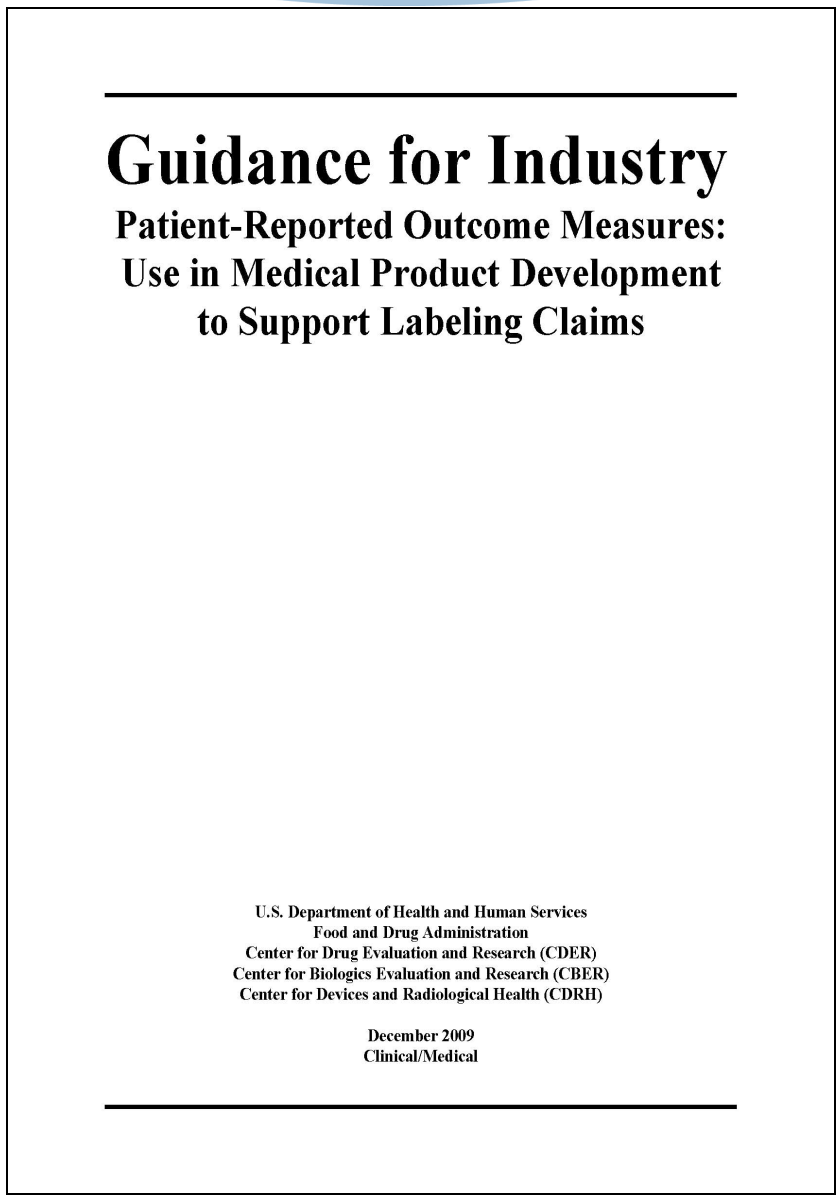
↓

Direct impacts
(i.e. difficulty sleeping)

↓

Less direct impacts (HRQL)
(i.e. visiting friends and family)

Assay Sensitivity: Outcome Measure Considerations



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

Feb 2006: Draft

Dec 2009: Final

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

Modified Wheel & Spokes

i. Hypothesize Conceptual Framework

- Outline hypothesized concepts & potential claims
- Determine intended population
- Determine intended application/characteristics (type of scores, mode & frequency of administration)
- Perform literature/expert review
- Develop hypothesized conceptual framework
- Place PROs within preliminary endpoint model
- Document preliminary instrument development

ii. Adjust Conceptual Framework & Draft Instrument

- Obtain patient input
- Generate new items
- Select recall period, response options & format
- Select mode/method of administration/ data collection
- Conduct patient cognitive interviewing
- Pilot test draft instrument
- Document content validity

iii. Confirm Conceptual Framework & Assess Other Measurement Properties

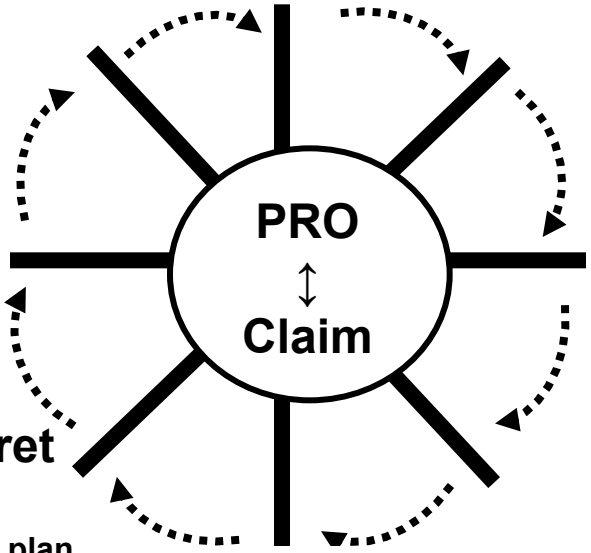
- Confirm conceptual framework with scoring rule
- Assess score reliability, construct validity, & ability to detect change
- Finalize instrument content, formats, scoring, procedures & training materials
- Document measurement development

iv. Collect, Analyze, & Interpret Data

- Prepare protocol & statistical analysis plan (final endpoint model and responder definition)
- Collect & analyze data
- Evaluate treatment response using cumulative distribution & responder definition
- Document interpretation of treatment benefit in relation to claim

v. Modify Instrument

- Change wording of items, populations, response options, recall period, or mode/method of administration/data collection
- Translate & culturally adapt to other languages
- Evaluate modifications as appropriate
- Document all changes



Content Validity

- Extent to which the instrument measures the concept of interest.
- Supported by evidence from qualitative studies that the measure is appropriate and comprehensible, and interpretable, relative to its intended measurement concept, population, and use.
- Other measurement properties are not meaningful without adequate content validity

Content Validity (cont'd)

- Specific to the population, condition, and treatment to be studied.

- Pertains to instructions, item, response options, recall period, data collection method, and instrument administration mode

Content Validity

- Content validity must be established for all measures, including:
  - Pain intensity measure
  - Analgesic measure

Pain Intensity Measure
Appropriate, Comprehensible, and Interpretable?

- Instructions, format, and training
- Item content
  - patient appropriate terminology (cancer related pain vs. bone pain)
  - worst pain vs. average pain during past 24 hours
  - single item vs. multi-item scales
- Response options
  - numeric rating scale vs. Likert scale vs. visual analog scale
- Recall period
  - weekly assessment vs. momentary assessment
- Data collection method
  - paper vs. electronic

Analgesic Measure

Appropriate, Comprehensible, and Interpretable?

- Instructions, format, and training
- Item content
  - Patient-appropriate terminology (i.e. avoid terms such as long or short acting opioids)
- Recording options
  - (i.e. recording box checks versus number of doses)
- Recall period
  - 24 hours
- Data collection method
  - paper vs. electronic

Applying Item Response Theory (IRT) to PRO Measurement

- When applied in the setting of instrument development, provides a tool to ensure that an optimum item set is constructed for the targeted population.
  - Addresses ceiling and floor effect problems
  - Optimizes sensitivity to change

- FDA is developing the statistical expertise to review PRO tools developed using IRT

Responder Burden

- Undue physical, emotional, or cognitive strain on patients, generally
- Decreases the quality and completeness of PRO data and threatens the interpretability of trial results
- Factors contributing to responder burden include:
  - Frequency and timing of assessments
  - Length of questionnaire or interview
  - Formatting/font size
  - Need for physical help for completion of measure

Responder Burden (cont'd)

- Cognitive debriefing can be useful in evaluating the responder burden for a single assessment, however…

- Responder burden can only be fully assessed when implemented in a clinical trial in which consecutive assessments are required

Other Factors that Decrease Assay Sensitivity

- **Inadequate study enrollment criteria**
  - Enrolling patients in a pain palliation trial without measurable baseline pain
  - Enrolling patients in a pain progression trial with significant baseline pain
  - Enrollment of patients with unstable pain
- **Inadequate measurements**
  - Content validity issues
  - Reliability issues
- **Inadequate Study Design**
  - Inappropriate frequency and timing of efficacy assessments (both pain and analgesic)
  - Poorly defined responder definition

Factors that Decrease Assay Sensitivity
(cont'd)

- **Use of medications that interfere with pain palliation response**
  - Poor documentation of concomitant or rescue analgesic use (lack of content validity of analgesic log)

- **Poor Compliance**
  - Toxicity of study drug

Assay Sensitivity: Conclusions

- Anything that affects the size of the treatment effect affects assay sensitivity
- In the superiority trial setting, rigor in measurement improves assay sensitivity because variability is decreased and impact of competing effect modifiers is minimized
- Other trial design considerations are also important to increase sensitivity and decrease variability in the outcome measure