

# Thinking with the End in Mind: From COA Instrument to Endpoint

*SIXTH ANNUAL  
PATIENT-REPORTED OUTCOME CONSORTIUM WORKSHOP*

April 29 - 30, 2015 ■ Silver Spring, MD

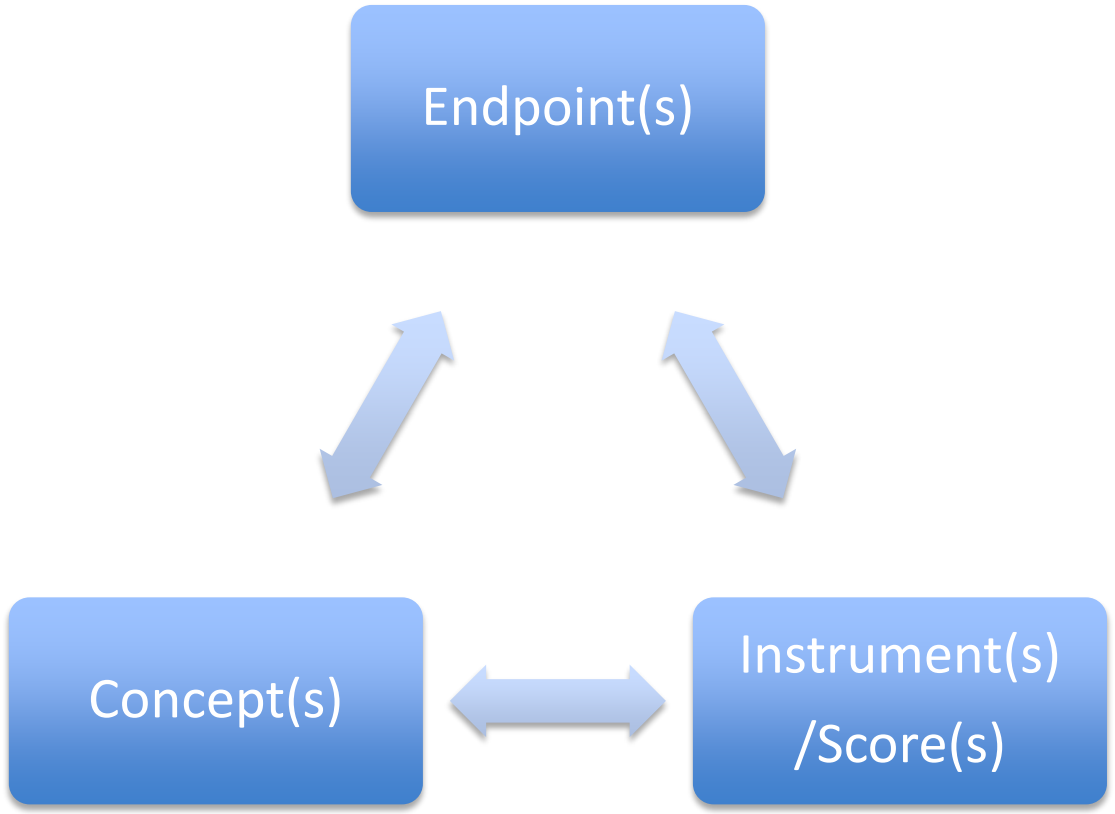


The views and opinions expressed in the following slides are those of the individual presenters and should not be attributed to their respective organizations/companies, the U.S. Food and Drug Administration, the Critical Path Institute, the PRO Consortium, or the ePRO Consortium.

These slides are the intellectual property of the individual presenters and are protected under the copyright laws of the United States of America and other countries. Used by permission. All rights reserved. All trademarks are the property of their respective owners.

- Assessments of symptoms and function are important endpoints to document treatment risk/benefit.
- The objective of the session is to discuss an approach that recommends that the endpoint development occurs alongside the instrument development.

# Concept-Endpoint-Instruments



## **Moderator**

- Jean Paty, PhD – Principal Advisory Services, Quintiles

## **Presenters and Panelists**

- Paul G. Kluetz, MD – Acting Deputy Director, Office of Hematology and Oncology Products, FDA
- David S. Reasner, PhD – Vice President, Data Science and Head, Study Endpoints, Ironwood Pharmaceuticals
- Laura Lee Johnson PhD – Associate Director, Division of Biometrics III, Office of Biostatistics, Office of Translational Sciences, FDA
- Elisabeth Piau-Louis, PharmD, MA – Principal Outcomes Research Scientist, OncologyGenentech

# Endpoints in Cancer Clinical Trials

**Paul Kluetz, M.D.**

**Office of Hematology and Oncology Products  
U.S. FDA**

***SIXTH ANNUAL***

***PATIENT-REPORTED OUTCOME CONSORTIUM WORKSHOP***

**April 29 - 30, 2015 ■ Silver Spring, MD**

# Why is PRO so Foreign to Oncologists as an Endpoint?



- We treat life-threatening diseases where median survival can be less than one year
  - Overall Survival- Gold Standard Endpoint
- We can Visualize our Disease and Objectively Quantify a Drug's Effect on the Tumor
  - Radiographic response and time to progression
- Oncologists and Statisticians have standard definitions and analyses for these endpoints!

# But OS and PFS Aren't the Whole Story...



- There is broad consensus that accurately describing a patient's experience is important to patients and physicians
- There has been difficulty in agreement on the optimal instruments, collection and interpretation of PRO data
- Efforts are underway to **OPTIMIZE** and **STANDARDIZE** PRO in oncology trials



# Which Questions Can PRO Answer?



- **Efficacy:** Does the drug provide superior improvement in disease related symptoms or functional deficits?
  - Pain, Total Symptom Score, Performance related outcomes
  - More conducive to formal statistical analysis (statistical superiority)
- **What do patients experience** while on therapy?
  - Adverse events from therapy (PRO-CTCAE?)
  - Physical function / Performance status
  - Likely to be more descriptive in nature

# For This Panel: I Will Focus on Efficacy



- **Efficacy:** Does the drug provide superior improvement in disease related symptoms or functional deficits?
  - Pain, Total Symptom Score, Performance related outcomes
  - More conducive to formal statistical analysis (statistical superiority)
- What do patients experience while on therapy?
  - Adverse events from therapy (PRO-CTCAE?)
  - Physical function / Performance status
  - Likely to be more descriptive in nature

# PRO Efficacy Endpoints: What is Important to the Clinical Reviewers at FDA?



- FDA has reviewed the instrument and feels that it is adequate (not perfect, but fit for purpose)
- Questions asked make clinical sense in evaluating efficacy
- Sponsor has provided data to support the rationale for the meaningful difference that creates the threshold for the endpoint event (improved pain, deterioration in function, etc.)
- There is a hypothesis and it is TESTED.
  - There is a pre-specified statistical analysis plan for how the data will be analyzed

# When an Instrument is Not the Endpoint



- Brief Pain Inventory- Short Form
  - Instrument BPI-SF: Has many questions
  - But the endpoint could be based on the single BPI-SF item #3: worst pain over past 24 hours.
  
- EORTC QLQ-C30
  - Has many questions and multiple domains
  - Could we pre-specify Time to Functional Deterioration as the Endpoint?

# Also: The Endpoint is MORE than Just the Question Being Asked

## CONCEPT: PAIN

### Instrument:

Brief Pain Inventory- Short Form (15 questions)

### Endpoint:

**PAIN PALLIATION** using Item #3: **Worse Pain in past 24 hours**

Definition includes clinically meaningful event threshold, concomitant analgesic use, whether confirmation on next assessment required, etc.

# When an Instrument is Not the Endpoint- Added Rationale is Needed



- If the intent is to use a portion of an instrument as your endpoint, strong rationale will need to be made for why this will still be a well-defined and reliable assessment.
  - Measurement properties of the item/domain, type of analysis (time to event, responder analysis), endpoint definition including threshold for response or deterioration, concomitant medications, etc.
- It can be done, but early consultation with the review division and SEALD is urged if you intend to consider this strategy

# From Instrument to Endpoint



- Jakafi PRO endpoint: Improvement in total symptom score (consisting of 6 items).

**Table 8: Improvement in Total Symptom Score**

	<b>Jakafi (N=148)</b>	<b>Placebo (N=152)</b>
Number (%) of Patients with 50% or Greater Reduction in Total Symptom Score by Week 24	68 (45.9)	8 (5.3)
P-value	< 0.0001	

- Jakafi did not statistically significantly reduce the severity of any of the 6 individual symptoms
  - Those hypotheses weren't statistically tested.
  - The individual symptoms were **DESCRIPTIVELY** analyzed to show there was no one symptom driving the result

# If you want to show a reduction in something... it has to exist at baseline



## For our Common Radiographic Endpoints:

- Tumor at Baseline: Either measure tumor response, or measure the time it takes to grow (progression)
  - Endpoint: Objective Response Rate (responder analysis)
    - Definition: 30% reduction in tumor measure
  - Endpoint: Progression Free Survival (time to event)
    - Definition: 20% increase in tumor measure OR death from any cause
- If you have NO tumor, you cannot assess a response rate and must use a time to appearance of disease
  - Endpoint: Disease Free Survival (time to event)
    - Definition: Radiographic appearance of disease or Death



# PRO Endpoints are Not Different



- Symptomatic Patients = > Symptom Improvement/  
Symptom Palliation
- Minimally Symptomatic Patients = > Time to  
Symptom Deterioration

# Palliation or Time to Deterioration?



- Successful oncology PRO labeling claims have been palliation endpoints with responder definitions conducted in blinded randomized clinical trials
  - Pain with Bone Seeking Radioisotopes
  - Total Symptom Score with Jakafi
- Less experience with Time to Deterioration (TTD) analyses which have specific challenges
  - Some have radiographic progression before symptoms worsen
  - Would require follow up beyond progression
  - Likely to be affected by subsequent therapies
  - Higher risk of missing data
- TTD can be done and can be supportive, but work must be done to optimize these types of analyses

# Must We Assess ALL Endpoints in One Single Trial?



- The best population to study a drug's effect on disease related symptoms is a Symptomatic Patient
- Many products are being tested in asymptomatic or minimally symptomatic populations
- **While TTD analyses can be done, consider a second smaller trial enriching for symptomatic patients and optimizing design for PRO disease symptom palliation primary endpoint**

## *What should we all be worried about?*

- A PRO result is labeled that is significantly biased and/or is NOT clinically meaningful to patients (**false positive**)
- Worst Case Scenario 1: False Positive PRO is the sole primary endpoint of a new drug or biologic anti-cancer product.
  - Risk: Ineffective drug is marketed to patients with toxicity and no benefit (toxic placebo)
- Worst Case Scenario 2: False Positive PRO is labelled, but drug has demonstrated effect on tumor based on PFS or OS:
  - Risk: Patients and physicians may choose the therapy over another more effective therapy (that does not have a labeled PRO benefit)

# How Can We Reduce This Risk?

## Optimize and Standardize PRO



- Core Concepts to Measure
- Narrow Group of Instruments for each Concept
- Clear Endpoints with Meaningful Definitions
- Clinical trial design- assessments, data collection
- Data analysis: Primary and Sensitivity Analyses
- Data presentation in Manuscripts and FDA Label

# Linking Assessments & Endpoints – Measurement Properties

**David S. Reasner, Ph.D.**  
**Ironwood Pharmaceuticals**

***SIXTH ANNUAL***  
***PATIENT-REPORTED OUTCOME CONSORTIUM WORKSHOP***

**April 29 - 30, 2015 ■ Silver Spring, MD**

# Assessments & Endpoints: Measurement Properties

## Goal

- Treatment of Disease X or
- Treatment of the Symptoms of Disease Y

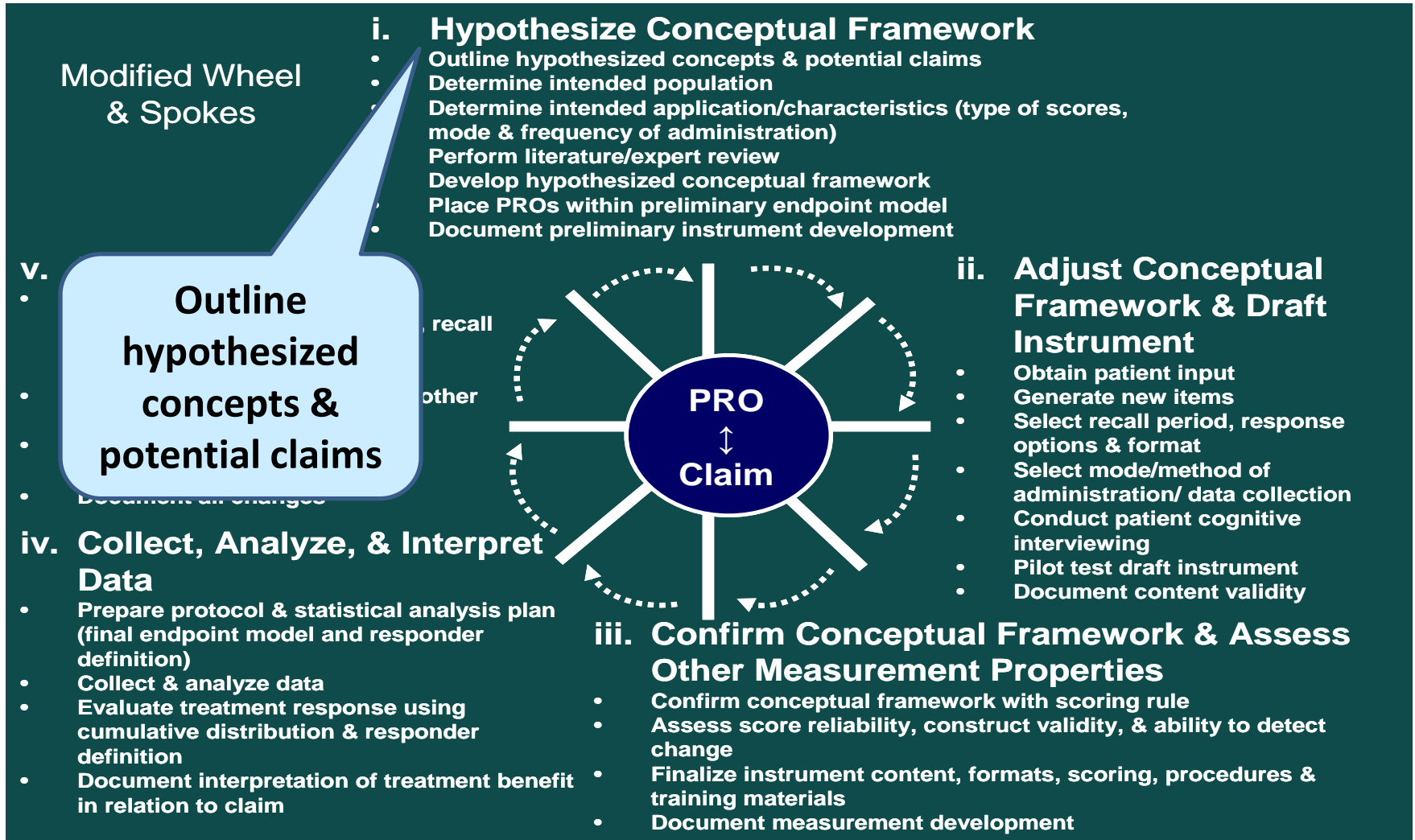
## Path Along the Way

- Patient as **the phenomenon**
- Patient as **the “instrument”**
- Psychometric **metadata**
- Well-defined and reliable assessments
- Clinically meaningful endpoints (i.e., relative improvement)



“By explicitly addressing the review issues identified in this guidance, sponsors can [increase the efficiency](#) of their discussions with the FDA during the medical product development process, [streamline the FDA’s review](#) of PRO instrument adequacy and resultant PRO data collected during a clinical trial, and [provide optimal information about the patient perspective](#) for use in making conclusions about treatment effect at the time of medical product approval.”





## Sections

- I. Instrument
- II. Targeted Claims or TPP [Target Product Profile]
- III. Endpoint Model
- IV. The PRO Instrument's Conceptual Framework
- V. Content Validity Documentation
- VI. Assessment of Other Measurement Properties
- VII. Interpretation of Scores
- VIII. Language Translation and Cultural Adaptation
- IX. Data Collection Method
- X. Modifications
- XI. PRO-Specific Plans Related to Clinical Trial Design and Data Analysis
- XII. Key References

Appendix A – User Manual

Appendix B – Item Tracking Matrix

Appendix C – Transcripts

*PRO Guidance Appendix – The “Dossier” [eCTD 5.3.5.3]*

# Target Product Profile: A Strategic Development Process Tool



## Key Section List

1. **Indications and Usage**
2. Dosage and Administration
3. Dosage Forms and Strengths
4. Contraindications
5. Warnings and Precautions
6. Adverse Reactions
7. Drug Interactions
8. Use in Specific Populations
9. Drug Abuse and Dependence
10. Overdosage
11. Description
12. Clinical Pharmacology
13. Nonclinical Toxicology
14. **Clinical Studies**
15. References
16. How Supplied/Storage and Handling
17. Patient Counseling Information

Characteristics of PRO instruments that are reviewed:

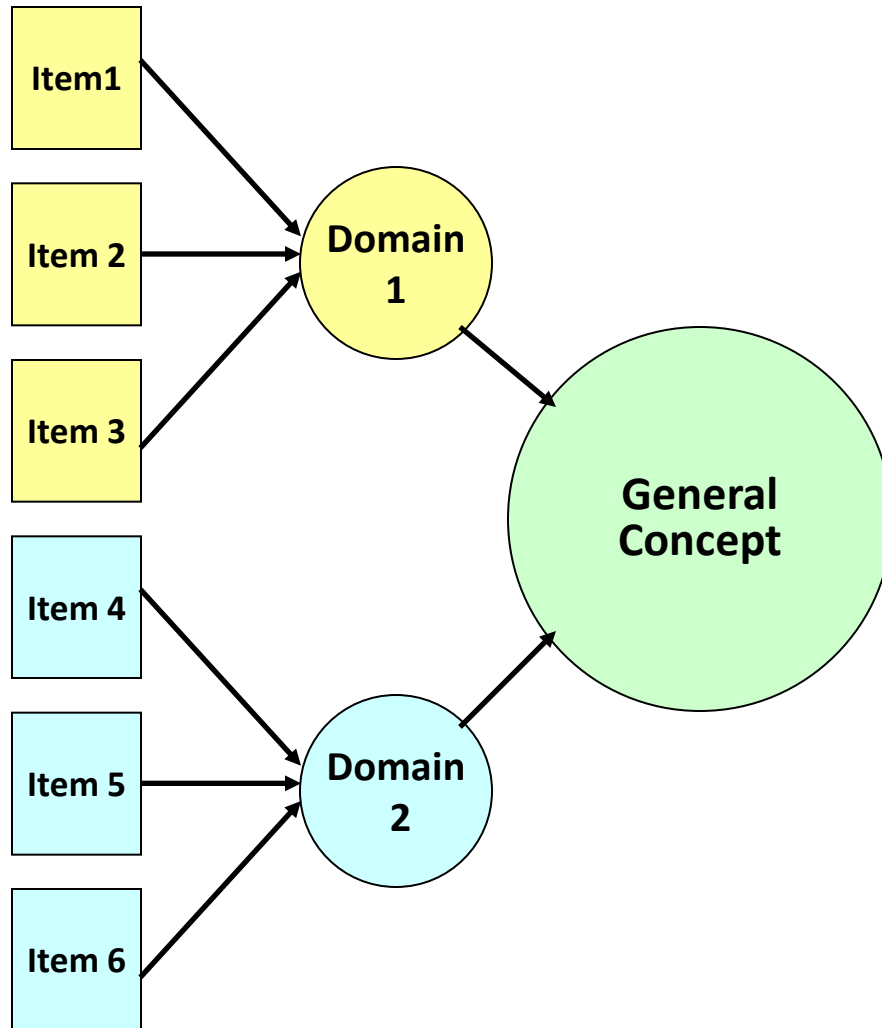
- Concepts as elicited from patients
- Conceptual framework as documented in the patient population and in the words of the patients
- [Content validity \(i.e., items\) and other measurement properties](#)
- Administration mode as appropriate to the patient population
- [Scoring](#) of the questionnaire to create instrument
- Collection of [analysis variables](#) and [derivation of analysis endpoints](#)

Note: The FDA will review documentation of PRO instrument development and testing [in conjunction with clinical trial results](#) to determine whether a labeling claim is substantiated.

A few definitions related to measurement properties:

- Content Validity: Evidence that the instrument measures the concept of interest including evidence from qualitative studies that the items and domains of an instrument are appropriate and comprehensive [i.e., saturation] relative to its intended measurement concept. Testing other measurement properties will not replace or rectify problems with content validity.
  - Necessary but not sufficient
- Construct Validity: Evidence that relationships among items, domains, and concepts conform to *a priori* hypotheses concerning logical relationships that should exist with measures of related concepts or scores produced in similar or diverse patient groups (e.g., discriminant and convergent validity).
  - Evidence based on relations to other endpoints including predictive validity

# Conceptual Framework

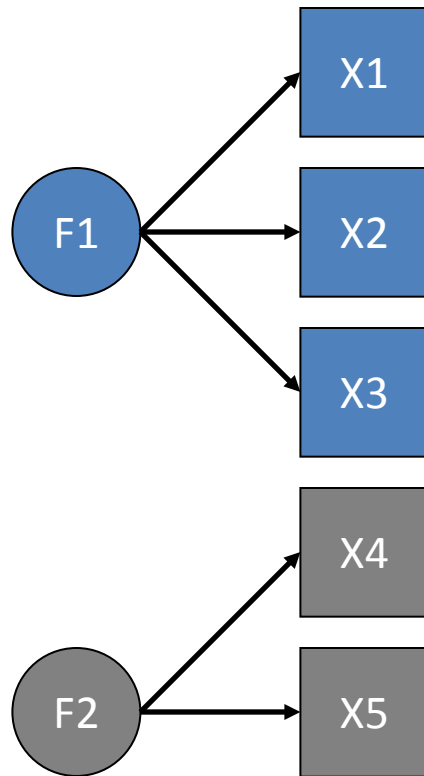


The conceptual framework of a PRO instrument may be straightforward if a single item is a reliable and valid measure of the concept of interest (e.g., pain intensity).

However, single-item measures of general concepts that include multiple items or domains rarely provide sufficient evidence to support claims about that general concept (e.g., a global question concerning a functional disorder defined by clusters of specific signs and symptoms).

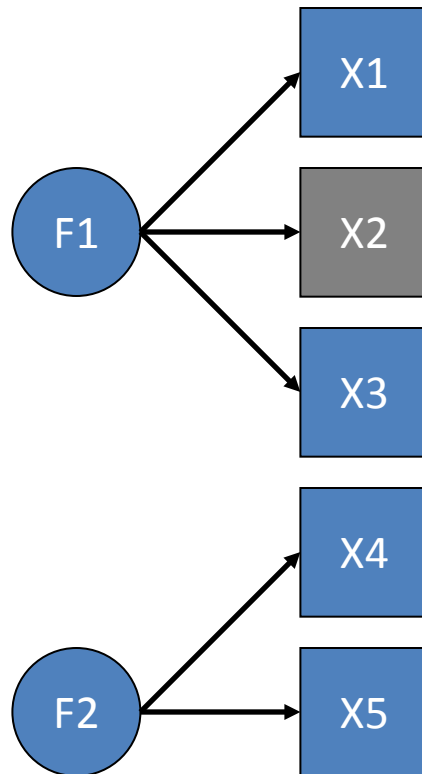
# Open Questions: Measurement Properties

Are the construct(s) underlying the patient-reported outcome appropriate to the targeted mechanism of action?



# Open Questions: Measurement Properties

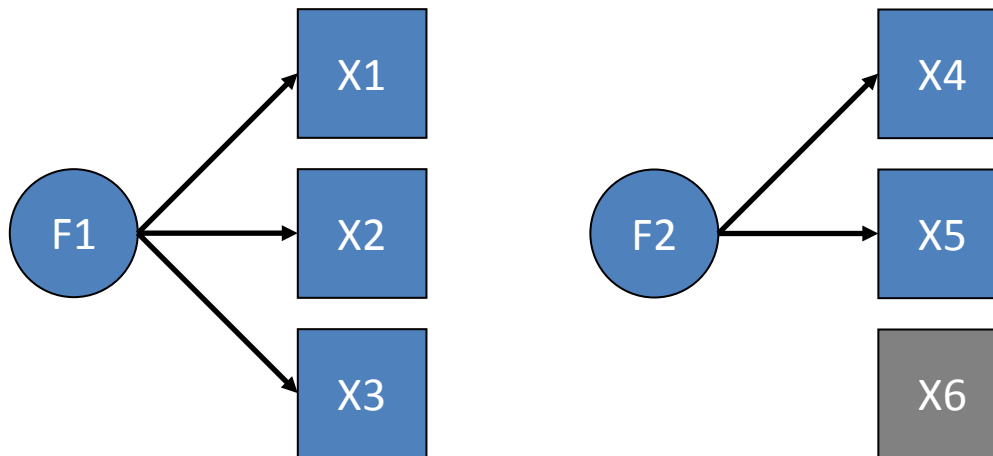
Does the patient in the context of use attend to all the available symptoms?





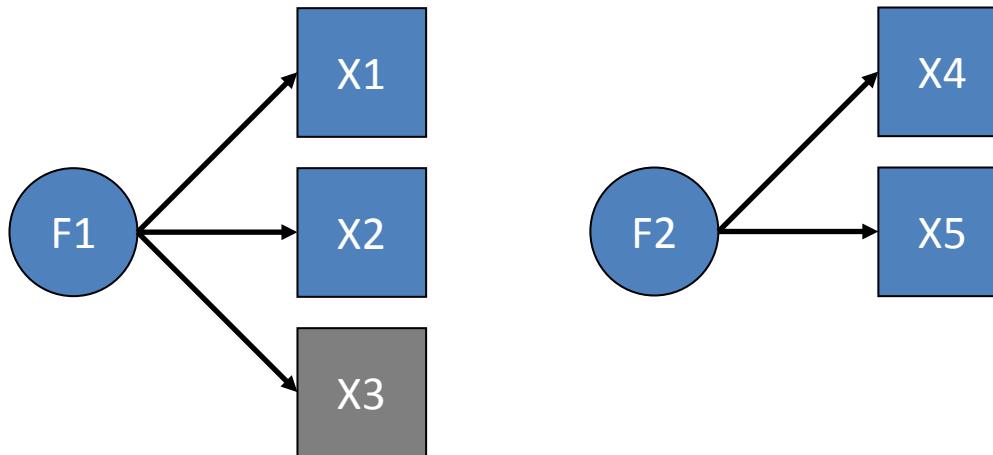
# Open Questions: Measurement Properties

Do clinicians reference signs and symptoms outside of the content-derived conceptual framework?



# Open Questions: Measurement Properties

Do clinicians reference all the signs and symptoms **inside** of the content-derived conceptual framework?



# Irritable Bowel Syndrome

## Clinical Evaluation of Drugs for Treatment



- The goal for efficacy endpoints in IBS clinical trials is to assess the treatment effect on the **core disease-defining signs and symptoms** of IBS in a well-defined and reliable way.
- The PRO measure(s) should capture all of the **clinically important signs and symptoms** of the IBS target population (e.g., IBS-C or IBS-D).

# Irritable Bowel Syndrome

## Clinical Evaluation of Drugs for Treatment



### Abdominal Pain

If a drug is developed specifically to improve only one of the major signs or symptoms of IBS (e.g., abdominal pain), based on the drug's mechanism of action, it is still important to assess the other important signs and symptoms to document that the drug has not negatively affected those components.

Daily Diary	How would you rate your <u>abdominal pain</u> at its worst over the last 24 hours? Enter a number from 0 to 10, where 0 represents no abdominal pain and 10 represents very severe abdominal pain.	Push button with numbers shown in buttons	Range: From 0 to 10, incrementing by 1  Left Label: No abdominal Pain  Right Label: Very severe abdominal pain
----------------	--	---	---

# Irritable Bowel Syndrome

## Clinical Evaluation of Drugs for Treatment



### Stool Frequency (Six different questions – C|S)

For IBS-C, the defecation component of the proposed primary endpoint can be evaluated by assessing stool frequency.

<p>Bowel Movement</p>	<p>Do you have a bowel movement to report since X?</p> <p><i>Note = Y = If phase = Pre-Treatment</i></p> <p><i>Y = the most recent date/time of the following 00:00 day of assignment , <u>or</u> when subject records No more BM, completion date/time of most recent BM Diary completed in Pre-treatment <u>or</u> 00:00 of the previous day</i></p> <p><i>If phase = Treatment or Post-Treatment</i></p> <p><i>Y = the most recent date/time of the following: Completion date/time of Eligibility Review Report, <u>or</u> when subject record no more BM, completion date/time of most recent BM Diary completed in Treatment or Post-Treatment phase <u>or</u> 00:00 of the previous day</i></p> <p><i>Note business rule #.#.# applies</i></p> <p><i>Display all date/times as DD Mmm YYYY HH:mm</i></p>	<p>Multiple choice (choose one)</p> <p>Offset = 1</p> <p>Ascending values</p> <p>Solid button</p>	<p>Yes = 1</p> <p><i>If yes next screen</i></p> <p>No = 2</p> <p><i>If no go to 1<sup>st</sup> question daily diary</i></p>
-----------------------	---	---	---

# Irritable Bowel Syndrome

## Clinical Evaluation of Drugs for Treatment



### Provisional Endpoints for IBS Clinical Trials:

- Weekly Stool frequency, as measured by the number of complete spontaneous bowel movements (CSBMs) per week
- Weekly Percent Change from Baseline in Abdominal pain intensity, as measured by a numeric rating scale (i.e., 0 to 10) that asks patients to rate their worst abdominal pain
- Overall Responder: A responder in greater than 50% of the treatment period weeks where response is both a decrease in abdominal pain intensity of at least 30% and an increase of at least 1 CSBM per week, both relative to baseline

# PRO Instrument Development & Modification: Endpoint Model

(final endpoint model and responder definition)

Place PROs within preliminary endpoint model

## 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

## 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

## 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



## 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

# Endpoint Model: Parallel Development



- “Sponsors should **define the role a PRO endpoint is intended to play** in the clinical trial (i.e., a primary, key secondary, or exploratory endpoint) so that the instrument development and performance can be reviewed in the context of the intended role, and appropriate statistical methods can be planned and applied. It is critical to plan these approaches in what can be called an endpoint model.
- “PRO instrument adequacy depends on its role as depicted in the endpoint model. **The endpoint model explains the exact demands placed on the PRO instrument** to attain the evidence to meet the clinical trial objectives and support the targeted claims corresponding to the concepts measured.”








# Endpoint Model: Planned Analyses & Hierarchy





- “We intend to determine the adequacy of clinical trial data to support claims in light of the **prespecified method** for endpoint analysis. We usually view unplanned or *post hoc* statistical analyses conducted **after unblinding as exploratory** and, therefore, unable to serve as the basis of a labeling claim of effectiveness.”
- “**A single hierarchy of endpoints** as diagrammed in an endpoint model (see Figures 1 and 2 in section III.A., Endpoint Model) is determined by the trial’s stated objectives and the clinical relevance and importance of each specific measure independently and in relationship to each other.”

# Endpoint Model: Treatment of Disease X

Concept	Link	Endpoint
Indication:		Primary:
Treatment of Disease X		Physiological Effect
Supportive Concepts:		Secondary:
Improvement in Symptoms/Signs of Disease X	   	Symptoms Diary [PRO] Signs Diary [PRO] Physical Exam Physical Performance [possibly PRO]

# PRO Claim with an Indication

Concept	Link	Endpoint
Indication:		Primary:
Treatment of Symptoms of Disease Y		Total Disease Y Symptoms Score [PRO]
Supportive Concepts:		Secondary:
Other Treatment Benefit		Physical Performance [Possible PRO] Disease Y-related Physical Limitations [PRO]

- Develop aspirational TPP, identify potential claims, and propose preliminary endpoint model
- Build a conceptual framework that supports the aspirational TPP
- Develop proposed scoring (items|instrument) that corresponds to conceptual framework
- Collect psychometric metadata and clinical data (e.g., Phase 2b trials)
- Align the conceptual framework, scoring, and potential claims
- Propose analysis endpoints that capture clinical meaning (e.g., End of Phase 2 milestone meeting)

**You Have a Measure.  
You Have a Tool.  
Do You Have an Endpoint for the Study?**

**Laura Lee Johnson, Ph.D.  
Division of Biometrics II, Office of Biostatistics  
U.S. FDA**

***SIXTH ANNUAL  
PATIENT-REPORTED OUTCOME CONSORTIUM WORKSHOP***

**April 29 - 30, 2015 ■ Silver Spring, MD**

- Measures
- Endpoint usually uses some summary
  - Time until an event defined by the measure
  - Difference in scores (e.g. Week 12 score – Baseline score)
    - Weekly average, but have a daily diary...but
  - A million other ways to make an endpoint
    - Some make sense and can be analyzed and interpreted on label, other endpoints are not as clear

# Responder Definition ≠ Responder Analysis



- To use or not to use a responder definition as a part of the endpoint
- Impact on power and analysis methods
- What if the responder definition changes?
- “Responder” sounds nice, but is there data to say that the status change matters clinically
  - Comparing % responded by a certain time?
  - Time until become a responder?
    - Can responder status change multiple times?
  - Something else?

- Interpretation and analysis issues
  - True for PROs, PET scans, etc
- What is the difference to plan power and sample size in a trial
  - Interim analyses and how to interpret



# Instrument is EVERYTHING Not Only Items and Responses and Instructions



- User Manuals
  - How the instrument should be scored
  - How to we handle missing items, missing scores, if people go off study, if people die
  - When should it be collected?
  - When should data not be used?
  - Many other issues, but more will be needed in the SOPs, SAP, protocol, etc.

(picture of all the documents that came with my house, all the manuals etc.)

- Still did not tell me who to contact about electricity etc.
- When study team members, patients, and others use a tool or an ePRO you are thinking of the User Manual. Also think of it for the study designer, data manager, analyst.

# Statistical Issues That May Affect Results: Multiplicity & Non-Inferiority



- Type I error inflation due to multiplicity
  - What is (are) the score(s) being used?
  - Is this a single score, composite endpoint, something else?
  - Looking at items and a summary score?
- Non-Inferiority
  - If the study design is for non-inferiority or equivalence, how is that margin or bound set for the endpoint?

# Statistical Issues That May Affect Results: Missing Data & Subgroups



- Missing data
  - High percentage of dropouts
    - Is it the burden of the COA? The ePRO? Or something else?
  - Inappropriate imputation for missing values
    - Appropriate should take in to account tool development for example IRT or CTT?
- Inconsistency of results across subgroups
  - Is it the interventions or the measure?

- When working on protocols and when asking FDA for advice, keep in mind our Statistical Reviews include information on the following
  - Breaking the blind: can this happen by accident?
  - Unblinded or unplanned interim analyses
  - Change of primary endpoint during conduct of the trial
    - Dropping items?
    - Linking studies?
  - Dropping/adding treatment arms
  - Sample size modification
  - Planned and unplanned study design adaptations
    - Usually this means to randomization, etc.

- Talk to Regulators
- Talk to the statisticians who will design and analyze the data\*
  - Many will say number? Ok...and not ask questions about the PRO itself. Be careful and pro-active!
- Start thinking in early development of the COA how it will be used in randomized longitudinal clinical trials
  - Every validation study
  - Qualitative work
  - Every step needs to keep the end in mind

# Discussion

## **Moderator**

- Jean Paty, PhD – Principal Advisory Services, Quintiles

## **Presenters and Panelists**

- Paul G. Kluetz, MD – Acting Deputy Director, Office of Hematology and Oncology Products, FDA
- David S. Reasner, PhD – Vice President, Data Science and Head, Study Endpoints, Ironwood Pharmaceuticals
- Laura Lee Johnson PhD – Associate Director, Division of Biometrics III, Office of Biostatistics, Office of Translational Sciences, FDA
- Elisabeth Piau-Louis, PharmD, MA – Principal Outcomes Research Scientist, OncologyGenentech

# Irritable Bowel Syndrome-Quality of Life Measure - (IBS-QOL)



Endpoint (Subscale)	Items	Lowest & Highest Possible Scores	Possible Score Range
Dysphoria (DY)	1, 6, 7, 9, 10, 13, 16, 30	8, 40	32
Interference With Activity (IN)	3, 18, 19, 22, 27, 29, 31	7, 35	28
Body Image (BI)	5, 21, 25, 26	4, 20	16
Health Worry (HW)	4, 15, 32	3, 15	12
Food Avoidance (FA)	11, 23, 28	3, 15	12
Social Reaction (SR)	2, 14, 17, 34	4, 20	16
Sexual (SX)	12, 20	2, 10	8
Relationships (RL)	8, 24, 33	3, 15	12
Overall (OV)	All items	34, 170	136



# Irritable Bowel Syndrome-Quality of Life Measure - (IBS-QOL)



- Subscales are scored through simple summative scaling.
- All items are negatively framed (on a 1-5 point scale), with the highest response scale equaling the worst quality of life.
- When scored, all items are reversed so that, as IBS-QOL scores increase, quality of life increases.
- All final raw scores are transformed to a 0 to 100 scale using the following formula:

$$\textit{Scale Score} = \frac{\textit{sum of the items} - \textit{lowest score}}{\textit{possible raw sum range}} \times 100$$