Qualification of Treatment Benefit Measurements

IMMPACT-XV

Laurie Burke, Director
Study Endpoints and Labeling Development Staff
Office of New Drugs, CDER, FDA

22 June 2012 – Washington DC

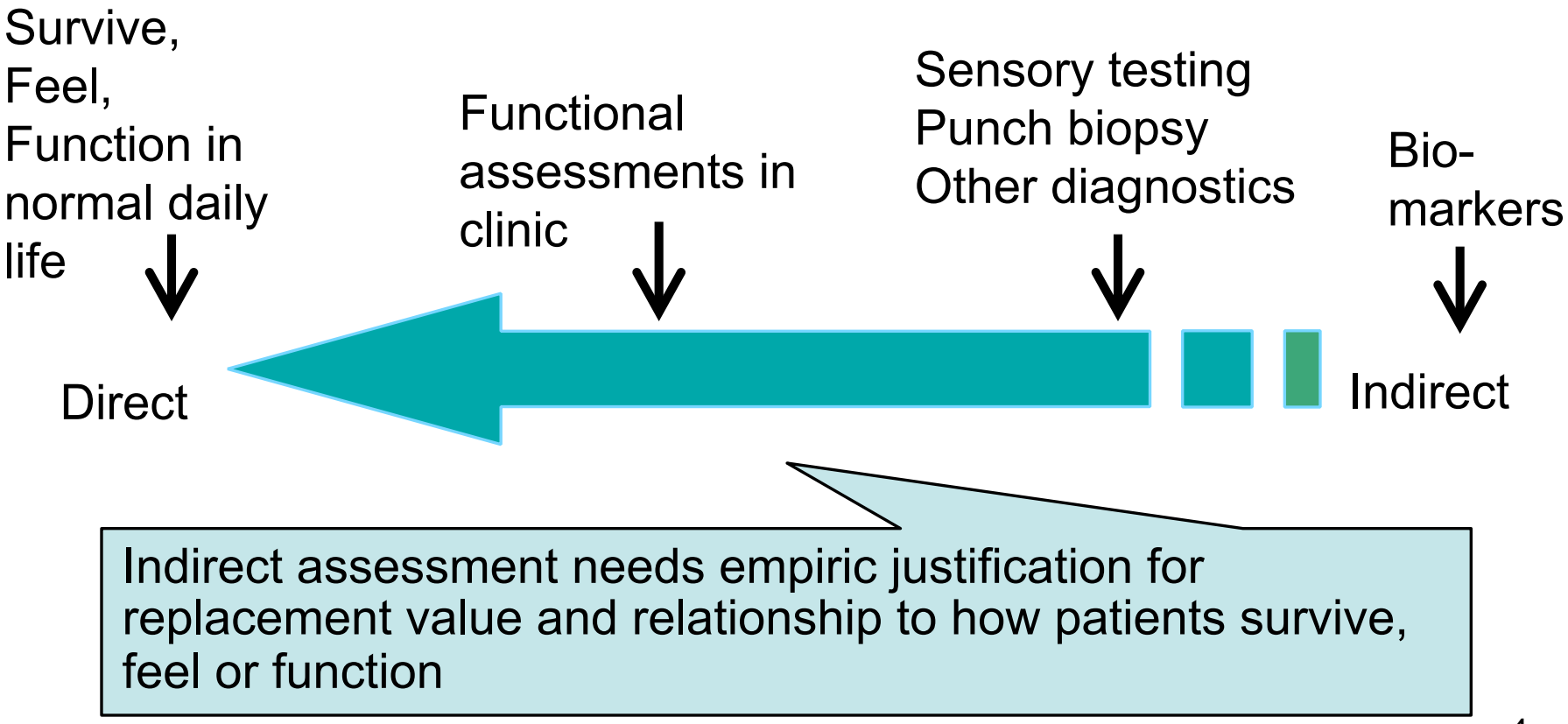
Overview

- Treatment benefit
- Treatment benefit measurements
- Planning for measurement
  - Context of use
  - Concept of measurement
- Evidence for well-defined and reliable measurement
  - Content validity
  - Other measurement properties
  - Relationship between an indirect measurement and treatment benefit

Treatment Benefit

- The impact of treatment on how a patient "survives, feels, or functions" in daily life
  - Measured as effectiveness or comparative safety
  - Used interchangeably with "clinical benefit"
  - Can be measured **directly** or **indirectly**
- Described in labeling or promotion as a claim that describes the benefit measured in the context of use defined by the study protocol

Direct versus Indirect Evidence of Treatment Benefit



Ref: Fleming T, COA Workshop, FDA, Oct 2011

What is "Indirect" Evidence of Treatment Benefit?

- The concept being measured is different from the directly meaningful concept—it's a replacement
- "Indirectness" is relative and permanent
- FDA reviews evidence of the relationship between the indirect concept and how patients survive, feel or function
  - Generally requires longitudinal studies to demonstrate that relationship
  - Often described as evidence of clinical meaningfulness

PRO, ClinRO and ObsRO Assessments

- All influenced by human choices
  - Conscious or unconscious
- Clinician, other observer, or patient rater
  - Judgment, cooperation, motivation involved
- Only patients can assess symptoms **directly**
- Most ClinROs and ObsROs **indirectly** assess "feels or functions" (as a replacement for direct assessment)
- Indirect assessment cannot be avoided because
  - Not all patients can rate themselves
  - Not all functioning can feasibly be measured directly

Patient-reported outcome (PRO) assessment

- An assessment based on a report that comes directly from the patient without amendment or interpretation about the status of particular aspects of or events related to a patient's health condition
- Only PRO assessments can measure symptoms
- PRO assessments can measure signs or other indirect evidence of treatment benefit
- Examples:
  - 0-10 NRS of pain intensity
  - PF-10 of lower limb functioning

Clinician-reported outcome (ClinRO) assessment

- An assessment made by an observer with some recognized professional training that is relevant to the assessment
- May include an evaluation and interpretation of the patient's condition based on clinical judgment.
- Examples include:
  - Subjective impression following physical examination
  - Physical performance measures
  - Quantitative sensory testing

Observer-reported outcome (ObsRO) assessment

- An assessment that is determined by an observer who does not necessarily have a relevant background of professional training
- Often used when the patient is unable to self-report (e.g., infants, young children, cognitively incapacitated)
- Used to capture observable concepts only (e.g., signs or behaviors—NOT symptoms)
- Examples include:
  - Parent report of infant behavior
  - Caregiver report of patient observations

Development of a Treatment Benefit Measurement

Step 1: Define context of use

Step 2: Define concept(s) of measurement and reporter (direct or indirect)

Observable Concept

No Clinical Judgment Needed

Self-report?

No → **ObsRO**

Yes → **PRO**

Clinical Administration or Judgment → **ClinRO**

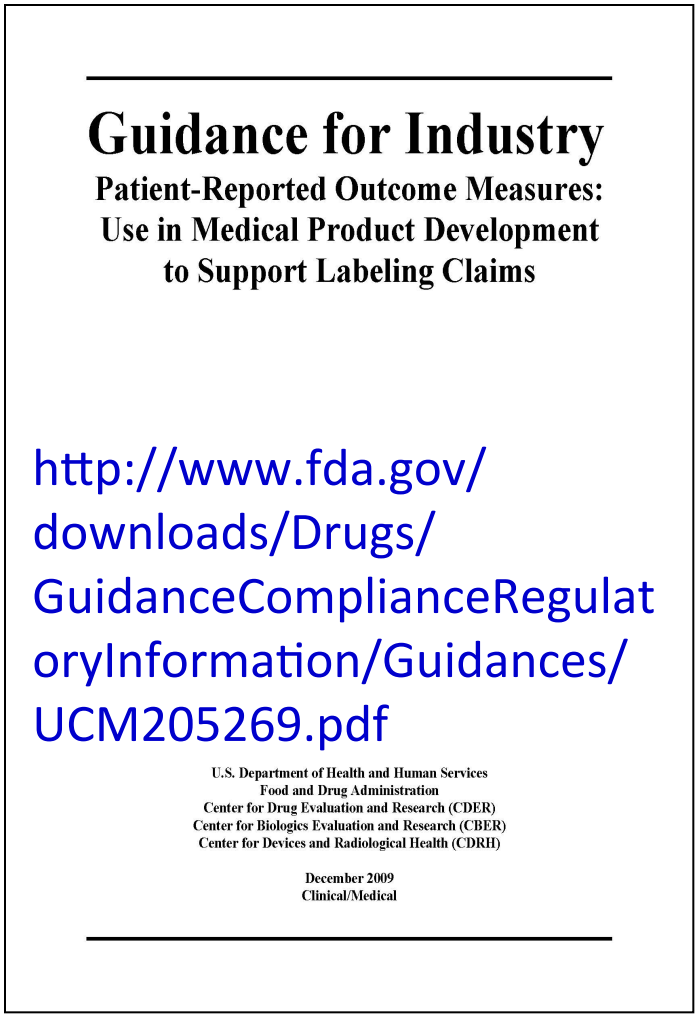
Non-Observable Concept → **PRO**

Physiologic or lab findings that can be measured without human assessment → **Biomarker**

Step 3: Document evidence that the measurement is a valid and reliable assessment of the concept of measurement in the context of use

Step 4: If indirect, document evidence that the measurement is an adequate replacement for how patients feel or function in their daily lives

Review of PRO, ClinRO, and ObsRO Measurements



Guidance for Industry

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

http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM205269.pdf

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

- Defines how the Agency interprets "well-defined and reliable" (21 CFR 314.126) for PRO measures intended to provide evidence of treatment benefit
- Summarizes good measurement principles applicable to any PRO, ClinRO or ObsRO assessment

What Are the Elements of Context of Use for COAs?

- Outcome measurement (not a diagnostic or screening CoU)
- Disease definition
  - Explicit and specific to targeted clinical trial population
  - Matches the inclusion/exclusion criteria
    - Disease severity
    - Demographics
    - Other important aspects of heterogeneity
  - More detailed than diagnostic or stratification criteria
  - May vary by subgroup (e.g., age)
- Clinical setting (e.g., inpatient vs. outpatient)
- General plan for study design (study objectives, endpoint model)
- General plan for data interpretation
- Targeted labeling claims (consistent with the concept of measurement)

Context of Use: Endpoint Model

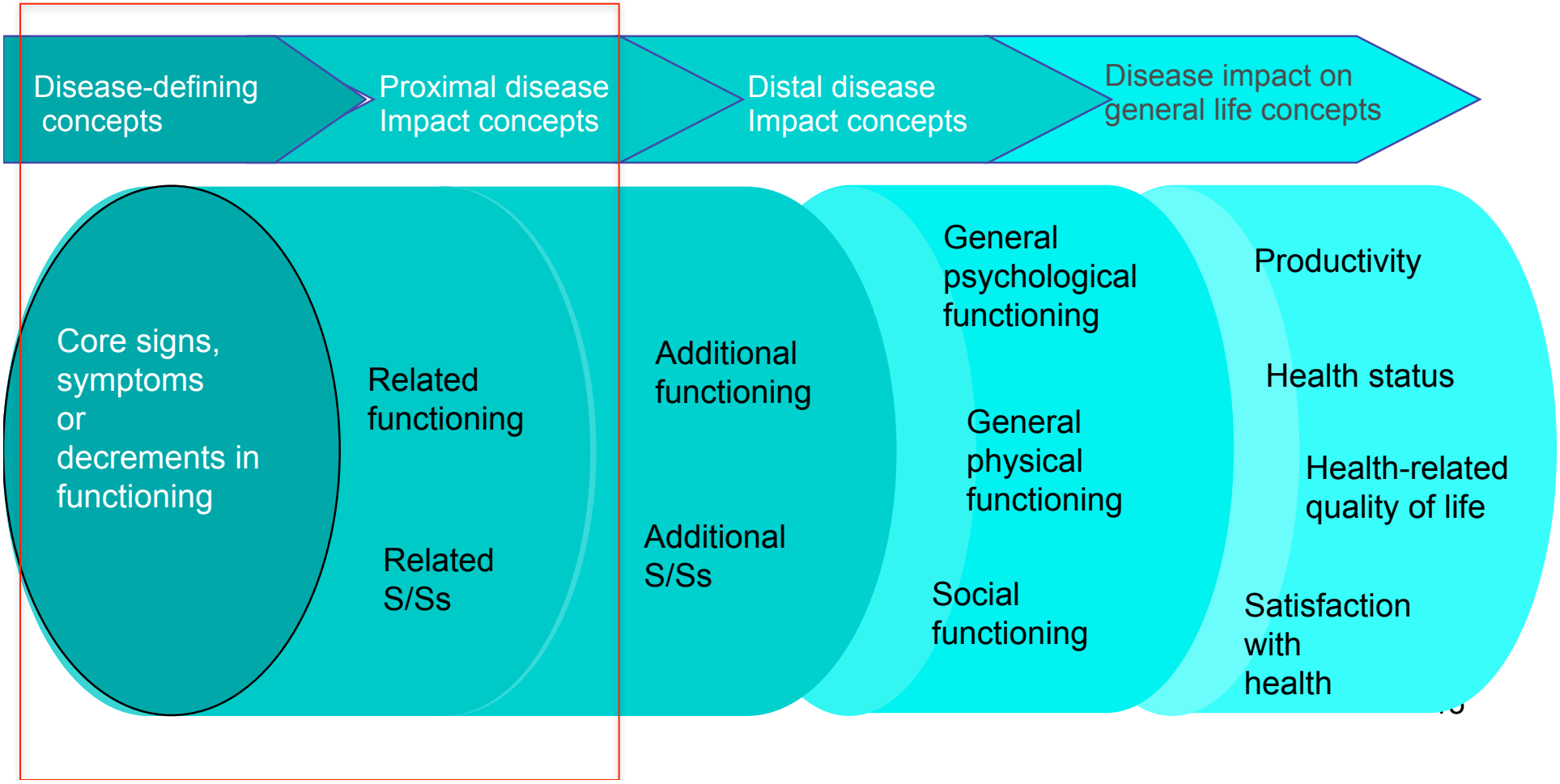
An Endpoint Model displays the role and hierarchy of relevant outcome concepts in clinical trials (i.e., all primary and secondary endpoints)

Endpoints		Concepts	COA/Biomarker/Survival
Primary	→	Concept A	OA 1
Secondary with Hierarchy	→	Concept B	OA 2
	→	Concept C	OA 3
	→	Concept D	OA 4
	→	Concept E	OA 5
Exploratory	→	Other concept	Other OA

What COAs Are Eligible for Qualification?

- Context of use is well-defined
- Measurement concept is appropriate for the context of use
- Measurement is intended to support primary or secondary endpoints related to treatment benefit
- Review of evidence provides confidence that the assessment adequately measures the measurement concept and the evidence is specific to:
  - **The concept of measurement**
  - **The context of use**

Concepts of Measurement Representing Direct Evidence of Treatment Benefit (Indirect Assessments Do Not Appear Here)



Natural History Studies

- Can provide the basis for describing the disease
  - Track course of disease over time
  - Provide information about variability/heterogeneity
  - Identify demographic, genetic, environmental and other variables that correlate with disease and outcomes in the absence of treatment
  - Example: Baron et al, 2009. (PDN and PHN)
- Contribute to scientific foundation upon which drug development programs can be built
- Independent of individual investigational agents
- Most informative when NH study data are available early in development Ideally before design of efficacy trials
- Patient and caregiver involvement is important
  - Engage all stakeholders early and on an ongoing basis

- Institute of Medicine. 2010. *Rare Disease and Orphan Products. Accelerating Research and Development*

What Is Content Validity?

- Empiric evidence that the score is a measurement of the intended concept in the specified context of use
  - Claims in labeling are based on confidence that the claimed concept (direct or indirect evidence of treatment benefit) was measured validly and results were interpretable in the context of use studied
  - Traditional statistical tests of validity (internal consistency, correlations with other measures, known group differences) do not tell us what a score represents
- Established before evidence of construct validity, reliability or sensitivity to change can be interpreted

What Is the Relationship between Content Validity and Context of Use?

- Content validity is specific to the context of use in which the evidence was generated
- If the existing measurement is to be adapted for a new context of use, additional content validity evidence may need to be developed
- FDA reviews content validity within each context of use
  - There's no such thing as a "validated instrument."

What Is Evidence of Content Validity?

Qualitative

Includes literature review and expert opinion

- Protocol and hypothesis driven
- Input from target responder population to document understandability and comprehensiveness (interviews, focus groups)
- Multiple rounds of qualitative research necessary to support
  - Development of the content
  - Refinement of the content
  - Confirmation of validity with the final content and in the final format

Quantitative

Can be used iteratively with qualitative evidence to finalize measurement content

Includes evidence that:

- Scores represent a single concept
- Scale represents less severe to more severe
- Response options are correctly ordered and spaced from less severe to more severe
- The range and distribution of scores is adequate for the context of use?

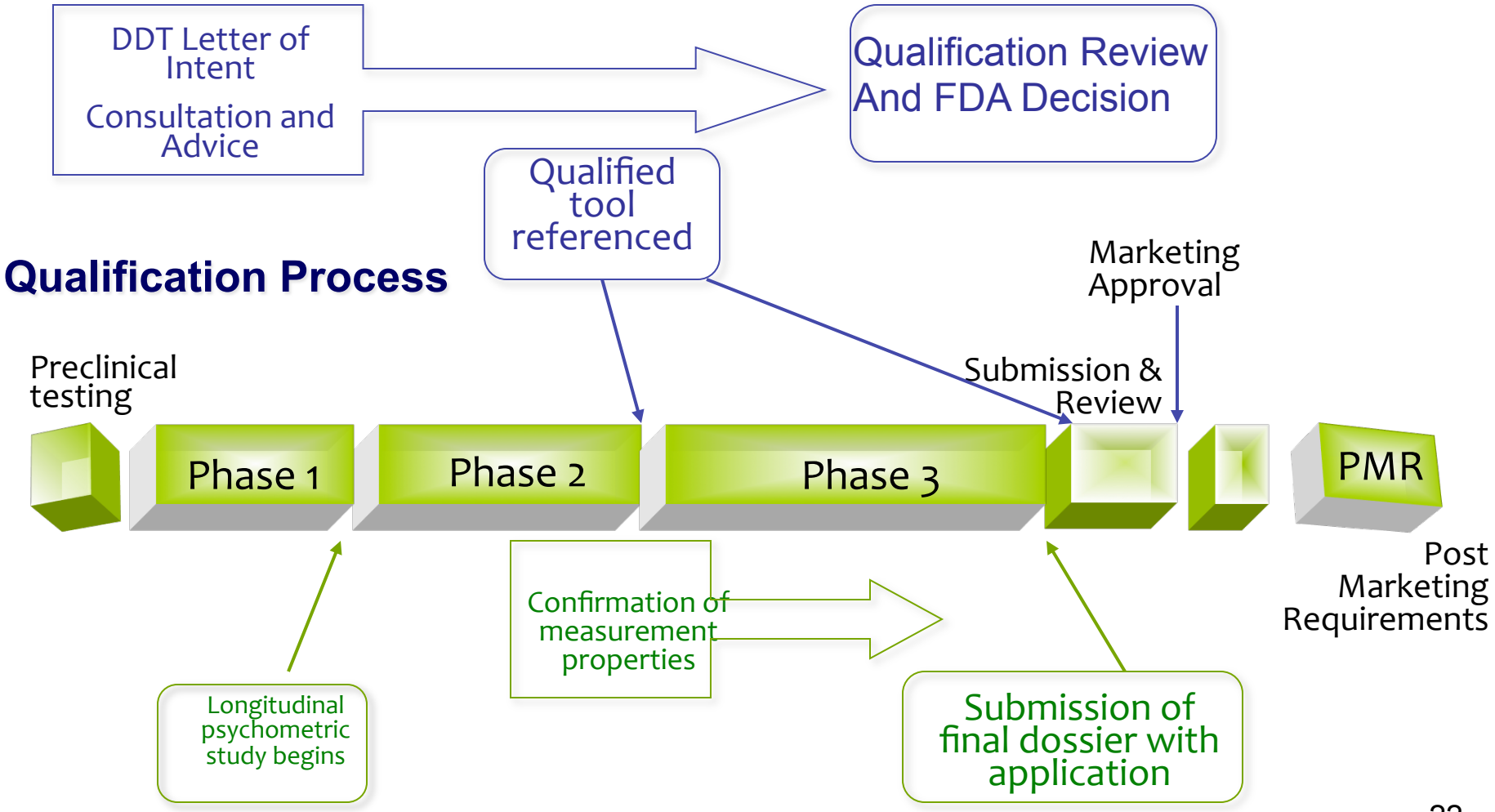
Other Measurement Properties

- Longitudinal studies to establish other measurement properties:
  - Construct validity (if a hypothesized relationship to other measures exists)
  - Reliability (e.g., test-retest in stable patients)
  - Ability to detect change
    - Mean change from baseline
    - Responder definition
- All need to be demonstrated with the final version of the instrument that has been demonstrated to have adequate content validity

Replacement Value Evidence for Indirect Measurements of Treatment Benefit

- Part of the evidence of "well-defined and reliable" for indirect measurements
  - Demonstrate the relationship of the indirect measure to how patients survive, feel or function
- Generally requires longitudinal studies to demonstrate predictive validity
- Often described as evidence of clinical meaningfulness
- Most diagnostic tests will correlate in the .5-.7 range because they are not directly measuring how patients function or feel

DDT Development: Two Processes



Review of a COA Measurement

- What concept is represented by the score?
- What is the context of use proposed?
- Is the concept clinically meaningful/important/relevant in the proposed context of use?
- Does the concept directly measure how patients feel and function in daily life?
- If not, what is the replacement value of the measure?
- Does the COA measure the concept in the appropriate context of use in a valid way?
- What are its other measurement properties (traditional validation data) in the context of use?