Panel Discussion 4: Selection and Development of Clinical Outcome Assessments (COAs) for Use in Pediatric Clinical Trials

THIRD ANNUAL
PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP

April 4, 2012 ■ Silver Spring, MD

Co-sponsored by

CRITICAL PATH INSTITUTE
collaborate · innovate · accelerate

FDA

Objectives

- To discuss practical aspects of incorporating COAs to document treatment benefit in pediatric trials.
- To report on challenges/opportunities when developing a pediatric instruments for use in drug development programs

Agenda

- Moderator
  - Melissa Tassinari PhD, DABT
- PROs in Pediatric Drug Development: Industry Perspectives & Needs
  - Paul Wang, MD
- Selection and Development of Clinical Outcome Assessments (COA) for Use in Pediatric Clinical Trials
  - Linda Abetz-Webb, MA; Donald Patrick, PhD
- FDA response
  - Jessica Lee; Elektra Papadopoulos, MD
- Open floor discussion

PROs in Pediatric Drug Development: Industry Perspectives & Needs

Paul Wang, MD
Seaside Therapeutics, Inc.

THIRD ANNUAL
PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP

April 4, 2012 ■ Silver Spring, MD

Co-sponsored by

CRITICAL PATH INSTITUTE
collaborate · innovate · accelerate

FDA

Industry motivations for pediatric drug development

- **Primary focus on pediatric population (rare)**
  - Small populations; limited commercial potential
  - Exceptions: vaccines, Respiratory Distress Syndrome, single gene disorders (e.g., Fabry's, cystic fibrosis)
    - Major morbidity/mortality are typical
    - PROs generally not needed
- **PREA, BPCA, and EC 1901/2006**
  - 6 month exclusivity extension
  - Regulatory mandate

Pediatric Drug Development Timeline

- **Pediatric trials commonly deferred pending demonstration of safety/efficacy in adults**
- **Pediatric program timeline**
  - 3-5 years to complete pediatric program
  - Pre-clinical: developmental tox; formulation development
  - Pediatric PK study(ies): 6-12 months (?)
  - Pediatric safety/efficacy study(ies): 2-3 years (?)
- **EU: PIP requires early pediatric planning**
  - PIP required after adult PK completed
  - PIP should specify efficacy endpoint

⇨ **PRO development must start very early in the drug development program**

PRO development: Should not be sponsor-specific

- **Work should begin before industry sponsors ready to commit to pediatrics**
- **PRO should not be proprietary**
  - Trials of multiple agents for a single indication (same endpoint) are helpful
    - ADHD & bipolar: multiple agents approved
    - Depression: many failed trials
  - Trials of non-pharmacologic interventions can use the same PRO

Example:
PRO development (vetting) in Autism

- **No drugs approved for core symptoms of autism**
- **Few rigorous efficacy trials for non-drug interventions**
  - No gold-standard outcome measure(s)
- **Review of available PROs**
  - Workgroup sponsored by Autism Speaks; Jan11 – Mar12
  - Criteria
    - Construct validity
    - Clinical meaningfulness
    - Psychometric properties
    - Burden
    - Sensitivity to change
  - Discussion with FDA Mar-2012

Operational considerations for PROs in pediatric trials

- **Trial personnel**
  - Inexperienced in GCP clinical trials
  - Over-stretched academics vs. non-specialist private sites
  - Staff turnover (academic sites)
- **Patient population**
  - Age range
  - Family structure (single-parents; multiple households)
  - Non-English speakers (in USA and globally)
- **Operational context**
  - Investigators meeting (USA and global)
  - Clinician rater training & certification (initial & on-going)
  - Parent rater training

⇨ **Ideal PRO characteristics**

  - Single version across full age range
  - Validated in English, Spanish(es), and other languages
  - Training program available (live & recorded)
  - Robust across raters, languages, ages, severity

PRO sensitivity to change

- **Best**
  - Prior positive result in controlled drug trial
- **Good**
  - Prior positive result in controlled non-drug trial
- **OK**
  - Longitudinal change in natural history study
- **Better than nothing**
  - Prior positive result in uncontrolled trial

Example:
Aberrant Behavior Checklist in Fragile X

- **ABC - Community edition**
  - Developed & validated in general intellectually-disabled population
  - 56 items
  - 5 factor-derived subscales
- **ABC-Irritability subscale used for risperidone & aripiprazole programs**
  - No known validation in autism
  - Primary endpoint for (positive) pivotal trials
- **Primary endpoint in Phase 2 study in Fragile X syndrome**
  - Never validated in FXS
  - Chosen for regulatory precedent
  - No effect vs. placebo
- **Potential reasons for failed trial**
  - Drug doesn't work
  - Insensitive to change in this population
  - Invalid factor structure in this population

Sensitivity to change of ABC-Irritability in Fragile X

- **No drug effect on ABC-Irritability**
  - Scale items: Aggressive; tantrums; yells
  - STX209: -4.2 ± 0.85 (mean ± SE)
  - Placebo: -4.5 ± 0.85
  - $p = 0.823$
- **But: Visual Analog Scale for problem behaviors**
  - "Externalizing" category: Aggression, outbursts
  - STX209: -2.2 ± 0.47 (mean ± SE)
  - Placebo: -1.0 ± 0.48
  - $p < 0.05$

Factor structure of ABC scale in FXS

- **Originally validated in MR/ID population**
  - 5 factors, including ABC-Lethargy/Social Withdrawal
- **Validation in FXS population (Sansone et al., 2011, epub)**
  - > 600 subjects across 6 centers
  - Factor structure differs from original validation population
    - New factor score: ABC-Social Avoidance
- **Unvalidated vs. validated scoring algorithms**
  - ABC-Social Withdrawal (not valid in FXS)
    - STX209: -2.0 ± 0.88 (mean ± SE)
    - Placebo: -1.3 ± 0.87
    - $p = 0.604$
  - ABC-Social Avoidance
    - STX209: -1.2 ± 0.24
    - Placebo: -0.1 ± 0.24
    - $p < 0.01$

Selection and Development of Clinical Outcome Assessments (COA) for Use in Pediatric Clinical Trials

Donald Patrick[1] and Linda Abetz-Webb[2]
[1] University of Washington, Seattle, USA
[2]Adelphi Values, Bollington, UK

THIRD ANNUAL
PATIENT-REPORTED OUTCOME (PRO) CONSORTIUM WORKSHOP

April 4, 2012 ■ Silver Spring, MD

Co-sponsored by

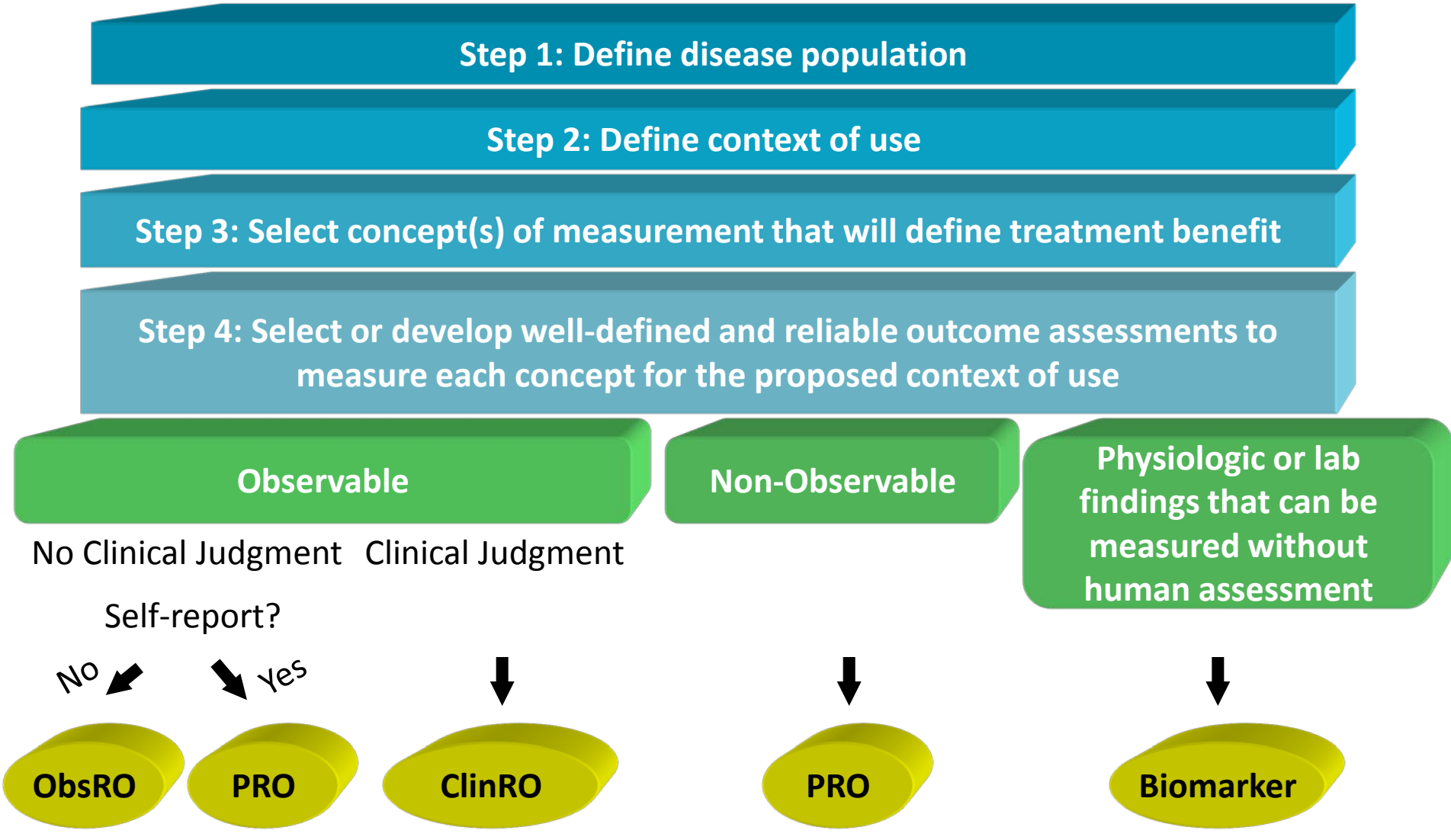
CRITICAL PATH INSTITUTE
collaborate · innovate · accelerate

FDA

Question 1

What kinds of Clinical Outcomes Assessments are used in developing pediatric instruments for use in medical product evaluation?

Endpoint Selection



Why is Self-Report Preferred When Possible?

- Some treatment effects are known ***only*** to the youth themselves, e.g., pain, mood, perceived health, daily function
- Biomarkers and ObsROs do not or may not accurately reflect how youth feel and function
- Well-developed measures reported by youth can be as reliable as observations reported by clinicians or parents
- Evidence base for measures at younger ages needed

"Why Can't We Just Adapt Our Adult Questionnaire?"

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Adult Question

Overall, how would you rate the RLS discomfort in your legs or arms?

- ❑ Very Severe
- ❑ Severe
- ❑ Moderate
- ❑ Mild
- ❑ None

(1 week recall questionnaire)

"Why Can't We Just Adapt Our Child Questionnaire?"

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Child Questions

1. At any time since you got up today, did your **legs**:

Feel like ants or bugs were crawling inside them?	☐ Yes	☐ No
Feel like something was pulling inside them?	☐ Yes	☐ No
Feel weird?	☐ Yes	☐ No
Feel like something is fizzing inside them?	☐ Yes	☐ No

2. How **bad** were these feelings in your **legs** today?


☐ **Very** bad


☐ **Bad**


☐ **A little** bad


☐ **Not** bad at all

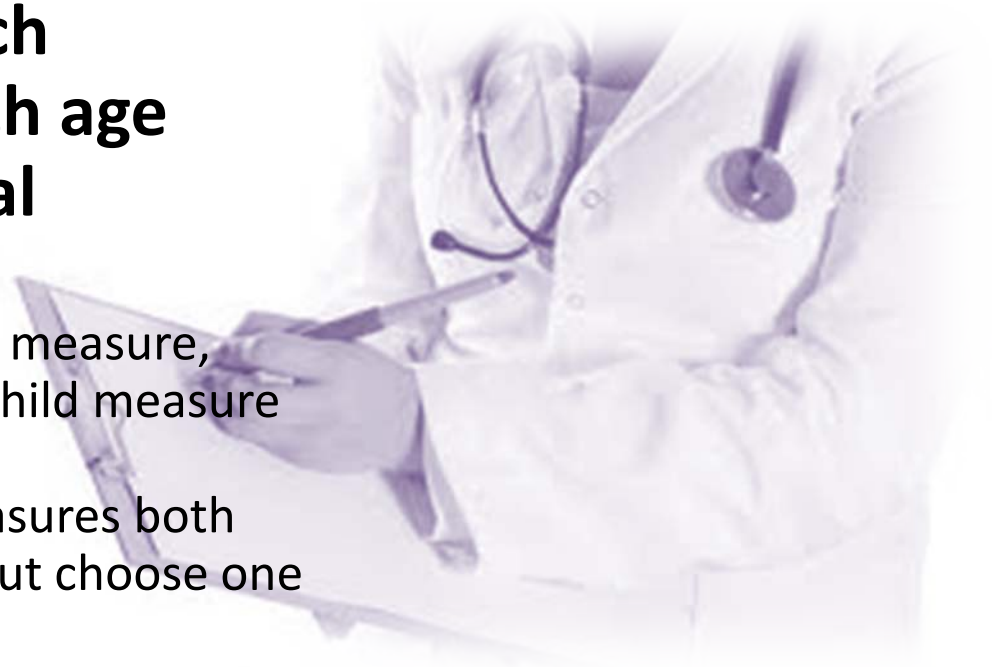
(recall period of "today")

Question 2

How do you decide what type of COA is appropriate for pediatric target populations from birth to age 18 including infants, toddlers, young (6-11), adolescent (12-18) ?

Who to ask when?

- **Consider age, disease, stage in terms of:**
  - Cognitive development
  - Motor development
  - Social/emotional development
  - Literacy
- **Need to determine which measures to use in which age groups, assuming normal cognitive development:**
  - 0-5: rely on caregiver observation measure, maybe interviewer administered child measure for 3-5
  - 6-8; 9-11: child and caregiver measures both have weaknesses – include both but choose one as primary?
  - 12-17: child measure generally primary, but caregiver measure can still add value



Stage in Pediatric PRO Instrument Design

PRO CONSORTIUM
CRITICAL PATH INSTITUTE

Desk Research (literature review) and Hypotheses

↓

Concept Elicitation through Patient/Parent Interviews – covering age ranges/developmental milestones

↓

Development/selection of questionnaire – specific to age ranges/developmental milestones/disease

↓

Content Validity Testing (Cognitive Debriefing) – across age ranges/developmental milestones/disease

↓

Item Reduction, Development of Scoring and Psychometric Validation – across age ranges/developmental milestones

↓

Instrument Ready for Implementation in Trial/Study

Input from expert clinicians throughout

Psychometrics

- No established guidelines
- Affected by concreteness of reported concept
- Age 7 often cited as bottom of age range
- Mixed validity & reliability results below age 11
- Ages 7 to 11: Combination of self- and observer-report may be best
- Age 11+ generally acceptable psychometrically
- Age is generally not best criterion, assessment of comprehension & willingness/motivation to respond is better

Question 3

What methods do you use and what are the challenges in developing an Observer-Reported Outcome measure or Parent-Observed measure to assure content validity and measurement properties?

Verifiable Observation

A sign or impact must be able to be detected by a sense or senses:

- Seen (*vision*)
- Heard (*auditory*)
- Smelled (*olfactory*), or
- Felt (*touch*)

Self-Report Concept	Observer-Report Concept (Sense)
Difficulty breathing	Gasping for breath? (see)
Feverish	Feverish (touch)
Tired	Lying down to rest more than usual (see)
Chills or sweats	Chills or sweats (see, touch)
Cough	Cough (see, hear)
Cough up mucous	Cough up mucous (see)
Tightness in chest	None
Wheeze	Wheeze (hear)
Difficulty sleeping	?
Worried	?
Irritable	Behaving fussy (see, hear)
Depressed	?
Frustrated	Behaving fussy (see, hear)
Time spent sitting or lying down	Time spent sitting or lying down (see)
Reduce usual activities	Reduce usual activities (see)
Miss school or work	Miss school or work (see)

Observable Signs: Examples in Ages 0-6

- ***Labored breathing (hear):*** Even as he's breathing at night, it'll sound like the breathing is more labored, it's harder for him to breathe, and it'll have that kind of crackle-pop kind of thing going on, as well. (#13)
- ***Fussiness (see):*** "Well, about a week and a half ago or so, he started just getting fussy, and he wouldn't finish all of his bottle which is not normal for him. And so his appetite was tapering off. He was getting fussier." (#2)

Reliability and Validity of Observation

- Inter-rater reliability is more appropriate than test-retest, although difficult to implement

- Test-retest is affected by consistency and frequency of observed phenomenon over time, as well as by measurement error

- ***KEY POINT*** -- Reliability not sufficient by itself: "How do we know we are measuring the right thing(s) for development, condition, and treatment context?"

Child Tells Parent

- Positives:
  - can provide useful additional supportive information
  - parents often rely on this in verbal, pre/middle school age children so has content validity
- Cons:
  - parent may ‘interpret’ what child says
    - ‘she just wants to get out of going to school; it’s not real’
  - child may not tell their parents things

Question 4

What are the challenges/solutions in developing a child reported measure (with a focus on some of the issues we encounter in eliciting information or interpreting qualitative data, as well as ensuring appropriate representation of different ages/cognitive/motor/emotional development abilities ?

Interviewing for Concept Elicitation and Content Validity Testing

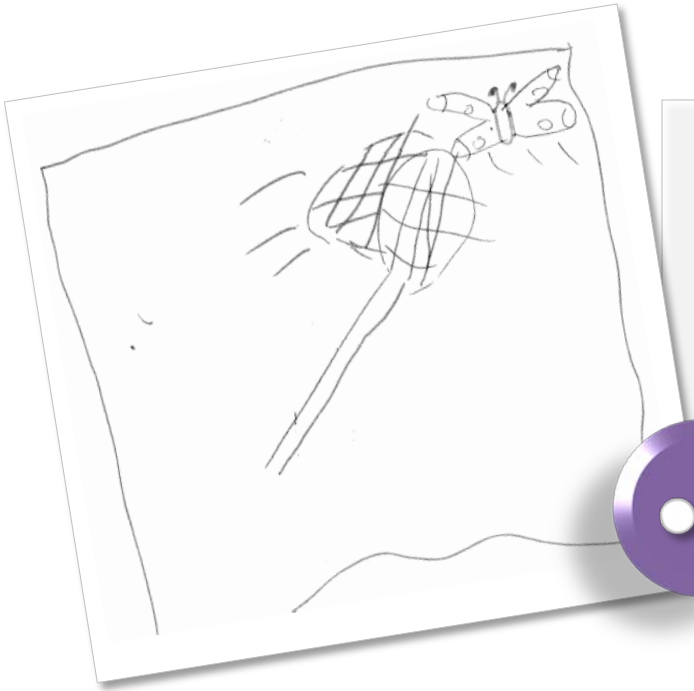
Challenges specific to paediatrics

- Limitations in attention span, memory, cognitive ability, language comprehension by age
- Children can be shy, lack vocabulary, may misunderstand your interview questions
- Parents are unable to report

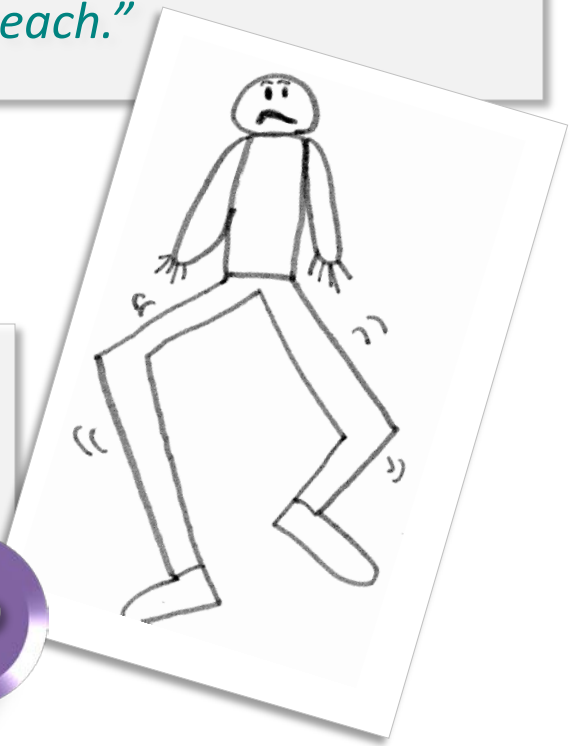
Solutions

- Carefully guided interview guides and well trained interviewers
- Creative interview techniques, toys and drawings
- Achieve saturation/cognitively debrief within each narrow age/developmental range
- Develop questionnaire for lowest age group (without insulting adolescents)

Drawings of Symptoms



094-F-13-C – "I drew a leg and a butterfly. And - well, that I just missed the butterfly… just like barely missed… Because I always try to find that spot and it never really gets there. It always slips out of reach."



066-M-8-C – "It's like my legs are wiggly and like…Probably in the morning or after school [frown on face as well]"

Picchietti, Arbuckle, Abetz et al, 2011; J Child Neurol.

Interviews vs. Focus Groups

Interviews strongly recommended over focus groups in 6-11 year olds

- Keeping the attention of more than one child extremely difficult
- Children very likely to be biased by other members of the group

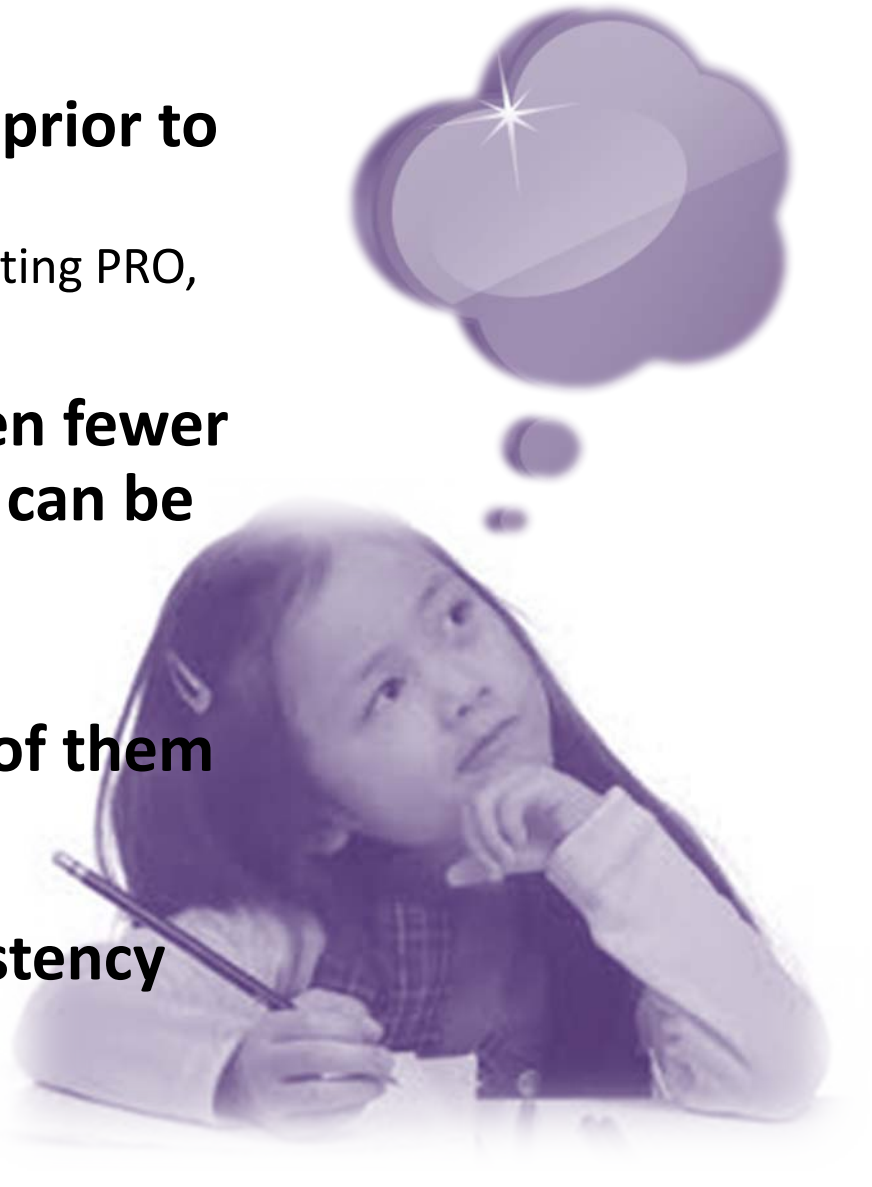
Focus groups possible in adolescents but interviews still recommended

- Younger adolescents may be intimidated or influenced by older members of group

For focus groups, consider appropriateness of mixed gender and mixed age groups

Content Validity Testing: Cognitive Debriefing

- **For diaries – recommend child completes diary for several days prior to cognitive debriefing**
  - Gives child real experience of completing PRO, avoids inquiry about hypothetical
- **Compared with adults for children fewer questions about comprehension can be asked**
- **Younger children can struggle to understand what is being asked of them in the interview**
- **When analysing, check for consistency between behaviour and across responses**



Psychometric Validation Studies and Clinical Trials

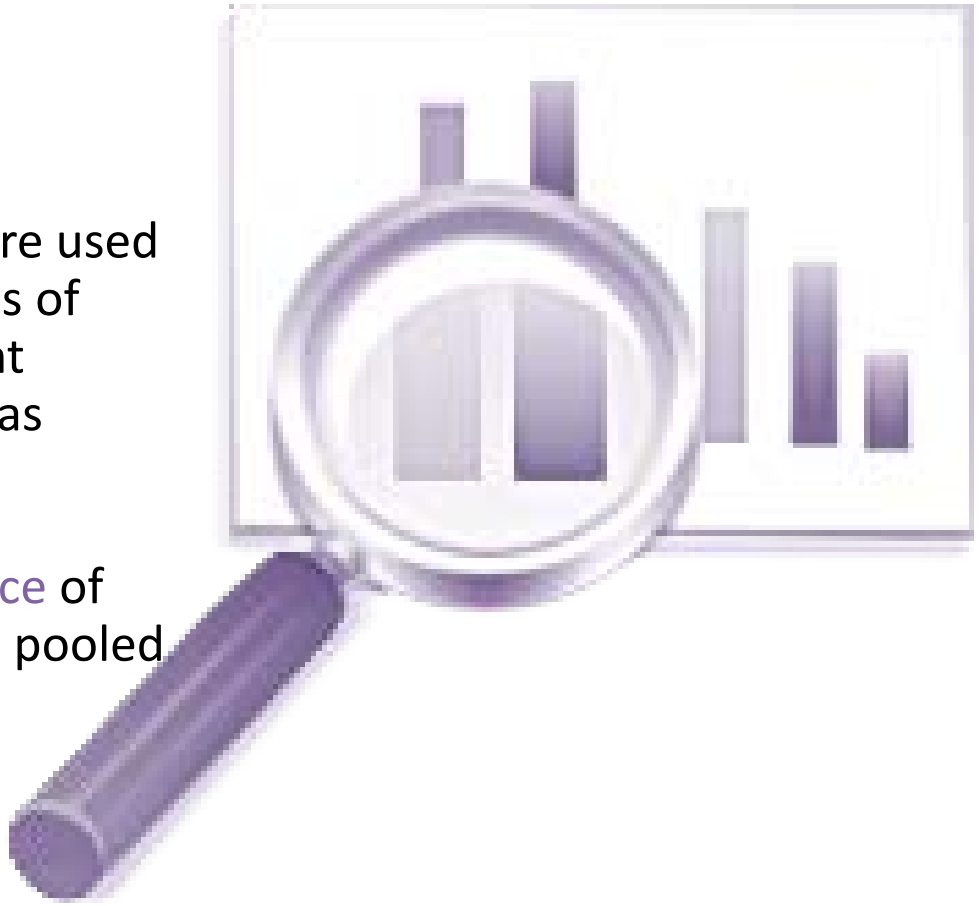
- **Sample must be stratified by age group**
  - Necessary to demonstrate validity/efficacy in each age group as well as overall sample
  - Increases the sample size required for validation
- **Recruiting patients for treatment intervention trials can be challenging**
  - Consider validating in observational study
  - BUT difficult to fully evaluate responsiveness in an observational study
  - Consider a hybrid approach?

Question 5

If you include more than one type of COA with similar or separate concepts for target labeling in the endpoint model, how do you decide on how these are to be reconciled? If you have more than one type of COA in an endpoint model, how do you decide if concepts overlap or do not overlap?

Pooling Data

- **Pooling data across age groups and COAs**
  - Not possible if different concepts are used in different age groups or if versions of questions are conceptually different
  - Ideal is for different versions to be as similar as possible
  - Otherwise suggest attempting to demonstrate conceptual equivalence of different versions If data cannot be pooled – larger sample size required



Use Mixed Methods to show Conceptual Equivalence

- Consider different age versions as being like different language (cultural) versions
- Use BOTH qualitative research and Rasch\IRT analysis to show conceptual equivalence across age ranges (as you would in linguistic validation)

Take Home Messages

- Consider age and developmental changes at all stages of PRO development, testing and in clinical trial design
- Obtain information from literature, clinicians, parents and most importantly the children themselves
- Recognise that different versions of a PRO might be required for different age/developmental groups
- This takes TIME: approximately 1-2 years. Start early

FDA Response

Jessica Lee, Elektra Papadopoulos