Psychometric Evaluation of Survey Data



Questionnaire Design and Testing Workshop

October 24, 2014, 3:30-5:00pm 10940 Wilshire Blvd. Suite 710 Los Angeles, CA

Class Roster

Priya Batra Reshma Gupta Lillie Hudson Roya Ijadi-Maghsoodi **Ippy Kalofonos** Linda Kim (audit) Christine Lam Lucinda Leung Yalda Rahimi Adam Schickedanz Megha Shah Jesus Ulloa























Patient-Reported Outcomes (PROs)

- "Any report of the status of a patient's health condition that comes directly from the patient, without interpretation of the patient's response by a clinician or anyone else"
 - Patient reports about their health
 - What they can do and how they feel
 - Patient evaluations of health care

PRO Development Process

Hypothesize Conceptual Framework

- Outline hypothesized concepts and potential claims
- Determine intended population
- Determine intended application/characteristics (type of scores,

PRO

Claim

- mode and frequency of administration)
- Perform literature/expert review
- Develop hypothesized conceptual framework
- Place PROs within preliminary endpoint model

Document preliminary instrument development

v. Modify Instrument

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

iv. Collect, Analyze, and Interpret Data

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

ii. Adjust Conceptual Framework and Draft Instrument

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

iii. Confirm Conceptual Framework and Assess Other Measurement Properties

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

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

Identify Concepts and Hypothesize Conceptual Framework

- Literature, media, and citizen reports used to identify concepts of interest and potential confounders
 - Functioning limitations
 - Satisfaction with surgery
 - Dry eye symptoms
 - Expectations of surgery
 - Coping
 - Optimism/pessimism
 - Depression/anxiety symptoms

PRO Development Process



- Document interpretation of treatment benefit in relation to claim
- and training materials Document measurement development

http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/ 7 Guidances/UCM205269.pdf

Adjust Conceptual Framework and Draft Instrument

- Evaluated published surveys of target concepts
- Obtained permission to use copyrighted items
- Wrote new items

Included Several Existing Measures in Draft Instrument

- National Eye Institute Refractive Error Quality of Life (NEI-RQL-42)
- National Eye Institute Visual Functioning Questionnaire (VFQ-25)
- Ocular Surface Disease Index (OSDI)
- Life Orientation Test Revised (LOT-R)
- Brien Holden Vision Institute Multidimensional Quality of Life (BHVI QOL) Scale for Myopia
- Work Productivity Activity and Impairment (WPAI)
- Patient Health Questionnaire (PHQ-4)
- Marlowe-Crowne Socially Desirable Response Set

Example of Visual Symptom Aberration Item

INSTRUCTIONS: The next few questions are about <u>starbursts</u>. By starbursts, we mean seeing rays of light coming out from lighted objects, such as in the car headlights in the images below. These images may not represent exactly what you see and your symptoms may be more or less severe than what is shown.



In the last 7 days, have you seen any starbursts?

- 1. Yes, but ONLY when NOT wearing glasses or contact lenses
- 2. Yes, but ONLY when wearing glasses or contact lenses
- 3. Yes, when wearing AND when not wearing glasses or contact lenses
- 4. No, not at all

Cognitive Interviews to Evaluate Draft Instrument

- Objectives:
 - To evaluate the content and ordering of the questionnaire, coverage of treatmentrelated issues pertinent to LASIK patients
 - To evaluate the usability of the electronic format of the questionnaire

Cognitive Interviews Conducted by RAND

- Conducted in Los Angeles, CA and Washington, DC
- Pre-operative patients (n=9)
 - Adults very likely to have LASIK in the next 6 months
- Post-operative patients (n=9)
 - 1 dissatisfied
 - 4 satisfied
 - 4 with visual symptoms
- General Exclusions
 - Eye care professionals, web site designers, and prior refractive surgery

Adjust Conceptual Framework: Based on Cognitive Interviews

- Modified questionnaire with the following changes
 - Ordering of items
 - Clearer wording of some items
 - Additional phrases to remind respondents of time frame
 - Instructions were shortened and bulleted
 - Formatting of web-based questionnaire to resemble other online surveys

Survey Measures

- Existing Measures
 - 7 NEI-RQL-42 scales (23 of 42 items)
 - NEI-VFQ-25 driving scale (3 items)
 - 2 of 3 Ocular Surface Disease Index scales (8 of 12 items)
 - Lost work and productivity due to eye problems (3 items)
- New Measures
 - Visual aberrations (4 scales)
 - Expectations of spectacle independence/vision clarity (6 items)
 - Satisfaction with vision (1 item)
 - Satisfaction with LASIK surgery (8 items)
- Optimism (10 items)
- Health Proneness (10 items)
- Depression and Anxiety (4 items)

PRO Iterative Development



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

Assess Measurement Properties: Reliability

Degree to which the same score is obtained when the *target* or thing being measured (person, plant or whatever) has not changed.

✓ Internal consistency (items)

✓ Need 2 or more items

✓ Test-retest (administrations) correlations
 ✓ Need 2 or more time points

Reliability

Model	Reliability	Intraclass Correlation
Two-way random	$\frac{N(MS_{BMS} - MS_{EMS})}{NMS_{BMS} + MS_{JMS} - MS_{EMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS} + k(MS_{JMS} - MS_{EMS})/N}$
Two- way mixed	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{EMS}}{MS_{BMS} + (k-1)MS_{EMS}}$
One- way	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS}}$	$\frac{MS_{BMS} - MS_{WMS}}{MS_{BMS} + (k-1)MS_{WMS}}$
BMS = WMS = JMS EMS =	 Between Ratee Mean Squ Within Mean Square Item or Rater Mean Squa Ratee x Item (Rater) Mea 	are N = n of ratees k = n of items or raters re n Square

Alpha Reliability Formulas



Internal Consistency Reliability and Item-Scale Correlations for 23 Multi-Item Scales

• PROWL-1

- Median alpha = 0.78 (range: 0.55-0.98)

• PROWL-2

- Median alpha = 0.81 (range: 0.63-0.97)

 Item-scale correlations (hypothesized scales vs other scales) support item discrimination across scales

Item-scale correlation matrix

	<u>Depress</u>	<u>Anxiety</u>	<u>Anger</u>
ltem #1	0.50*	0.50	0.50
Item #2	0.50*	0.50	0.50
Item #3	0.50*	0.50	0.50
Item #4	0.50	0.50*	0.50
Item #5	0.50	0.50*	0.50
Item #6	0.50	0.50*	0.50
Item #7	0.50	0.50	0.50*
Item #8	0.50	0.50	0.50*
Item #9	0.50	0.50	0.50*



*Item-scale correlation, corrected for overlap.

Item-scale correlation matrix

	<u>Depress</u>	<u>Anxiety</u>	<u>Anger</u>
ltem #1	0.80*	0.20	0.20
Item #2	0.80*	0.20	0.20
Item #3	0.80*	0.20	0.20
Item #4	0.20	0.80*	0.20
Item #5	0.20	0.80*	0.20
Item #6	0.20	0.80*	0.20
Item #7	0.20	0.20	0.80*
Item #8	0.20	0.20	0.80*
Item #9	0.20	0.20	0.80*



*Item-scale correlation, corrected for overlap.

PROWL-1 Item-Scale Correlations Example

Item Number	Ocular Surface Disease	NEI-RQL Clarity Vision
Q65 (eyes sensitive to light)	0.38*	28
Q66 (eyes feel gritty)	0.32*	20
Q67 (painful or sore eyes)	0.32*	15
Q68 (blurred vision)	0.46*	<u>57</u>
Q69 (poor vision)	0.47*	<u>53</u>
Q70 (uncomfortable—wind)	0.44*	15
Q71 (uncomfortable—humidity)	0.45*	15
Q72 (uncomfortable—air cond.)	0.31*	15
Q5 (how clear is your vision?)	21	<u>0.12*</u>
Q34a (distorted vision)	35	0.52*
Q35a (blurry vision)	45	0.65*
Q36a (trouble seeing)	48	0.62* 22

Assess Measurement Properties: Validity

- Content validity: Does measure "appear" to reflect what it is intended to (expert judges or patient judgments)?
 - Do items operationalize concept?
 - Do items cover all aspects of concept?
 - Does scale name represent item content?
- Construct validity
 - Are the associations of the measure with other variables consistent with hypotheses?

Threats to Validity

- Those with higher levels of expectations about surgery will be less satisfied with surgery
 - The correlations between expectations and satisfaction with surgery were <u>not</u> statistically significant at the 1-month, 3-month, and 6month follow-ups in PROWL-1.

Threats to Validity

 Those with lower health proneness at baseline will be less satisfied with surgery

 Only the correlation of health proneness with 3-month satisfaction with surgery was statistically significant and it was a small correlation (r = 0.14, p = 0.0443) in PROWL-1.

Threats to Validity

- Those with depressive/anxiety symptoms at baseline will be less satisfied with surgery
 - Only the correlation of the PHQ-4 with 6month satisfaction with surgery was statistically significant and it was a small correlation (r = -0.19, p = 0.0043), PROWL-1.

Support for Validity

• Those with a greater degree of visual aberrations will be less satisfied with surgery

- Correlations statistically significant in hypothesized direction at 1-month, 3-month and 6-month follow-up in PROWL-1:
 - Glare (r's = 0.34, 0.36, 0.43)
 - Starbursts (r's = 0.27, 0.24, 0. 32)
 - Haloes (r's = 0.37, 0.34, 0.49)
 - Double images (r's = 0.43, 0.37, 0.39)

Usability Results

	PROWL-1	PROWL-2	
Minutes to Complete (median)	25	20	
Length of Questionnaire About right A little too long	46% 40%	64% 33%	
No problems using computer True False	86% 10%	90% 8%	
Ease of taking questionnaire by computer vs paper Easier Harder	54% 14%	69% 5%	

PRO Iterative Development Process



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

Item Response Theory (IRT)

IRT models the relationship between a person's response Y_i to the question (i) and his or her level of the latent construct θ being measured by positing

$$\Pr(Y_i \ge k) = \frac{1}{1 + \exp(-a_i\theta + b_{ik})}$$

b_{ik} estimates how difficult it is for the item (i) to have a score of k or more and the discrimination parameter a_i estimates the discriminatory power of the item.

Item Responses and Trait Levels



www.nihpromis.org

Computer Adaptive Testing (CAT)











always. Indeed, the item depicted in Figure 2, item 31, was 1 of the 5 items dropped from the communication scale based on the CTT analyses.

The information curve provides an indication of the amount of information the scale yields at different points along the underlying continuum. Information is inversely related to

Because the participation rate was 50%, some caution is warranted in interpreting the study results. Nonetheless, separate analyses of a CAHPS® item similar to item 276 revealed that the negative wording of this item confuses respondents. As a result, when CAHPS® 3.0 was released, the item was worded in terms of being seen within 15 minutes ۲



Differential Item Functioning (DIF)

- Probability of choosing each response category should be the same for those who have the same estimated scale score, regardless of other characteristics
- Evaluation of DIF by subgroups

DIF (2-parameter model)



PROMIS Physical Functioning vs. "Legacy" Measures



Thank You!



drhays@ucla.edu (310-794-2294). http://gim.med.ucla.edu/FacultyPages/Hays/