Interpretation of PROs in Clinical Practice
Solutions for Assessing Change and Diverse People

ISOQOL webinar – June 16 2021
Education | ISOQOL
**PRESENTERS**

**Richard Sawatzky** PhD, RN  
Professor and Canada Research Chair in Person-Centred Outcomes  
School of Nursing,  
Trinity Western University & Centre for Health Evaluation and Outcome Sciences,  
Providence Health Care, BC, Canada

**Jae-Yung Kwon** PhD, RN  
Assistant Professor  
School of Nursing  
University of Victoria  
BC, Canada
DISCUSSION PANEL

Theresa Coles  PhD  
Assistant Professor  
Population Health Sciences, Duke University, North Carolina, USA

Robert Klaassen  MD FRCPC  
Full Professor and Hematologist/Oncologist  
University of Ottawa, Ontario, Canada

Kara Schick-Makaroff  PhD, RN  
Associate Professor  
School of Nursing, University of Alberta, Edmonton, Canada
ACKNOWLEDGEMENTS

This webinar is offered as part of a Michael Smith Foundation for Health Research knowledge translation grant award

- Principal Investigators: R. Sawatzky & J.-Y. Kwon
- Researchers: L. Lix, T. Sajobi, K. Sibley, L. Russell
- Knowledge user team members: R. Brahmbhatt, L. Cuthbertson, S. Lauck, A. Tecson

The resources for this project were developed, in part, thanks to funding from the Canadian Institutes for Health Research and the Canada Research Chairs Program in support of Dr. Sawatzky’s Chair in Person-Centred Outcomes at Trinity Western University.
DISCLAIMER

The International Society of Quality of Life (ISOQOL) does not warrant, endorse, approve or certify information contained in this webinar, nor does it make any representation as to the accuracy, completeness, efficacy, or timeliness of such information.
LEARNING OUTCOMES

Explain the concepts of differential item functioning (DIF) and response shift (RS) and their implications for clinical decision-making

Critically evaluate and consider different ways to assess for DIF and RS in clinical practice
OUTLINE

1 BACKGROUND
   (20 min)
   • Validity evidence for use of PROs to assess change and diverse people
   • Introduction to response shift (RS) and differential item functioning (DIF)

2 INDIVIDUAL LEVEL
   (40 min)
   • Case studies regarding implications and solutions for clinical practice

3 AGGREGATE LEVEL
   (10 min)
   • Overview of aggregate level implications
   • Introduction to analytical approaches

4 CONCLUSION
   (20 min)
   • Breakout discussion
   • Further resources
MOTIVATION FOR THE WEBINAR

- Statistical methods for PROs
  - Differential Item Functioning
  - Response Shift

Analysis of PRO data → Knowledge translation regarding use of PRO data → Interpretation of PROs for clinical use in diverse populations

Our focus today
**Question:** Are you using PRO measures for clinical decision-making purposes?

- **YES** Individual point of care clinical practice
- **YES** Quality improvement
- **YES** Other
- **NO**
Patient story

“Bill” is a 70-year-old who recently lost his wife suffering from advanced cancer and is also trying to cope with multiple chronic illnesses including arthritis. He is visiting his rheumatologist and completes a patient-reported outcome measure.
SECTION 1
BACKGROUND

- Validity evidence for use of PROs to assess change and diverse people
- Response shift (RS) and differential item functioning (DIF)
INTRODUCING PRO MEASURES

VIDEO: https://www.youtube.com/watch?v=qC-XIquQu4g&list=PLKCOAAKefI8N8MN02zAZTm2fiFG4uNGO
USE OF PROS FOR COMPARISONS BETWEEN PEOPLE AND OVER TIME

**Between People**

*In the past 7 days, how often have you felt down, depressed, or hopeless?*

- Not at all
- Several days
- More than half the days
- Nearly every day

**Over time**

*What is your perceived health level from 0 to 10?*

5/10 → 7/10


---

**Measurement Validity**

**Validity Evidence**

- **Compare different people**
  - Differential item functioning (DIF)
    - Different people may not interpret and respond to PRO question in the same way

- **Compare over time**
  - Response shift (RS)
    - People may not interpret and response to PRO questions in the same way at different points in time

**Consequences**

- **Personal Consequences**
- **Societal Consequences**
- **Intended Consequences**
- **Unintended side effects**
PRO MEASUREMENT

PROM
- questionnaire

Psychometric measurement model
- responses to questions

Score
- scaled scores

Healthcare decisions
**PRO MEASUREMENT MODEL**

**PRO**

**QUESTIONS**

**Construct** = The PRO we want to measure

**Measurement model** = Mathematical algorithms that define how each question measures the construct differently

**Questionnaire** = Many different ways of asking people about the construct
COMPARING OVER TIME RESPONSE SHIFT

WHAT IS YOUR PERCEIVED HEALTH LEVEL FROM 0 TO 10?

5/10 ➡️ 7/10

Video: https://www.youtube.com/watch?v=sbkcqJ-Pcso&list=PLKCOAAKeflJ8N8MN02zAZTm2fiFG4uNGO&index=2
INVARINANCE OVER TIME

Individual Level

Aggregate Level
RESPONSE SHIFT

Reconceptualization: Consistency in how the construct is conceptualized

Reprioritization: Consistency in relative importance of measurement indicators and construct

Recalibration: Consistency in scale of measurement

Time 1 ≠ Time 2
Are people consistent in how they interpret and respond to measurement questions at different points in time?

"a change in the meaning of one’s self-evaluation of a target construct as a result of change in":

<table>
<thead>
<tr>
<th>Recalibration</th>
<th>Internal standards of measurement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Reprioritization</td>
<td>Relative importance of domains or items</td>
</tr>
<tr>
<td>Reconceptualization</td>
<td>Definition of the target construct</td>
</tr>
</tbody>
</table>


COMPARING DIFFERENT PEOPLE
DIFFERENTIAL ITEM FUNCTIONING

IN THE PAST 7 DAYS, HOW OFTEN HAVE YOU FELT DOWN, DEPRESSED, OR HOPELESS?
NOT AT ALL ~ SEVERAL DAYS ~ MORE THAN HALF THE DAYS ~ NEARLY EVERY DAY

GROUP A
NOT AT ALL

GROUP B
NEARLY EVERYDAY

VIDEO: https://www.youtube.com/watch?v=AMb_oLKH2I&list=PLKCOAAKelfJ8N8MN02zAZTm2fIuNGO&index=3
INVARINACE BETWEEN DIFFERENT PEOPLE

Person A = Person B

Group A = Group B
Differential Item Functioning

Configural invariance: Consistency in how the construct is conceptualized

Metric invariance: Consistency in relative importance of measurement indicators and construct

Scalar invariance: Consistency in scale of measurement

Group A ≠ Group B

PRO

QUESTIONS

PRO

QUESTIONS
DIFFERENTIAL ITEM FUNCTIONING

Is it reasonable to assume that all people, regardless of their life context, will interpret and respond to items in the same way?

A difference between people in the meaning of one’s self-evaluation of a target construct

- **Scalar invariance**: Internal standards of measurement
- **Metric invariance**: Relative importance of domains or items
- **Configural invariance**: Definition of the target construct
Differential item functioning
- Differences in how people interpret and respond to questions
- Threatens the comparability of scores across individuals or groups

Response shift
- An individual’s frame of reference may change over time
- Threatens the comparability of scores over time
Fairness and equity in PRO measurement for:

- assessing diverse patients
- comparing different groups
- evaluating change over time
USE OF PROS AT INDIVIDUAL AND AGGREGATE LEVELS

Diverse people → PROMs → Healthcare Providers
SECTION 2
INDIVIDUAL LEVEL

Case studies on
considerations of response shift and differential item functioning at the individual level
Patient story

“Bill” is a 70-year-old who recently lost his wife suffering from advanced cancer and is also trying to cope with multiple chronic illnesses including arthritis. He is visiting his rheumatologist and completes a patient-reported outcome measure.
Arthritis Health Questionnaire (hypothetical example)

Please indicate how the following symptoms interfered with your day-to-day activities in the past week.

<table>
<thead>
<tr>
<th>Symptoms</th>
<th>Not at all</th>
<th>A little bit</th>
<th>Moderate</th>
<th>Quite a bit</th>
<th>Extremely</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Pain in joints</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2. Stiffness in joints</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3. Grating in joints</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

How much difficulty did you have conducting activities due to your arthritis in the past week?

<table>
<thead>
<tr>
<th>Physical function</th>
<th>Not difficult at all</th>
<th>A little difficult</th>
<th>Moderately difficult</th>
<th>Quite a bit difficult</th>
<th>Extremely difficult</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Walking briskly</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5. Vigorous activities (e.g., lifting weights, running)</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Social function</th>
<th>Not at all limited</th>
<th>A little limited</th>
<th>Moderately limited</th>
<th>Quite a bit limited</th>
<th>Extremely limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>6. Hobbies</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>7. Relationships with friends/family</td>
<td>5</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Overall health**

<table>
<thead>
<tr>
<th>Overall health</th>
<th>Very good</th>
<th>Very poor</th>
</tr>
</thead>
<tbody>
<tr>
<td>8. How would you rate your health overall?</td>
<td>10</td>
<td>9</td>
</tr>
</tbody>
</table>
RESPONSE SHIFT

How do we ensure valid assessments of change over time?
Choose one of the following responses to the question:
What definition of response shift do you agree with the most?

A. Change in the meaning of health and quality of life over time

B. People do not interpret the items in the same way at different points in time

C. Detection of items performing differently over time when the latent trait is held constant

D. All of the above
PANEL DISCUSSION OF CASE EXAMPLES

How do we identify response shift in clinical practice?

Why does response shift occur?

What do we do when there is response shift?
Bill adjusts his perspective of health in comparison to others over time.
Bill is having a 6-month follow up for hip replacement surgery. Before surgery, Bill responds to the overall health question as 6 out of 10. After surgery, he meets other patients who are in worse shape and subsequently answers slightly higher (7/10) even though he suffers as much pain as before. The then test shows that he rated his “overall health arthritis question” lower as 5 out of 10.

Respond to question in chat:
Have you encountered situations where previous and retrospective PRO scores do not match?
**Scenario example**

**Clinician:** How are you, Bill? Tell me on a scale from zero to ten where zero is worst health you can imagine and ten is the best.

**Bill:** There is still some pain issues but the stiffness has subsided a bit so I would say about 7 out of 10?

**Clinician:** Are you better or worse than how you were six months ago?

**Bill:** Well, in comparison to others, I feel fortunate so thinking back about how anxious I was before, I would give myself 5 out of 10.

**Clinician:** Let me check your file. Yes, when I asked you six months ago, you gave yourself a 6 out of 10. I made notes here about your pain level and how anxious this was making you feel.

**Discussion points for panel:**

How would this information influence your interpretation of the current PRO score?
• Acknowledge changes in perspective or frame of reference to help patients cope with illness.
• Some patients may report higher or lower self-reported health even though there may be no noticeable change by comparing themselves to those who are in worse or better shape.
• Conversely, some patients who have worsened or improved may report no change in their self-reported health.
Bill places more importance on certain health domains (e.g., maintaining social contact) over time.
Scenario example

Clinician: Also, from last time, I see that your symptoms haven’t changed but your “hobby” score has increased.

Bill: I started to volunteer at an outdoor community club.

Clinician: It seems like this new activity is very important to you.

Bill: Yes, it brought a sense of importance to my life.

Discussion points for panel:
How would this information influence the course of follow-up treatment?
• Focus treatment to mitigate adverse effects on domains (e.g., social function) more important to the patient.

• For some patients, importance of some health domains may increase or decrease over time.
RECONCEPTUALIZATION CASE EXAMPLE

Bill has redefined his health over time

Time 1

Symptoms
Physical
Social

Time 2

Symptoms
Physical
Social
Emotional well-being
Scenario example

Clinician: How has your eyesight been? You told me you began to lose your eyesight and are now seeing an ophthalmologist?

Bill: Yes, it has been terrible the last couple of months to adjust and needing help to do everything. However, I have learned to think more holistically.

Clinician: Oh, I see so that explains why your overall health has slightly improved even though your symptoms have worsened, is that correct?

Bill: Yes, I consider health to include emotional well-being as well.

Discussion points for panel:
How would this information influence the course of follow-up treatment?
• Consider holistic treatment from patients’ redefined health perspective that include aspects of mental health and well-being.

• Some patients may redefine how they perceive their health over time.
How do we ensure valid assessments when comparing different people?
Choose one of the following responses to the question:
What definition of differential item functioning do you agree with the most?

A. Differences in the meaning of health and quality of life between people

B. People with the same underlying health status do not interpret the items in the same way

C. Detection of items performing differently in subgroups when the latent trait is held constant

D. All of the above
PANEL DISCUSSION OF CASE EXAMPLES

How do we identify DIF in clinical practice?

Why does DIF occur?

What do we do there is DIF?
SCALAR INVARIANCE CASE EXAMPLE

Do the items have equivalent meaning for both people?

Vigorous activity
Both Bill and Emma (45-year-old office worker) respond to the “Vigorous activities” question as “moderately difficult”. For Bill, this response reflects being able to walk a few blocks. However, for Emma who is relatively younger, the same response means being able to cross country ski for 1 hour a day.

**Type in response to question:**
Why are there differences in meaning for Bill and Emma when answering the same arthritis questionnaire?
**Scenario example**

**Clinician:** I noticed that you responded to the “Vigorous activities” as 3 out of 5. What kind of vigorous activities were you thinking of?

**Bill:** Well, I am no longer able to go on these hiking trips that I used to so vigorous activities now means being able to walk around my neighborhood block without losing my breath.

**Clinician:** I see, is that something that you would like to aim towards?

**Bill:** Yes, but I don’t know...it’s been a challenge just to walk a few blocks.

**Discussion points for panel:**

How would interpretation of vigorous activity scores differ for Bill vs. Emma?

What are the implications of this for your treatment plan?
• Ask whether the wording choice has the same meaning when applied to different patients. This can help to manage expectations of treatment outcomes.

• Due to other contextual factors (e.g., cultural, developmental, or personality), some patients may not interpret the items in the same way as other patients.
Do the underlying health domains have equal importance for both people?
**Scenario example**

**Clinician:** You rated your overall health rating as 7 out of 10; however, you rated your physical function as “quite a bit difficult”. Can you tell me more about this discrepancy?

**Bill:** As I mentioned previously, ever since I started to volunteer at the outdoor community club, my social health has played a larger part in maintaining my quality of life.

**Clinician:** I see, so social function has now become more important? And you find you are doing well, socially?

**Bill:** Yes, that’s right.

**Discussion points for panel:**
What would you expect a clinician to do with this information? How would this information influence the care provided?
• Ask whether identified health domains are equally important when applied to different people. This can help to focus treatment on domains important to the patient.

• Due to other contextual factors (e.g., cultural, developmental, or personality), some patients may not interpret health domains to be equally important.
CONFIGURAL INVARIANCE CASE EXAMPLE

Is the underlying health construct defined in the same way for both people?
**Scenario example**

**Clinician:** You rated your overall health rating as quite high of 7 out of 10 even though you rated your physical function as “quite a bit difficult”. Can you tell me about this discrepancy?

**Bill:** Well, I think of health more holistically to include emotional well-being.

**Clinician:** I see, so you don’t just define your health in terms of symptoms, physical functioning and social health?

**Bill:** You can say that.

**Discussion points for panel:**

What would you expect a clinician to do with this information? How would this information influence the care provided?
• Ask whether health domains are defined in the same way (e.g., symptoms, physical and social function) when applied to different people. This can help to provide holistic treatment that includes mental health and well-being.

• Due to other contextual factors (e.g., cultural, developmental, or personality), some patients may not define health in the same way.
SECTION 3
AGGREGATE LEVEL

Introduction to response shift and differential item functioning detection methods for aggregate data
Types of measurement invariance

- Scalar invariance/recalibration
- Metric invariance/reprioritization
- Configural invariance/reconceptualization

Methods of analysis:

- Detection methods
- Explanatory methods
- Adjustment methods

Response shift over time
## METHODS FOR EXAMINING RESPONSE SHIFT

<table>
<thead>
<tr>
<th>Design based methods</th>
<th>“Ask the person”</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Qualitative methods</td>
<td>• Interviews about how people interpret and respond to measurement questions (response processes)</td>
</tr>
<tr>
<td>• Then test</td>
<td>• Statistical comparison of previous and retrospective scores</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model based approaches</th>
<th>“Infer from the data”</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Latent variable methods</td>
<td>• Structural equation models (Oort 2005)</td>
</tr>
<tr>
<td></td>
<td>• IRT / Rasch (Guilleux et al, 2015)</td>
</tr>
<tr>
<td></td>
<td>• Latent class analysis</td>
</tr>
<tr>
<td>• Mixed-effects regression models</td>
<td>• Using residuals (Mayo et al, 2008)</td>
</tr>
<tr>
<td></td>
<td>• Using interactions (Bernhard et al 2001)</td>
</tr>
<tr>
<td>• Relative importance analysis</td>
<td>• Importance measures based on logistic regression and discriminant analysis (Lix et al, 2013)</td>
</tr>
<tr>
<td>• Classification/Data mining techniques</td>
<td>• Recursive partitioning (Li &amp; Rapkin, 2009)</td>
</tr>
<tr>
<td></td>
<td>• Random Forest (Boucekine et al, 2013)</td>
</tr>
</tbody>
</table>

### METHODS FOR EXAMINING DIFFERENTIAL ITEM FUNCTIONING

<table>
<thead>
<tr>
<th>Design based methods</th>
<th>“Ask the person”</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Qualitative methods</td>
<td>• Interviews about how different groups of people interpret and respond to measurement questions (response processes)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Model based approaches</th>
<th>“Infer from the data”</th>
</tr>
</thead>
</table>
| • Latent variable group-based methods | • Multiple-group confirmatory factor analysis  
• Multi-group item response theory/RASCH  
• Latent variable mixture models |
| • Regression-based methods | • Logistic regression differential item functioning analysis  
• Multiple Indicators Multiple Causes (MIMIC) analysis |

Measurement invariance methods for patient-reported outcomes

Latent variable approaches for comparing diverse people (differential item functioning) and across measurement occasions (response shift)
Discuss strategies to address response shift and differential item functioning in your practice (15 mins)
How do we address response shift and differential item functioning in clinical practice?
THANK YOU

For further resources, please visit:

HealthyQOL.com

Measurement Validity

This page focuses on validity and methods for measuring health and quality of life in diverse populations using patient-reported outcomes (PROs).

The resources address the following measurement validity questions:

- How do we ensure valid comparisons of different people?
- How do we ensure valid assessments of change over time?

- Introduction to Measurement Validity of Patient-Reported Outcomes
- Learn About Analytical Methods (forthcoming)
- Research & Publications by Our Team