

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

BACKGROUND

1

(20 min)

- Validity evidence for use of PROs to assess change and diverse people
- Introduction to response shift (RS) and differential item functioning (DIF)

INDIVIDUAL LEVEL

2

(40 min)

- Case studies regarding implications and solutions for clinical practice

AGGREGATE LEVEL

3

(10 min)

- Overview of aggregate level implications
- Introduction to analytical approaches

CONCLUSION

4

(20 min)

- Breakout discussion
- Further resources

MOTIVATION FOR THE WEBINAR

Statistical methods for PROs

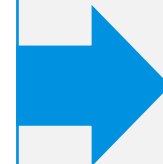
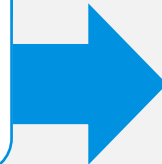
- Differential Item Functioning
- Response Shift

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

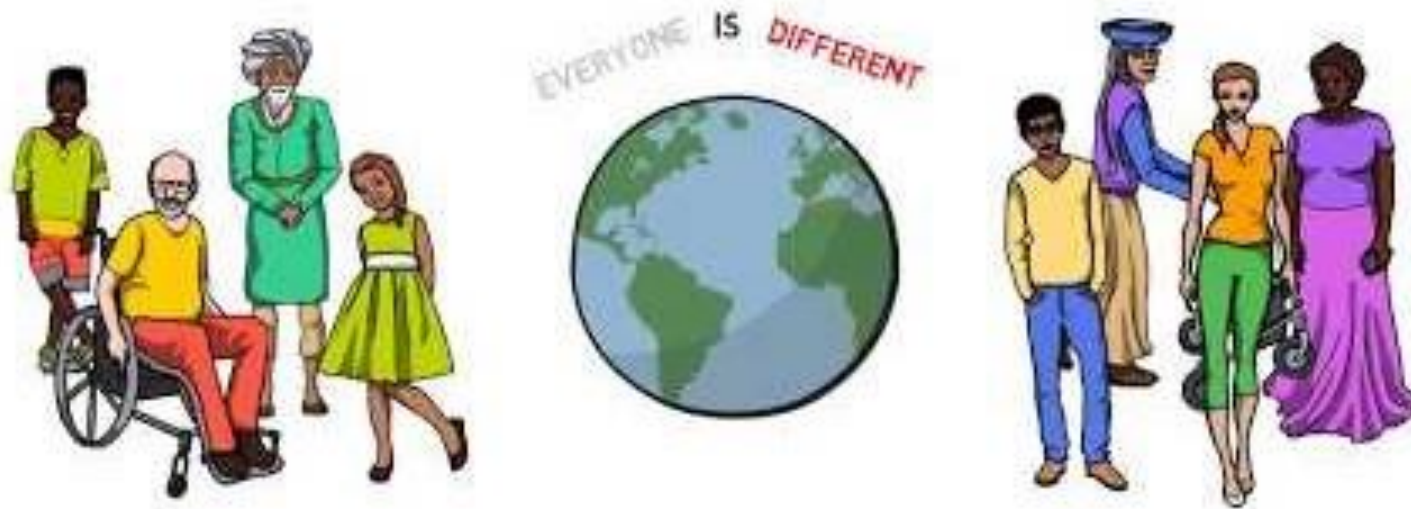
“Bill” is a 70-year-old who recently lost his wife suffering from advanced cancer and is also trying to cope with multiple chronic illnesses including arthritis. He is visiting his rheumatologist and completes a patient-reported outcome measure.

HOW HAS BILL'S HEALTH CHANGED OVER TIME & IN COMPARISON TO OTHER PATIENTS I HAVE SEEN?

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-XlquQu4g&list=PLKCOAAKeflJ8N8MN02zAZTm2fiFG4uNGO>

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

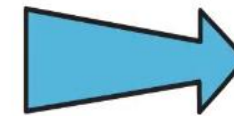


Over time



WHAT IS YOUR PERCEIVED HEALTH LEVEL FROM 0 TO 10?

5/10



7/10

MEASUREMENT VALIDITY

Compare
different
people

What evidence is needed to warrant
comparisons between people?

Compare
over time

What evidence is needed to warrant
comparisons over time?

**What are the value
implications, including
personal and societal
consequences, of using
PROs?**

VALIDITY EVIDENCE

CONSEQUENCES

MEASUREMENT VALIDITY

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

VALIDITY EVIDENCE

personal consequences

societal consequences

intended consequences

Unintended side effects

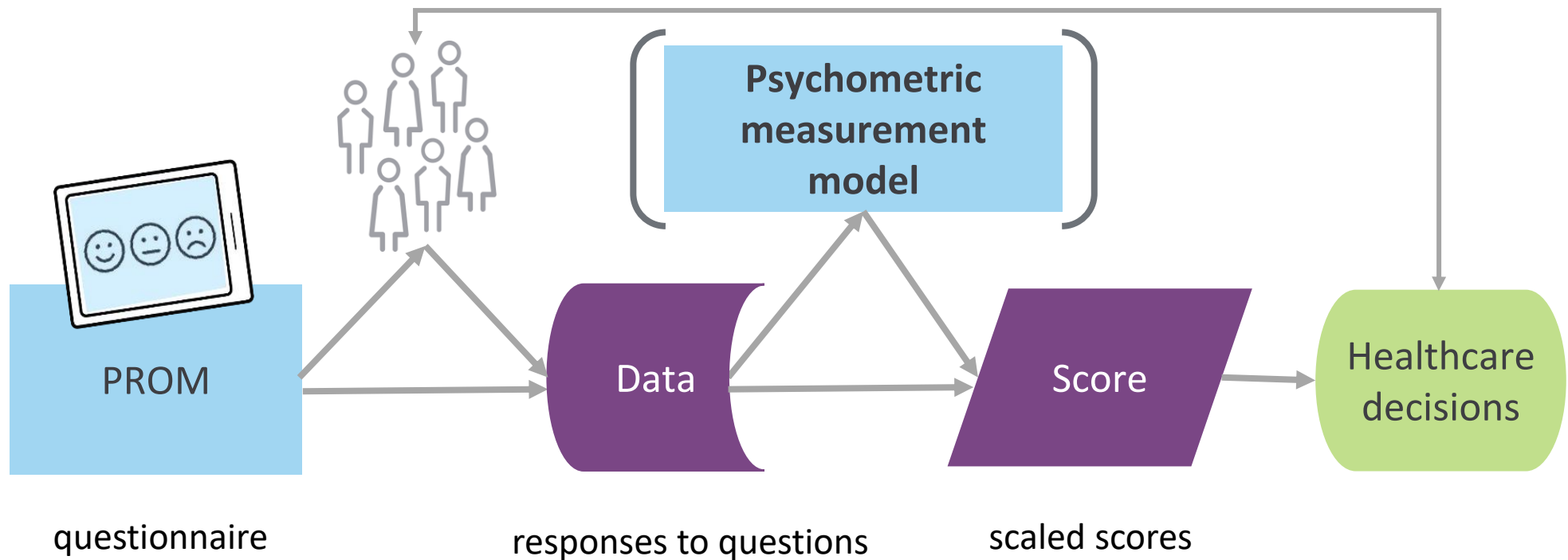
CONSEQUENCES



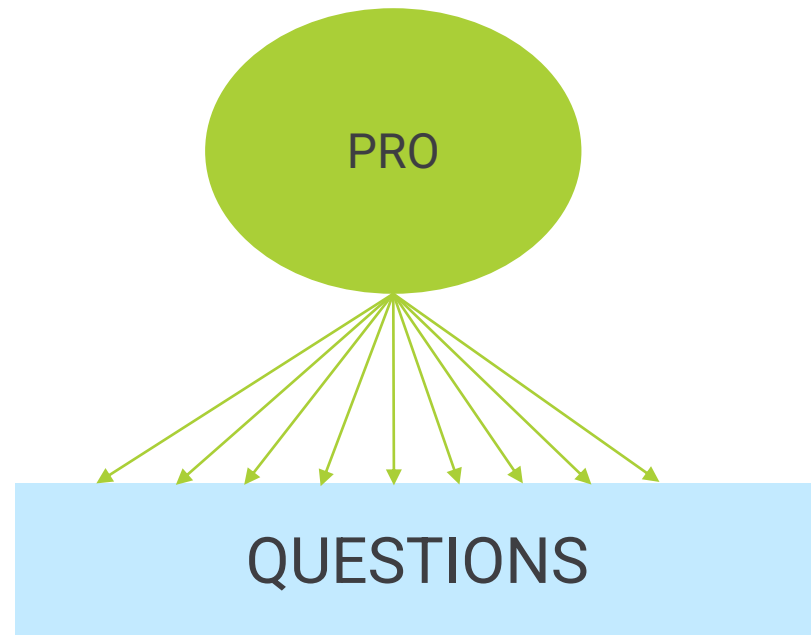
Sawatzky, R., Chan, E. K. H., Zumbo, B. D., Ahmed, S., Bartlett, S. J., Bingham, C., . . . Lix, L. M. (2017). Modern perspectives of measurement validation emphasize justification of inferences based on patient reported outcome scores. *Journal of Clinical Epidemiology*, 89, 154 – 159

Hubley, A.M., & Zumbo, B.D. (2011). Validity and the consequences of test interpretation and use. *Social Indicators Research*, 103, 219-230.

PRO MEASUREMENT



PRO MEASUREMENT MODEL



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

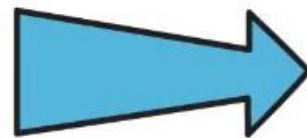
COMPARING OVER TIME

RESPONSE SHIFT



WHAT IS YOUR PERCEIVED HEALTH LEVEL FROM 0 TO 10?

5/10

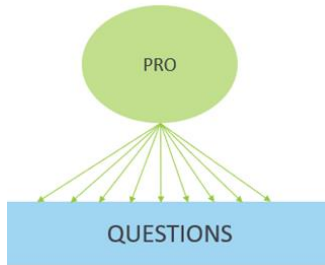


7/10

INVARIANCE OVER TIME



Time 1

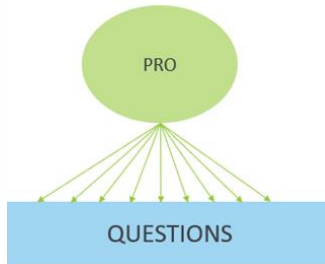


=

Time 2



Time 1



=

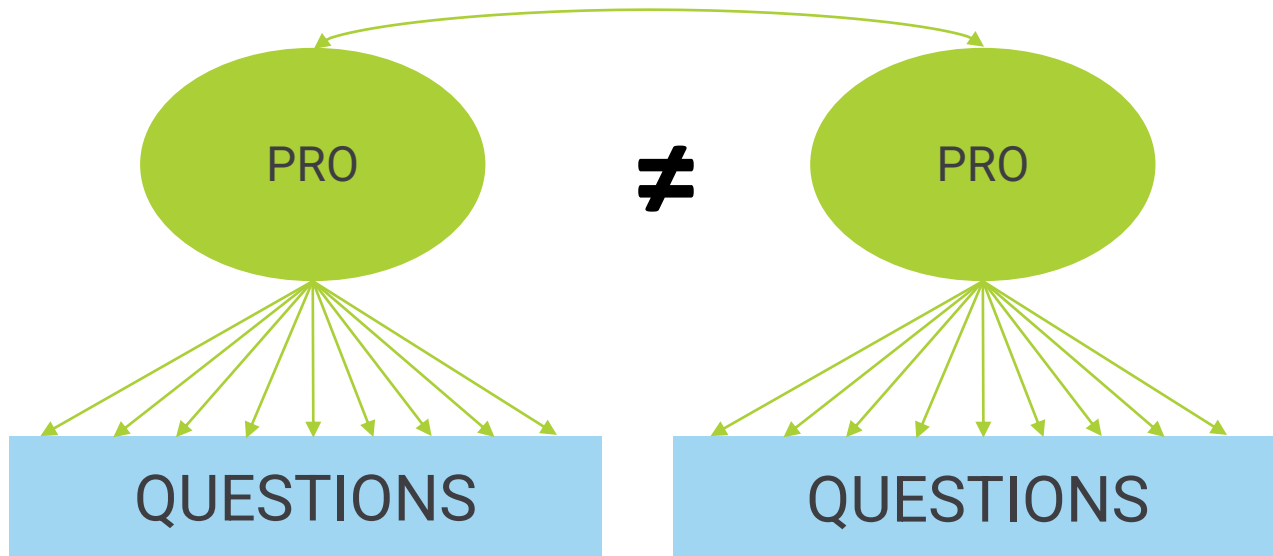
Time 2



RESPONSE SHIFT

Time 1

Time 2



Reconceptualization:
Consistency in how the construct is conceptualized

Reprioritization:
Consistency in relative importance of measurement indicators and construct

Recalibration: Consistency in scale of measurement

Response shift

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”:

Recalibration

Internal standards of measurement

Reprioritization

Relative importance of domains or items

Reconceptualization

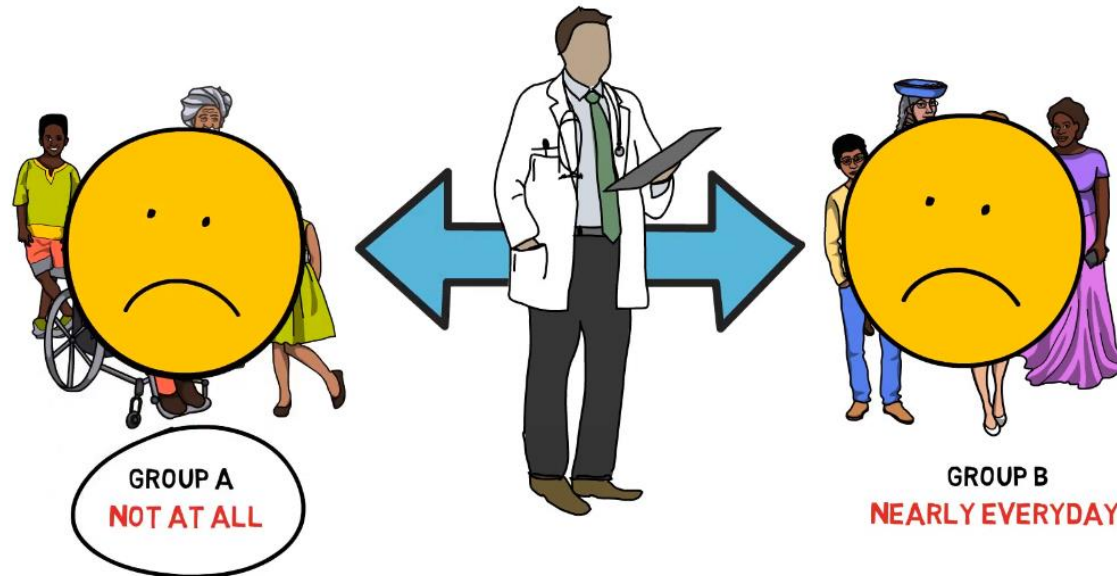
Definition of the target construct

COMPARING DIFFERENT PEOPLE

DIFFERENTIAL ITEM FUNCTIONING

IN THE PAST 7 DAYS, HOW OFTEN HAVE YOU FELT DOWN, DEPRESSED, OR HOPELESS?

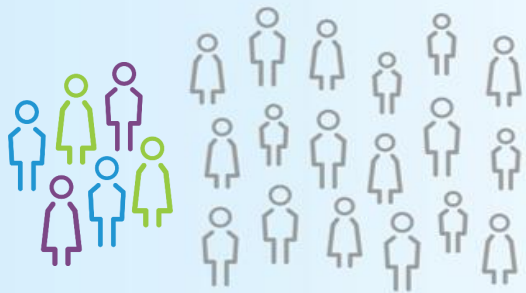
NOT AT ALL - SEVERAL DAYS - MORE THAN HALF THE DAYS - NEARLY EVERY DAY



INVARIANCE BETWEEN DIFFERENT PEOPLE



Individual Level



Aggregate Level

Person A



=

Person B



Group A

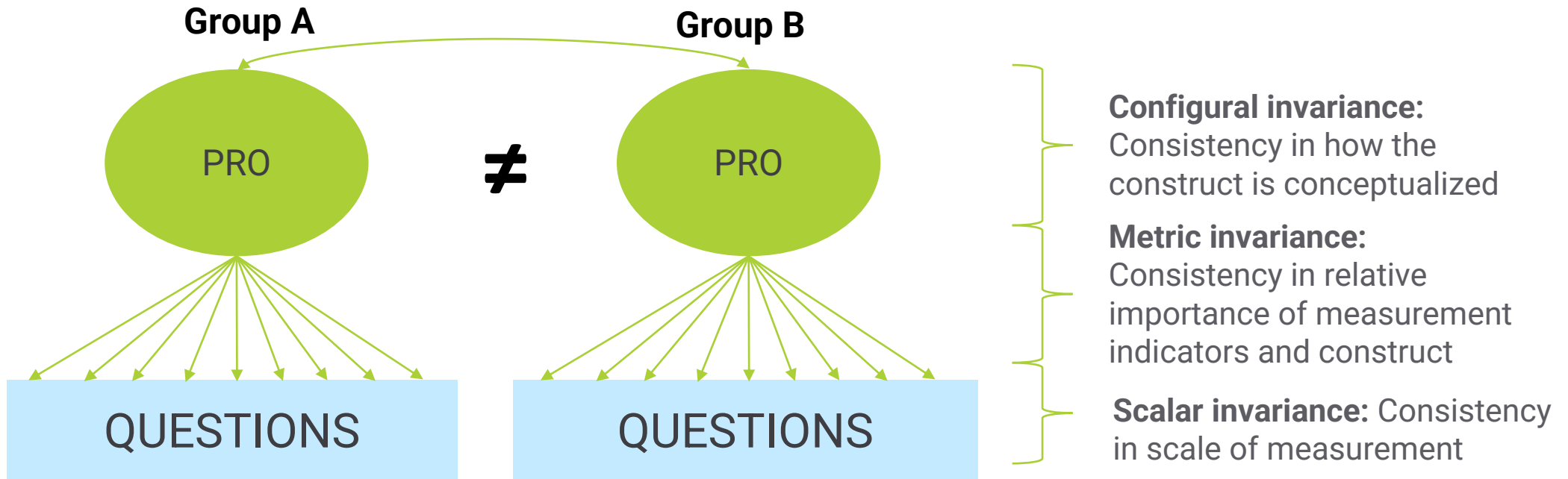


=

Group B



DIFFERENTIAL ITEM FUNCTIONING



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

MEASUREMENT INVARIANCE



**between
different people**

Differential item functioning

- Differences in how people interpret and respond to questions
- Threatens the comparability of scores across individuals or groups

over time

Response shift

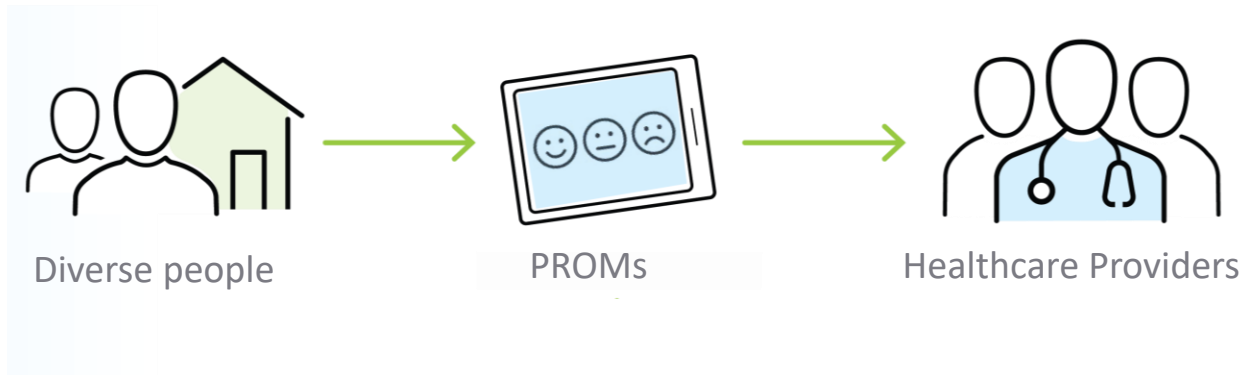
- An individual's frame of reference may change over time
- Threatens the comparability of scores over time

WHY IS ADDRESSING MEASUREMENT INVARIANCE IMPORTANT?

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



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.

CASE STUDY

REPORT OF
COMPLETED PRO:
ARTHRITIS HEALTH
QUESTIONNAIRE

Arthritis Health Questionnaire (hypothetical example)

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

Symptoms	Not at all	A little bit	Moderate	Quite a bit	Extremely						
1. Pain in joints	5	4	3	2	1						
2. Stiffness in joints	5	4	3	2	1						
3. Grating in joints	5	4	3	2	1						
How much difficulty did you have conducting activities due to your arthritis in the past week?											
Physical function	Not difficult at all	A little difficult	Moderately difficult	Quite a bit difficult	Extremely difficult						
4. Walking briskly	5	4	3	2	1						
5. Vigorous activities (e.g., lifting weights, running)	5	4	3	2	1						
Social function	Not at all limited	A little limited	Moderately limited	Quite a bit limited	Extremely limited						
6. Hobbies	5	4	3	2	1						
7. Relationships with friends/family	5	4	3	2	1						
Overall health	Very good				Very poor						
8. How would you rate your health overall?	10	9	8	7	6	5	4	3	2	1	0

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?

RECALIBRATION CASE EXAMPLE

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

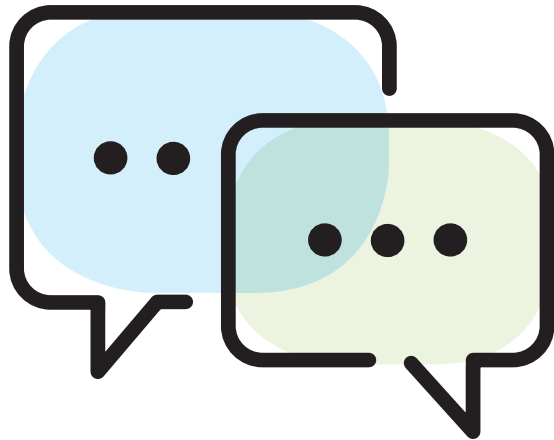
Clinician: How are you, Bill? Tell me on a scale from zero to ten where zero is worst health you can imagine and ten is the best.

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:

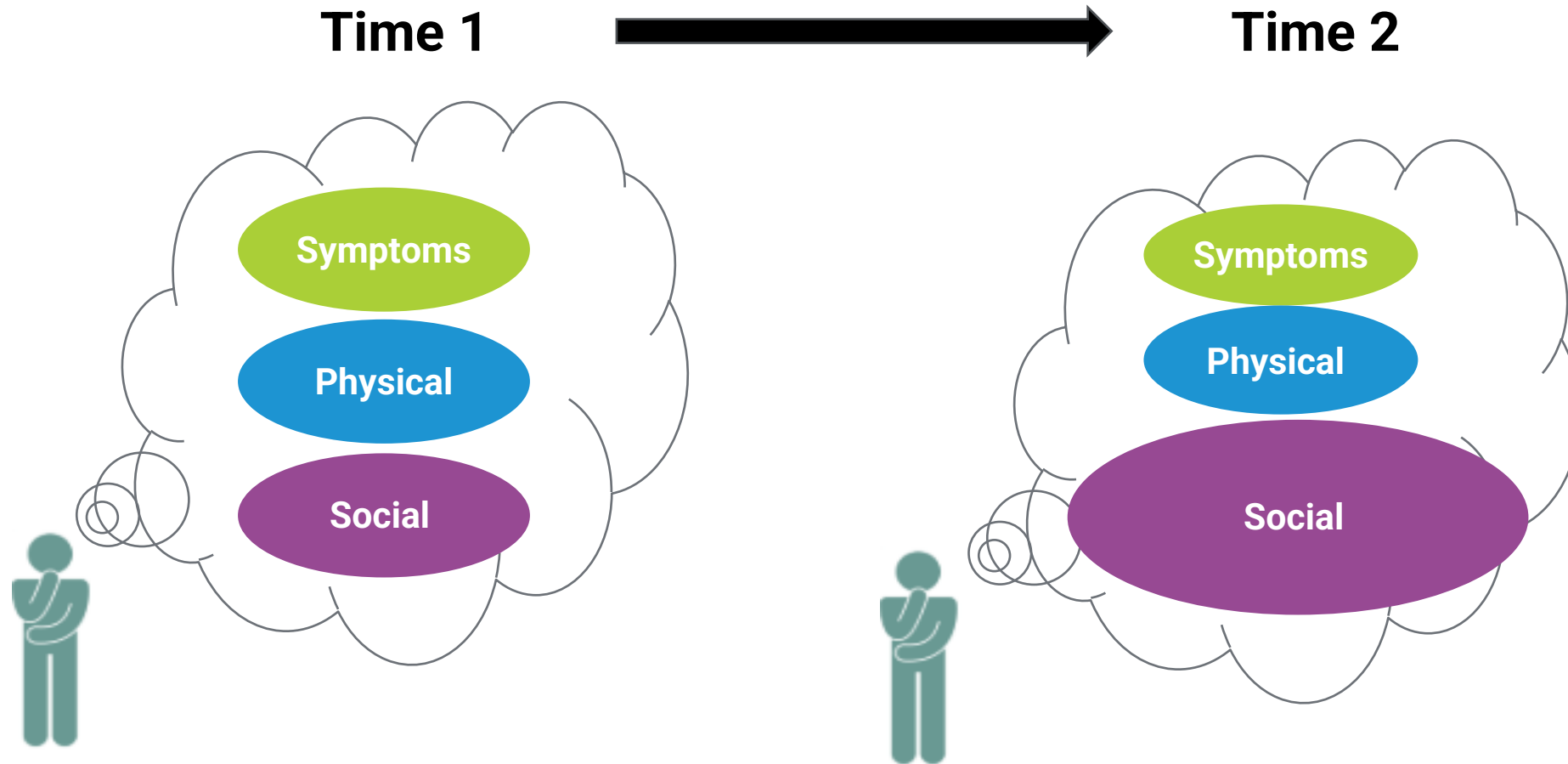
How would this information influence your interpretation of the current PRO score?

- 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.

POTENTIAL STRATEGIES AND
KEY POINTS

REPRIORITIZATION CASE EXAMPLE

Bill places more importance on certain health domains (e.g., maintaining social contact) over time



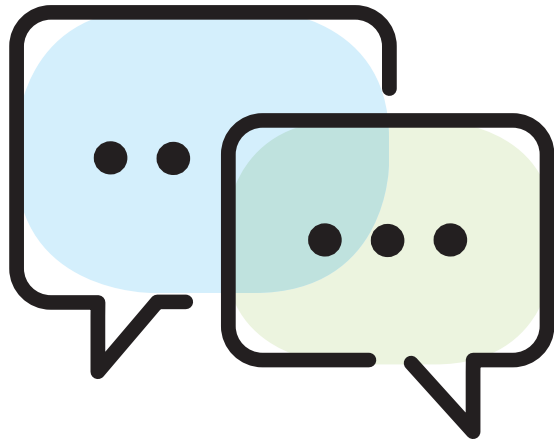
Scenario example

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:

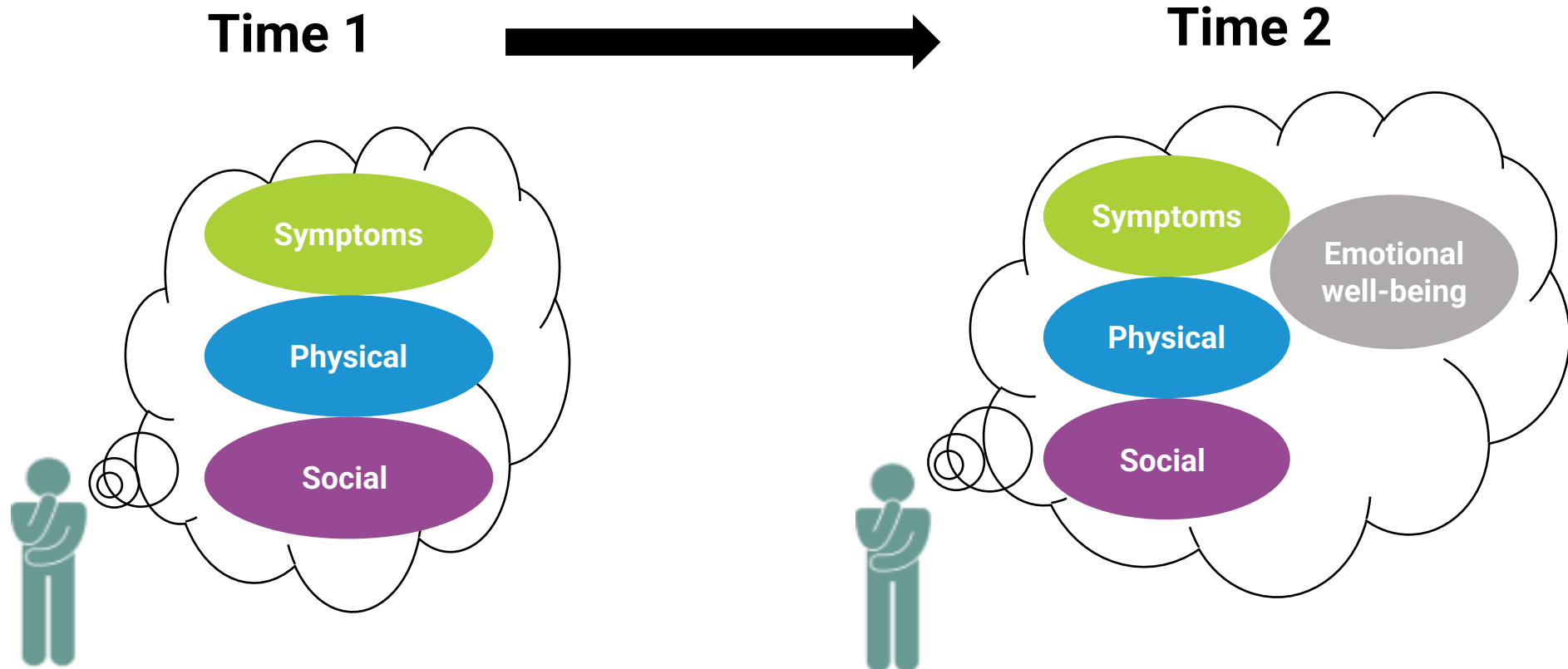
How would this information influence the course of follow-up treatment?

- 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.

POTENTIAL STRATEGIES AND
KEY POINTS

RECONCEPTUALIZATION CASE EXAMPLE

Bill has redefined his health over time



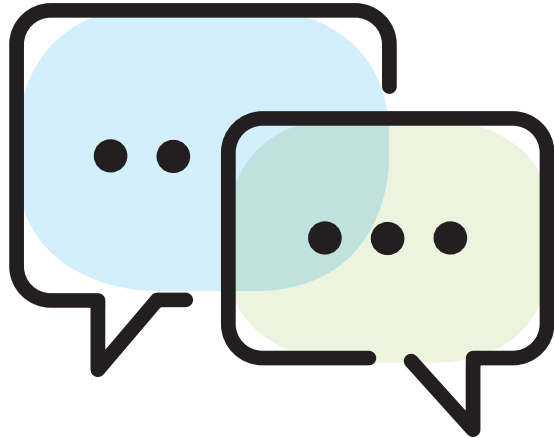
Scenario example

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.

POTENTIAL STRATEGIES AND
KEY POINTS

DIFFERENTIAL ITEM FUNCTIONING

**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?



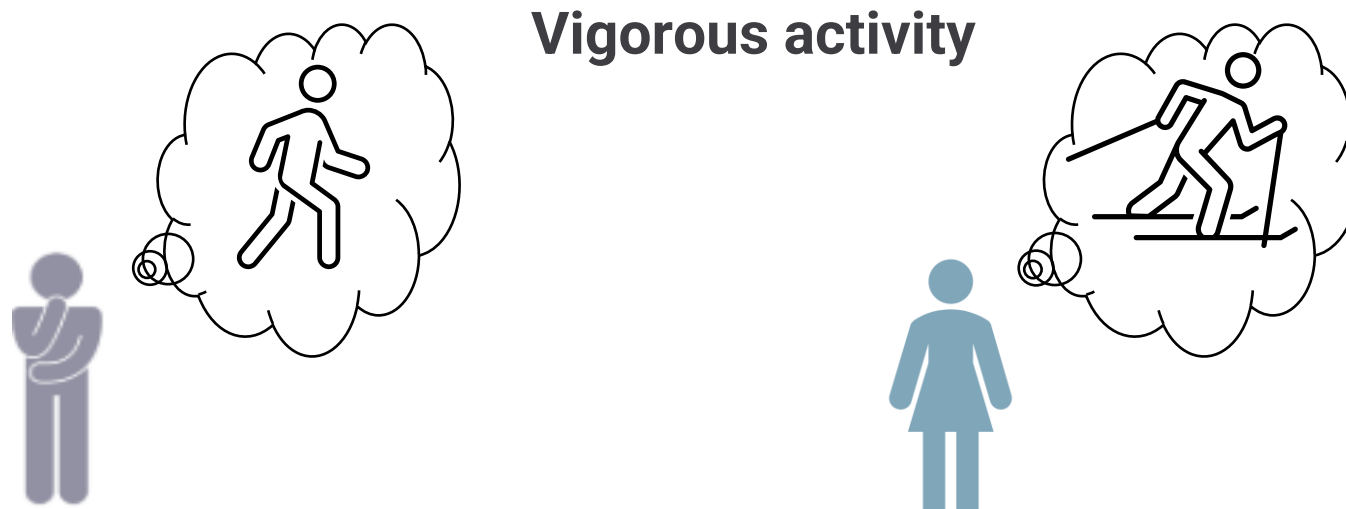
Why does DIF occur?



What do we do there is DIF?

SCALAR INVARIANCE CASE EXAMPLE

Do the items have equivalent meaning for both people?



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?

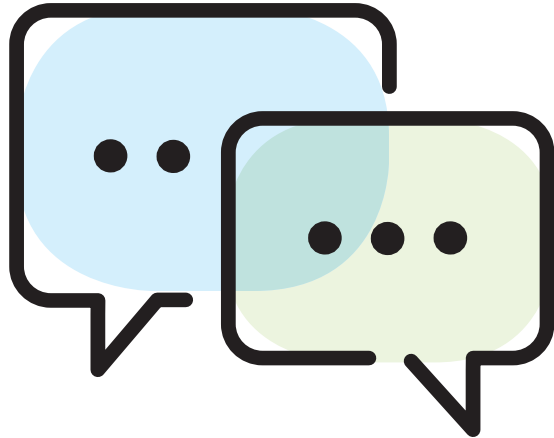
Scenario example

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?

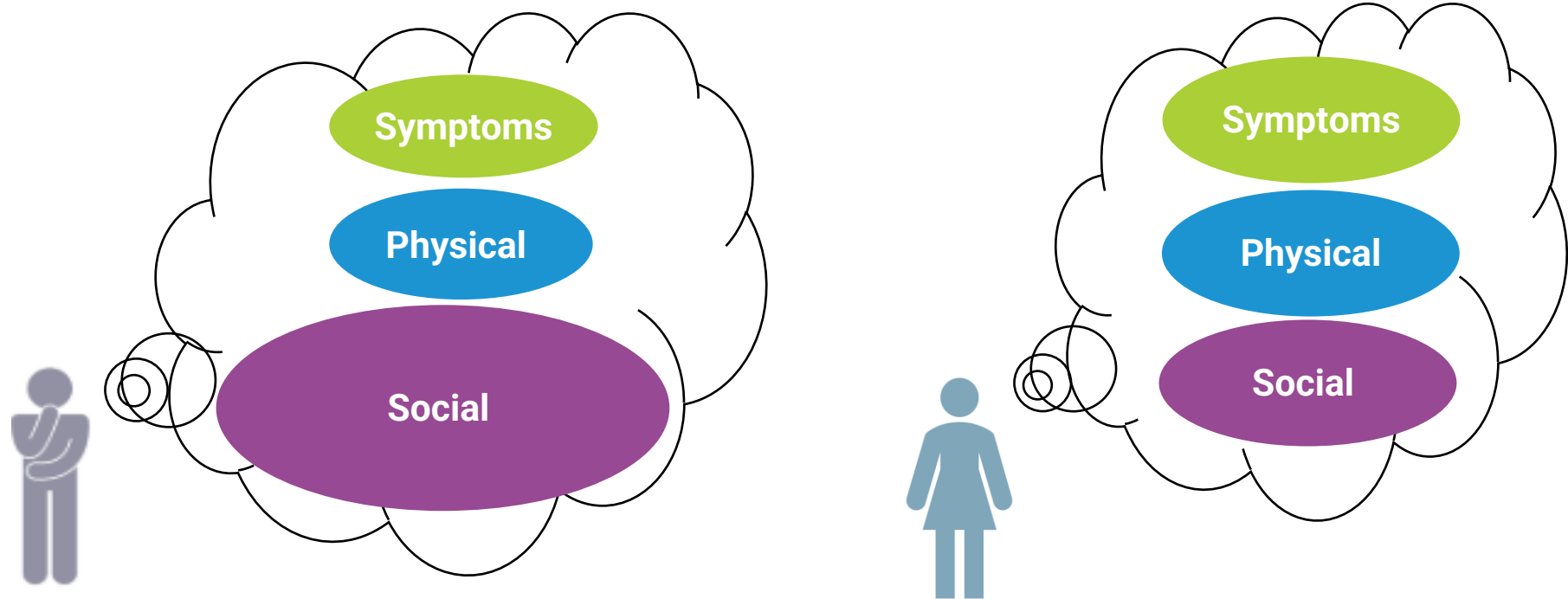
What are the implications of this for your treatment plan?

- 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.

POTENTIAL STRATEGIES AND
KEY POINTS

METRIC INVARIANCE CASE EXAMPLE

Do the underlying health domains have equal importance for both people?



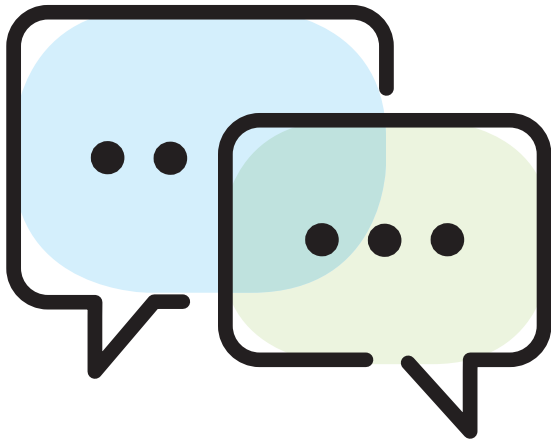
Scenario example

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?

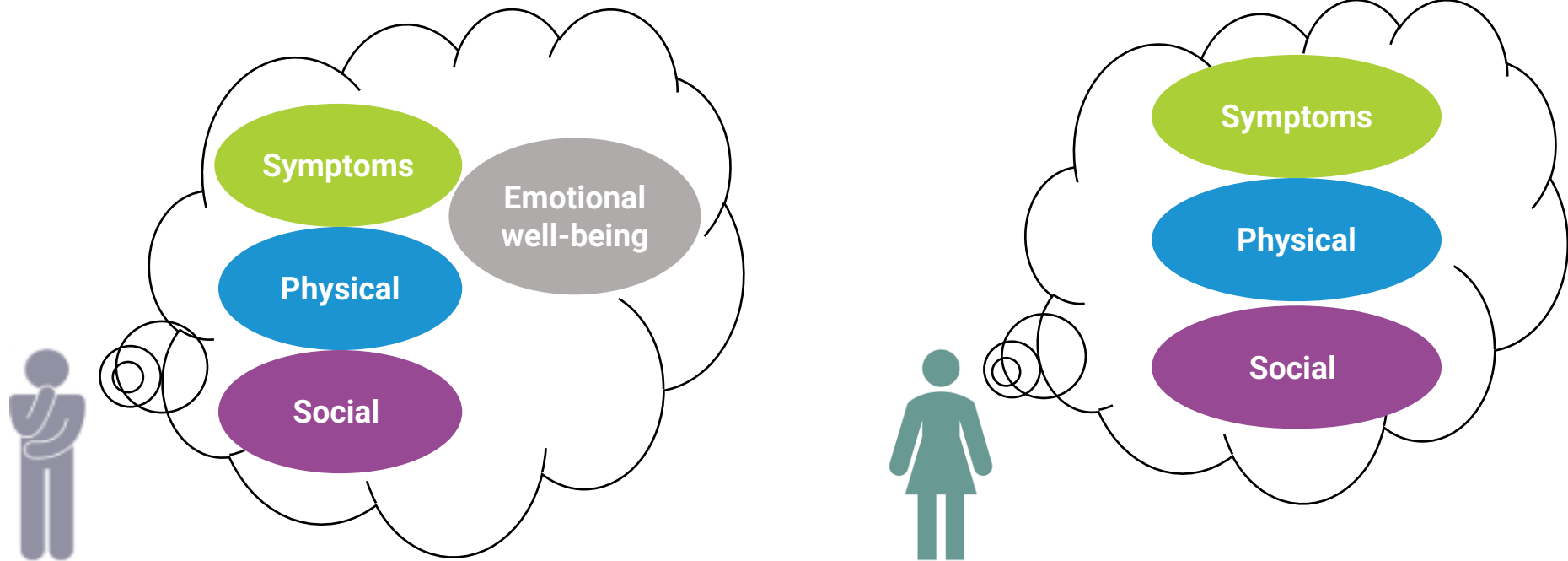
How would this information influence the care provided?

- 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.

POTENTIAL STRATEGIES AND
KEY POINTS

CONFIGURAL INVARIANCE CASE EXAMPLE

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



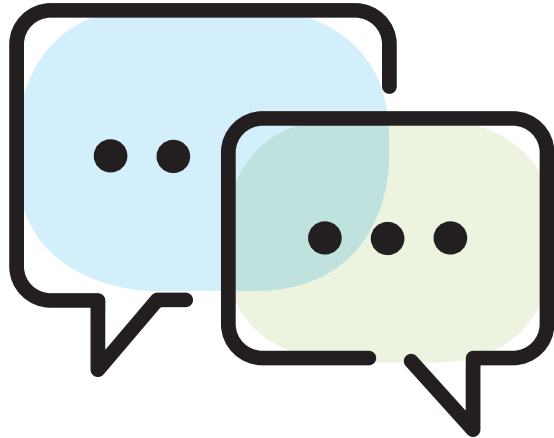
Scenario example

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.

POTENTIAL STRATEGIES AND
KEY POINTS

SECTION 3

AGGREGATE LEVEL

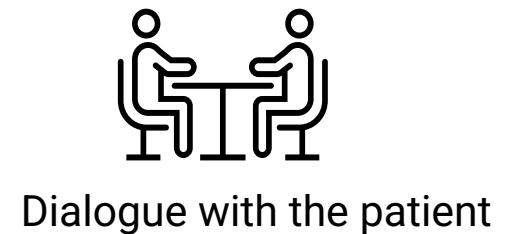
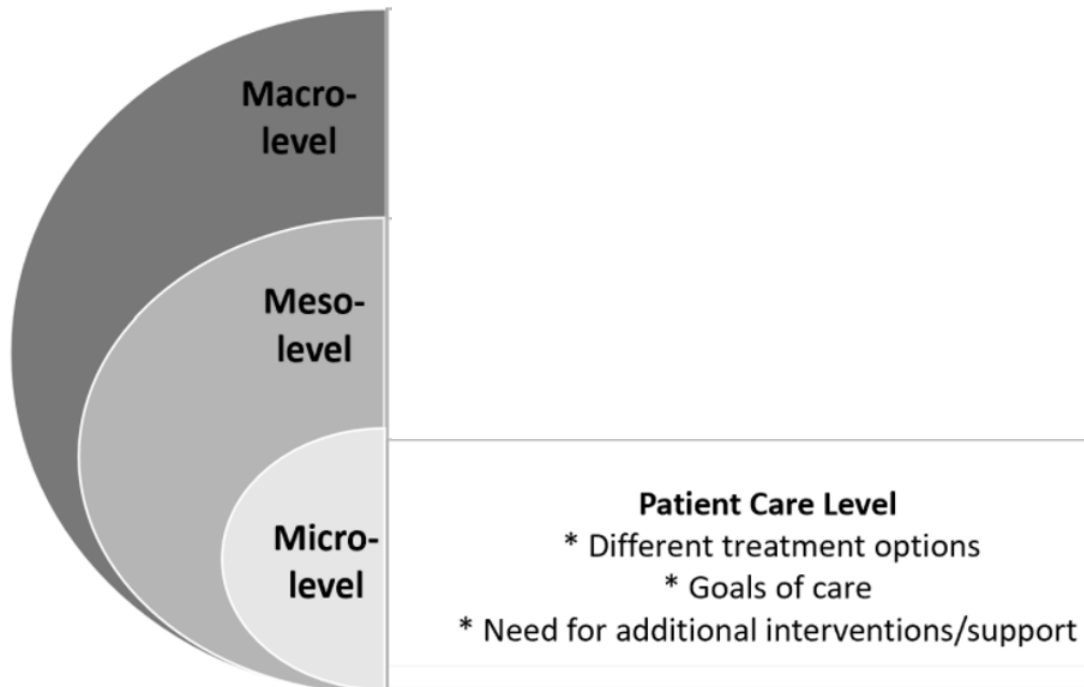
Introduction to response shift and differential item functioning detection methods for aggregate data

IMPLICATIONS OF RESPONSE SHIFT + DIFFERENTIAL ITEM FUNCTIONING

Levels of decision-making

Implications

Solutions



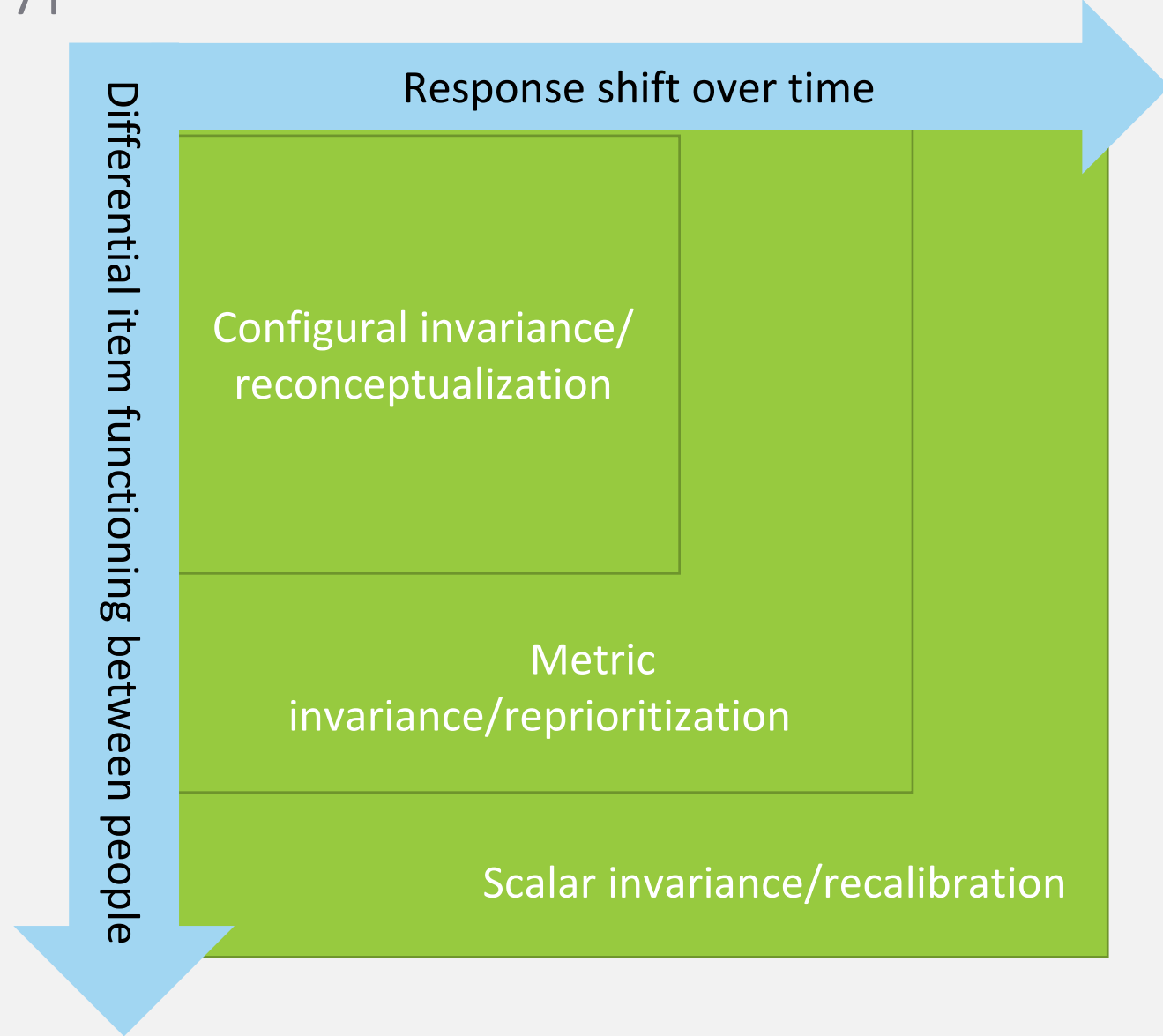
Types of measurement invariance

METHODS OF ANALYSIS

Detection methods

Explanatory methods

Adjustment methods



METHODS FOR EXAMINING RESPONSE SHIFT

Design based methods		"Ask the person"	
• Qualitative methods		• Interviews about how people interpret and respond to measurement questions (response processes)	
• Then test		• Statistical comparison of previous and retrospective scores	
Model based approaches		"Infer from the data"	
• Latent variable methods		• Structural equation models (Oort 2005)	
		• IRT / Rasch (Guilleux et al, 2015)	
		• Latent class analysis	
• Mixed-effects regression models		• Using residuals (Mayo et al, 2008)	
		• Using interactions (Bernhard et al 2001)	
• Relative importance analysis		• Importance measures based on logistic regression and discriminant analysis (Lix et al, 2013)	
• Classification/Data mining techniques		• Recursive partitioning (Li & Rapkin, 2009)	
		• Random Forest (Boucekine et al, 2013)	

METHODS FOR EXAMINING DIFFERENTIAL ITEM FUNCTIONING

Design based methods		“Ask the person”	
<ul style="list-style-type: none">Qualitative methods		<ul style="list-style-type: none">Interviews about how different groups of people interpret and respond to measurement questions (response processes)	
Model based approaches		“Infer from the data”	
<ul style="list-style-type: none">Latent variable group-based methods		<ul style="list-style-type: none">Multiple-group confirmatory factor analysisMulti-group item response theory/RASCHLatent variable mixture models	
<ul style="list-style-type: none">Regression-based methods		<ul style="list-style-type: none">Logistic regression differential item functioning analysisMultiple Indicators Multiple Causes (MIMIC) analysis	

Measurement invariance metho... 

<https://www.healthyqol.com/methods>

Slide Title

- Title
- Introduction
- Expository analysis
 - Title
 - Syntax
 - Data
 - Sample description
 - VR-36 General Health
- Background
 - Webinar
 - PROMs video
 - Latent variables
 - Measurement model
- Examples
 - Continuous items
 - Binary/ordinal items

Measurement invariance methods for patient-reported outcomes

Latent variable approaches for comparing *diverse people (differential item functioning)* and *across measurement occasions (response shift)*



SECTION 4 DISCUSSION

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?

DISCUSSION

THANK YOU

For further resources, please visit:

HealthyQOL.com

The screenshot shows the HealthyQOL.com website. At the top, there is a navigation bar with four items: 'What are QOL Assessments?', 'Applications in Healthcare', 'Measurement Methods', and 'About/Contact'. The main content area features the title 'Measurement Validity' in a large, bold font. Below the title, there is a paragraph: 'This page focuses on validity and methods for measuring health and quality of life in diverse populations using patient-reported outcomes (PROs).' This is followed by another paragraph: 'The resources address the following measurement validity questions:'. Below this, there is a bulleted list with two items: 'How do we ensure valid comparisons of different people?' and 'How do we ensure valid assessments of change over time?'. At the bottom of the page, there is a list of three expandable sections: 'Introduction to Measurement Validity of Patient-Reported Outcomes' (with a plus sign), 'Learn About Analytical Methods (forthcoming)' (with a plus sign), and 'Research & Publications by Our Team' (with a minus sign).

What are QOL Assessments? Applications in Healthcare Measurement Methods About/Contact

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 -