Machine-Learning Methods for Differential Item Functioning in Patient-Reported Outcomes

Facilitators

Dr. Yuelin Li
Attending,
Memorial Sloan Kettering Cancer Center
New York, USA
e-mail: liy12@mskcc.org

Dr. Lisa M. Lix
Professor and Canada Research Chair
University of Manitoba
Winnipeg, Canada
e-mail: lisa.lix@umanitoba.ca

Dr. Tolulope T. Sajobi
Associate Professor of Biostatistics
University of Calgary
Calgary, Canada
e-mail: ttsajobi@ucalgary.ca

The facilitators have no conflicts of interest with respect to the material that is presented
in this workshop

Acknowledgements

3

Purpose and Objectives

- **Purpose:** To introduce data-driven methods to detect differential item functioning in patient-reported outcomes

- **Objectives:**
  1. To examine machine-learning models to explore and detect differential item functioning in high-dimensional data.
  2. To describe the types of data and research problems that will benefit from the application of machine-learning models for detection of differential item functioning.
  3. To demonstrate the implementation of machine-learning methods using existing software packages, with a particular emphasis on R software.

Time	Topic	Facilitator
9:00 – 9:15	Welcome and Overview	Lisa Lix
9:15 – 9:40	Machine-Learning Methods for Patient-Reported Outcomes Data	Yuelin Li
9:40 – 10:05	Item Response Theory Methods for Detection of Differential Item Functioning	Tolulope Sajobi
10:05 – 10:25	Item-Focused Machine-Learning Models for Detection of Differential Item Functioning	Lisa Lix
10:25 – 10:40	Break	
10:40 – 10:55	**Continued:** Item-Focused Machine-Learning Models for Detection of Differential Item Functioning	Lisa Lix
10:55 – 11:20	Person-Centered Polytomous IRT for Detection of Differential Item Functioning	Tolulope Sajobi
11:20 – 11:45	Extending Machine-Learning Methods to Detect Response Shift in Patient-Reported Outcomes Data	Tolulope Sajobi & Yuelin Li
11:45 – 12:00	Concluding Remarks Q&A	Lisa Lix

Workshop Materials

https://www.ucalgary.ca/research/person-centered-methods-lab/research/resources/isoqol

Content:

- Lecture notes with list of relevant references

- R scripts

- Dataset for case example

Dataset for Case Example

- Regional joint replacement registry from Canada;

- 1391 patients having a total hip replacement

- Patients had complete responses on the SF-12 (version 2) physical health (PH) and mental health (MH) component items

- Females: 51.5%

- Age: 17 years to 92 years; mean of 64.7 years (SD 11.3)

Physical Health Component Items

Item: Label	Response options, n (%)				
	Excellent	**Very Good**	**Good**	**Fair**	**Poor**
P1: General health	24 (1.7)	138 (9.9)	602 (43.3)	512 (36.8)	115 (8.3)
	Limited a lot	**Limited a little**	**Not limited at all**		
P2: Limited in moderate activity	960 (69.0)	354 (25.4)	77 (5.5)		
P3: Climbing several flights	1014 (72.9)	308 (22.1)	69 (5.0)		
	All of the time	**Most of the time**	**Some of the time**	**A little of the time**	**None of the time**
P4: Accomplished less (physical health)	543 (39.0)	482 (34.7)	237 (17.0)	89 (6.4)	40 (2.9)
P5: Limited in work and other activities	555 (39.9)	495 (35.6)	258 (17.1)	71 (5.1)	32 (2.3)
P6: Pain interference with normal work	18 (1.3)	95 (6.8)	261 (18.8)	632 (45.4)	385 (27.7)

Mental Health Component Items

Item: Label	Response options, n (%)				
	All of the time	**Most of the time**	**Some of the time**	**A little of the time**	**None of the time**
M1: Accomplished less (emotional problems)	145 (10.4)	248 (17.8)	325 (23.4)	283 (20.3)	390 (28.0)
M2: Did work or other activities less carefully than usual (as a result of any emotional problems)	135 (9.7)	212 (15.2)	313 (22.5)	301 (21.6)	430 (30.9)
M3: Felt calm and peaceful	74 (5.3)	229 (16.5)	401 (28.8)	605 (43.5)	82 (5.9)
M4: Energy level	195 (14.0)	394 (28.3)	484 (34.8)	281 (20.2)	37 (2.7)
M5: Felt downhearted and depressed	35 (2.5)	94 (6.8)	389 (28.0)	466 (33.5)	407 (29.3)
M6: Physical health or emotional problems interfered with social activities	136 (9.8)	239 (17.2)	434 (31.2)	278 (20.0)	304 (21.9)

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

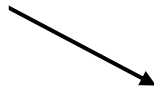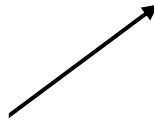| Scalar invariance | Internal standards of measurement |
| Metric invariance | Relative importance of domains or items |
| Configural invariance | Definition of the target construct |

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