ISSN 2282-0930 • Epidemiology Biostatistics and Public Health - XIII[th] SISMEC Congress - Vol. 20 Suppl. 1

Data Harmonization of Psychosocial Questionnaires across Population Cohorts: A Differential Item Functioning Analysis

Giusti Emanuele Maria[(1)], *Ferrario Marco Mario*[(1)], *Veronesi Giovanni*[(1)]

(1) EPIMED Research Center, Department of Medicine and Surgery, University of Insubria, Varese, Italy

CORRESPONDING AUTHOR: *Giusti Emanuele Maria, emanuelemaria.giusti@uninsubria.it*

INTRODUCTION

Data harmonization is the process of achieving comparability of similar measures collected by separate cohorts by aligning their definitions and measurement formats [1]. In epidemiologic research, and in particular in predictive modelling for chronic non-communicable diseases, it may facilitate the identification of derivation and external validation cohorts [2]. This process may be particularly challenging when psychosocial variables are amongst the predictors of interest, as different self-report questionnaires measuring the same psychosocial variable might capture different aspects, and even small differences in item wording affect responses [3]. Therefore, measurement invariance; i.e.; whether respondents from different populations (e.g.; at different time periods, or different cultures) with the same latent trait (e.g. depression) level respond similarly, should be preliminarily established to avoid bias in subsequent analyses [4]. Currently, there is little guidance on how to assess this property for data harmonization purposes. A promising framework is Item Response Theory (IRT), a probabilistic approach for modelling the relationship between a latent trait and observed item responses [5].

We designed an international project pooling Italian and German cohorts with recruitment time spanning over a 10-year period, with the aim to identify a restricted set of psychosocial items able to increase the predictive power of established risk prediction models for Cardiovascular Diseases (CVD). Here, we leverage IRT to assess Differential Item Functioning (DIF) in harmonised items, thereby testing whether item characteristics are invariant across the Italian cohorts.

AIMS

To harmonise two questionnaires used in different cohorts to assess depressive symptoms and to evaluate the measurement invariance of the harmonised items across cohorts.

METHODS

The MONICA Brianza study includes three independent cohorts recruited from the Brianza population, Lombardy region, over a 10-year period (MONICA87: 1986–1987; MONICA90: 1989–1990; MONICA93: 1993–1994). Each cohort includes a 10-year age- and gender-stratified random sample of the target 25- to 64-years old population. Overall, 4932 individuals participated to the study (69% of invited). Depressive symptoms were measured in the MONICA87 cohort using the Beck Depression Inventory (BDI) [6], a 22-item instrument with a 0 (absence of symptom) to 4 (severe symptom) response format, and in the MONICA90-93 cohorts using a 14-item version of the Maastricht Vital Exhaustion Questionnaire (EX) [7], coded on a scale of 0 ("No") – 1 ("Not sure") – 2 ("Yes").

PROTOCOL STEPS FOR DATA HARMONIZATION AND ANALYSIS

We developed an a priori protocol for data harmonization based on the following steps. First, the content of the BDI and EX items were analysed to select item pairs measuring the same depressive symptoms. Second, for each item pair, the response format was dichotomised ensuring that the collapsed response categories have the most similar possible frequencies of endorsement across cohorts. Third, the resulting scale was analysed through a unidimensional Confirmatory Factor Analysis and, in case of departures from unidimensionality, a minimum residual exploratory factor analysis (EFA) to identify subscales. Fourth, a 2-parameter logistic IRT model was fitted on the resulting subscales to estimate for each item a discrimination parameter (ability to differentiate between individuals with close levels of the latent trait), and a difficulty parameter (level of the latent trait at which the item has a 50% probability of endorsement). Item fit was evaluated through S-X2 statistics [8]. Finally, DIF analyses were conducted to evaluate whether

DOI: 10.54103/2282-0930/29218

item parameters were invariant across cohorts through Likelihood ratio tests comparing nested models with increasing equality constraints on discrimination and difficulty parameters.

RESULTS

525 individuals in the MONICA87 and 48 individuals in the MONICA90-93 cohorts were excluded due to unavailability of the questionnaires at baseline date. The final sample (51% female, mean age ±SD: 45.9±11.3) included 1134 subjects from MONICA87 and 3225 from MONICA90-93.

The item content analysis led to the selection of 10 item pairs, which were used to form a harmonised scale of depressive symptoms (Dep). One item pair was discarded due to overlap with another item pair, both assessing sleep disturbance (tetrachoric correlation: 0.78). Standard fit indices from the CFA suggested departures from unidimensionality. EFA identified two latent variables, one underpinning a "Neurovegetative and arousal disturbance" subscale (4 items; e.g., sleep problems; loss of energy), and one an "Affective and cognitive disturbance" subscale (5 items; e.g., suicidal thoughts). Both subscales met the assumptions for the IRT model and all items showed adequate model fit. DIF analysis (Figure 1) confirmed measurement invariance for six items (e.g.; Dep03, panel A). Items Dep02 (panel B) and Dep05 (panel C) exhibited different discrimination and difficulty parameters across cohorts, suggesting non-uniform DIF. Finally, the different difficulty parameters for Dep10 (panel D), indicating uniform DIF, suggests that people from different cohorts with the same latent trait have different probability to endorse the item.

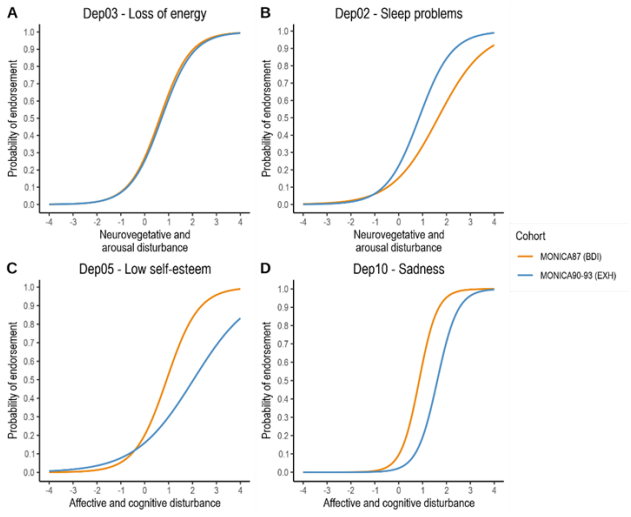

Figure 1. Item Characteristic Curves (ICC) representing the probability to endorse items across the relevant latent trait (mean=0, SD=1) across cohorts. Panel A: Example of the ICC of an item without DIF. Panel B, C, D: ICC of items showing DIF

CONCLUSIONS

The harmonization of two depression questionnaires identified 6 items with measurement invariance and revealed DIF in items that could have biased subsequent analyses if undetected. In epidemiological research, measurement invariance of psychosocial questionnaires should be thoroughly checked using comprehensive methods. Analysis of DIF within an IRT framework offers a promising approach, and can be further evaluated for cross-cultural evaluations.

REFERENCES

1. Fortier I, Raina P, Van den Heuvel ER, et al. Maelstrom Research guidelines for rigorous retrospective data harmonization. International Journal of Epidemiology 2017;46:103–5.
2. Veronesi G, Gianfagna F, Giampaoli S, et al. Validity of a long-term cardiovascular disease risk prediction equation for low-incidence populations: the CAMUNI-MATISS Cohorts Collaboration study. Eur J Prev Cardiol 2015;22:1618–25.
3. Steinmann I, Sánchez D, van Laar S, et al. The impact of inconsistent responders to mixed-worded scales on inferences in international large-scale assessments. Assessment in Education: Principles, Policy & Practice 2022;29:5–26.
4. Protzko J. Invariance: What Does Measurement Invariance Allow Us to Claim? Educ Psychol Meas 2024:00131644241282982.
5. DeMars C. Item response theory. Oxford University Press; 2010.
6. Beck AT, Ward CH, Mendelson M, et al. An Inventory for Measuring Depression. Archives of General Psychiatry 1961;4:561–71.
7. Appels A. Psychological prodromata of myocardial infarction and sudden death. Psychother Psychosom 1980;34:187–95.
8. Orlando M, Thissen D. Likelihood-based item-fit indices for dichotomous item response theory models. Applied Psychological Measurement 2000;24:50–64.

DOI: 10.54103/2282-0930/29218