

O48 Semantic analysis of the PROMIS-29

Joel Alcantara¹, Ryan Whetten²

¹ICPA, Media, USA. ²University of Avignon, Avignon, France

Objective: The PROMIS-29 has been validated through various statistical means (i.e., convergent and discriminant analyses, known-groups validity, and test-retest reliability, factor analysis) to assess its reliability and validity. Another approach to validating patient reported outcomes (PROs) is semantic validation where the meaning of words and phrases are scrutinized within a survey to ensure that survey questions are understood as intended, and that responses accurately reflect the intended construct being measured. The aim of this study was to perform a semantic analysis of the PROMIS-29 using recent methods developed in the field of Natural Language Processing (NLP).

Methods: A sentence transformer was applied to each item of the PROMIS-29 to capture their semantic meaning, including the relationships between words and the context in which they appear. A sentence transformer is a neural network language model trained to produce high dimensional vector representations (or embeddings) of sentences (rather than individual words) to preserve their contextual meaning and significantly enhance the performance of NLP applications for accurate and meaningful text analysis. The PROMIS-29 items in vector form were then analyzed using K-means cluster analysis.

Results: The k means cluster analysis revealed four distinct clusters. First, a mental health cluster, composed mostly of items from the Anxiety and Depression subdomain of the PROMIS-29. Second, a cluster containing items from Sleep Disturbance. Third, a cluster representing items with Pain Interference and Pain Intensity. Lastly, we observed a cluster of items from Physical Function and Ability to Participate in Social Roles and Activities. Items from the Fatigue subdomain do not form a distinct cluster but instead were distributed with clusters from Pain Interference and Sleep Disturbance. This highlights the semantic relationship between Fatigue, Pain Interference and Sleep Disturbance.

Conclusions: The wording of questionnaire items can affect a respondent's interpretation and subsequently their responses. We analyzed the PROMIS-29 questionnaire from a semantic perspective, analyzing the similarities of item using a language model. We believe these results to be insightful and potentially provide an additional if not new methods of analyzing the PROMIS-29 and other survey instruments.