Bull. Aust. Math. Soc. **105** (2022), 173–174
doi:10.1017/S0004972721000769

WEIGHTED INTER-RATER AGREEMENT MEASURES FOR ORDINAL OUTCOMES

QUOC DUYET TRAN

(Received 13 August 2021; first published online 1 October 2021)

2020 *Mathematics subject classification*: primary 62C10; secondary 60J22, 65C05.

Keywords and phrases: applied statistics, Bayesian approaches, inter-rater agreement, agreement measures, grey zone effects.

The measurement of the degree of agreement among different assessors, which is called inter-rater agreement, is of critical importance in the medical and social sciences. Inter-rater agreement is measured by using different statistics for various types of data. In this work, inter-rater agreement is considered for an ordinal scale with two assessors. For an ordinal scale, agreement measures are used along with weight functions based on the values of scores representing the structure of hierarchy among the levels of the ordinal row and column variables. The choice of row and column scores impacts the estimated degree of agreement. The weighted agreement measures are prone to the unbalanced structures of the table and/or the existence of grey zones in the agreement table owing to the assessment behaviour of the raters.

In this study, we focus on the impact of such grey zones on the accuracy of inter-rater agreement measures for the ordinal outcomes. While there are altered versions of the inter-rater agreement measures for the ordinal tables, there is no explicit agreement about the use of a standardised method. In an inter-rater agreement study, where two raters score different facets of the subject of interest or have different degrees of expertise, there might be a grey zone among the levels of the categories. A grey zone may cause misrepresentation of the degree of consensus among raters and impact the decisions made based on the inter-rater agreement measures. In this context, it is important to know how the presence of a grey zone influences the inter-rater agreement measures to use the most reliable measure of agreement against the grey zones.

There are many combinations of agreement measures and weight functions proposed in the literature. However, the reliability of an agreement measure and weight

Thesis submitted to RMIT University in December 2020; degree approved on 10 March 2021; senior supervisor Haydar Demirhan, joint supervisor Anil Dolgun.

function combination against the grey zones has not been thoroughly studied in the literature. There is a lack of a general definition of a grey zone in an agreement table and no methodology developed for random generation of agreement tables in Monte Carlo simulation studies. Estimation of weights from the agreement table to compute inter-rater agreement measures has not been considered in the literature.

The thesis starts with the introduction in Chapter 1, where we give an overview of inter-rater agreement studies in the literature and describe our research questions, objectives and contributions. Next, in Chapter 2, we provide general information on inter-rater agreement measures in both classical and Bayesian approaches and outline Markov Chain Monte Carlo methods used in this thesis. In Chapter 3, we assess the accuracy of the inter-rater agreement measures when there is no grey zone and the cell counts are distributed across the cells of the agreement table in an unbalanced pattern (unbalanced table structure) by considering a large number of agreement measure and weight function combinations in a Monte Carlo setting. Then, in Chapter 4, we propose a formal definition for grey zone context and a method for the random generation of agreement tables with a grey zone in Monte Carlo studies. We extend our first Monte Carlo study for the tables having grey zones to identify reliable measures of agreement against the existence of grey zones. In Chapter 5, we propose a Bayesian approach to estimate the row and column scores of the agreement tables from the data and compute the weights based on the scores reflecting the information on the hierarchy among the ordinal levels. We illustrate the details of prior specifications and the impact of this approach using a real agreement table with grey zones. We also conduct a Monte Carlo simulation study to assess the accuracy of the proposed approach and compare it to the classical approach for the weighted agreement measures. By the proposed approach, in Chapter 6, we conclude that our studies improve the accuracy of the weighted agreement measures and mitigate the impact of the grey zones in the estimation of the strength of agreement between two raters in inter-rater agreement studies.

Some of this research has been published in [1, 2].

References

[1] D. Tran, A. Dolgun and H. Demirhan, 'Weighted inter-rater agreement measures for ordinal outcomes', *Comm. Statist. Simulation Comput.* **49**(4) (2020), 989–1003.

[2] Q. Tran, A. Dolgun and H. Demirhan, 'The impact of grey zones on the accuracy of agreement measures for ordinal tables', *BMC Med. Res. Methodol.* **21** (2021), Article no. 70, 19 pages.

QUOC DUYET TRAN, An Giang University, VNU-HCM,
Long Xuyen City, An Giang Province, 076, Vietnam
and
School of Science, RMIT University, Melbourne, Victoria 3000, Australia
e-mail: phyteachagu@gmail.com

https://doi.org/10.1017/S0004972721000769 Published online by Cambridge University Press