**S3 Appendix. Validity assessments of instruments for spinal muscular atrophy.**

|  |  |  |
| --- | --- | --- |
| Type of Assessment | Definition | Evaluation |
| Intra-rater reliability | The extent to which a single individual, reusing the same instrument, consistently produces the same results while examining a single set of data[1] | Tests to evaluate intra-rater reliability include: evaluation of rating consistency for a given patient set assessed multiple times over a period of time by a single practitioner  Evaluated by statistics such as: intraclass correlation coefficient, and Pearson correlation coefficient[2] |
| Inter-rater reliability | The extent to which 2 or more raters using the same instrument agree; addresses the issue of consistency of the implementation of a rating system[3] | Tests to evaluate inter-rater reliability include: evaluation of the rating consistency for a given patient (e.g. video of patient performing tasks in the instrument) by multiple practitioners  Evaluated by statistics such as: percentage agreement, kappa, and intraclass correlation coefficient, with higher values indicating higher degree of agreement[3] |
| Other validity measures | The suitability of the instrument for use in a given population | Tests to evaluate instrument validity include: evaluation of correlation with other established instruments used in the patient population, assessment of sensitivity to detect differences between patient subtypes |

REFERENCES

[1] Dictionary M. Intrarater Reliability. 2020 [cited 2020 October 16]; Available from: https://medical-dictionary.thefreedictionary.com/intrarater+reliability.

[2] Koo TK, Li MY. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J Chiropr Med 2016;15:155-63.

[3] Lange RT. Inter-rater Reliability. In: Kreutzer J S, DeLuca, J,Caplan, B ed. Encyclopedia of Clinical Neuropsychology. New York, NY: Springer New York, 2011: 1348-1348.