Inter rater reliability

2024-03-13

Table of contents

Short version 1
Litterature 2
Data 2
Inter-rater-inagreement examples 2
Calculations 3
Conclusion 4

Short version

Agreement is acceptable for all but atrophy. See calculations in the end.

My suggestion will be to re-plan. Instead of retraining, I will assign two assessors to each subject. On disagreement, consensus will be sought.

With 8 assessors helping out, each will get to perform a mean of ~250 assessments

Table 1: **Alternative**: exclude atrophy as a biomarker. The HARNESS initiative as well as the FINESSE framework recommends exactly WMH, lacunes, microbleeds and atrophy as biomarkers.(Smith et al. 2019; Markus et al. 2022)

Pros	*Cons*
Increased accuracy	Increased workload
Improved data quality	Possibly increased time use
Improved chance of using the data for further projects	Decreased chance of finishing on time

Litterature

Staals et al. (2014) reports Intraclass Correlation Coefficient of 0.68-0.92 depending on the biomarker. They use different aids to maximise the likelihood of agreement (reference pictures etc.).

Depending on sources, Fleiss-Kappa or Intraclass Correlation Coefficient (ICC) are argued as the best meassure. Hallgren (2012) recommends using the ICC with multiple assessors and when scores are ordinal. Multiple performance measures are included.

Here are a few discussions on the topic:

- researchgate
- cookbook-for-r

Data

svd_user	n
ABF	52
AGD	52
AMG	1
GA	18
JKM	52
KMØ	52
MFH	52
NLP	52
RAB	52
SBV	52

Inter-rater-inagreement examples

record_id	svd_user	svd_quality	svd_microbleed	svd_microbleed_location____1	svd_microbleed_l
svd_3	AGD	2	1	0	
svd_3	SBV	2	0	0	
svd_3	NLP	1	2	0	
svd_3	JKM	2	1	0	
svd_3	KMØ	2	3	0	
svd_3	RAB	1	2	0	
svd_3	MFH	1	1	0	

svd_3	ABF	2	2	0

record_id	svd_user	microbleed	lacunes	wmh	atrophy	score
svd_3	AGD	1	1	1	0	3
svd_3	SBV	0	1	1	0	2
svd_3	NLP	1	1	1	0	3
svd_3	JKM	1	1	1	0	3
svd_3	KMØ	1	1	1	0	3
svd_3	RAB	1	1	1	0	3
svd_3	MFH	1	1	1	0	3
svd_3	ABF	1	1	1	0	3

Calculations

Overall reliability measures on all variables

```
Warning: Using an external vector in selections was deprecated in tidyselect 1.1.0.
i Please use `all_of()` or `any_of()` instead.
  # Was:
  data %>% select(.x)

  # Now:
  data %>% select(all_of(.x))

See <https://tidyselect.r-lib.org/reference/faq-external-vector.html>.
```

```
Joining with `by = join_by(Variable)`
```

Variable	Agreement	Krippendorffs_Alpha	Fleiss_Kappa	Brennan_Predigers_
svd_quality	0.7884615	0.3909154	0.389447748	0.78
svd_microbleed	0.6923077	0.5344182	0.533296331	0.69
svd_microbleed_location____1	0.7884615	0.6058272	0.604877423	0.78
svd_microbleed_location____2	0.8269231	0.5509859	0.549903896	0.82
svd_microbleed_location____3	0.9230769	0.4543851	0.453070386	0.92
svd_siderose	0.9795918	0.0000000	-0.002557545	0.97
svd_lacunes	0.4375000	0.4987262	0.497417440	0.43
svd_wmh	0.1730769	0.5336626	0.532538933	0.17

svd_atrophy	0.1960784	0.2637533	0.261944363	0.19

Reliability on simplified 0-4 scale.

```
Joining with `by = join_by(Variable)`
```

Variable	Agreement	Krippendorffs_Alpha	Fleiss_Kappa	Brennan_Predigers_Kappa	IntraclCorrC
microbleed	0.7307692	0.6595911	0.6587709	0.7286493	0.6635
lacunes	0.6041667	0.7007082	0.6999267	0.6010499	0.7050
wmh	0.6538462	0.7193592	0.7186830	0.6511205	0.7231
atrophy	0.7058824	0.3095880	0.3078916	0.7035665	0.3192
score	0.2553191	0.5176905	0.5164044	0.2553096	0.8003

Conclusion

For the simplified score, the Intraclass Correlations Coefficients for microbleed, lacunes, wmh, atrophy and score is 0.663578226496803, 0.705020920502092, 0.723151204342727, 0.319244904777821 and 0.800345579462708 respectively.

For reference, the scoring system is included below.

Hallgren, Kevin A. 2012. "Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial." *Tutorials in Quantitative Methods for Psychology* 8 (1): 23. https://doi.org/10.20982/tqmp.08.1.p023.

Markus, Hugh S., Wiesje M. van Der Flier, Eric E. Smith, Philip Bath, Geert Jan Biessels, Emily Briceno, Amy Brodtman, et al. 2022. "Framework for Clinical Trials in Cerebral Small Vessel Disease (FINESSE)." *JAMA Neurology* 79 (11): 1187. https://doi.org/10.1001/jamaneurol.2022.2262.

Smith, Eric E., Geert Jan Biessels, François De Guio, Frank Erik De Leeuw, Simon Duchesne, Marco Düring, Richard Frayne, et al. 2019. "Harmonizing Brain Magnetic Resonance Imaging Methods for Vascular Contributions to Neurodegeneration." Edited by Jorge Jovicich and Giovanni B. Frisoni. *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring* 11 (1): 191–204. https://doi.org/10.1016/j.dadm.2019.01.002.

Staals, Julie, Stephen D. J. Makin, Fergus N. Doubal, Martin S. Dennis, and Joanna M. Wardlaw. 2014. "Stroke subtype, vascular risk factors, and total MRI brain small-vessel disease burden." *Neurology* 83 (14): 1228–34. https://doi.org/10.1212/WNL.0000000000000837.