The article is very well written and addresses the important topic of finding a short screening tool for frailty that has acceptable measurement properties. The authors have taken the Groningen Frailty Indicator which is a well-known and well used frailty indicator and modified the scoring of the tool (they have identified it consists of three scales instead of one) which is in agreement with the concept of frailty, which is seen as a multidimensional concept. In addition, in clinical practice one would have a closer look at which items the older adults scores "frail" on for identifying potential areas of intervention rather than using the overall scale score, and their three scales approach fits with this approach.

I only have a few questions.

The total sample consisted of 1508 older adults, and you used a subsample of 119 persons to examine test-retest reliability, how was the subsample of 119 persons selected? Were they a random sample of the total sample? Please clarify.

With regard to the statistical analysis, could you clarify the missing data section? In the section it is described that those with more than 5 items on the GFI missing were excluded, and it is mentioned that 17 persons are excluded for these reason. However, in the next sentence, you describe that 1277 have no missing data, 194 missed one GFI item, 27 had 2 items missing, 4 had three missing and 6 had 4 or 5 items missing. So it seems that at most 6 persons should have been excluded?

Level of interest: An article of importance in its field

Quality of written English: Acceptable

Statistical review: Yes, and I have assessed the statistics in my report.

Declaration of competing interests:

I declare that I have no competing interests