



# Herhaalde metingen van twee testen onrechtstreeks met elkaar vergelijken: mogen we hier conclusies uit trekken?

Barbara Michiels, Vakgroep Eerstelijns- en Interdisciplinaire Zorg, Centrum voor Huisartsgeneeskunde, Universiteit Antwerpen en Tom Poelman, Vakgroep Volksgezondheid en Eerstelijnszorg, UGent

In dit nummer van Minerva bespreken we een studie waarbij een herhaaldelijk uitgevoerde cognitieve zelftest (de SAGE-zelftest) vergeleken wordt met een herhaaldelijk uitgevoerde MMSE-test bij patiënten die men gedurende bijna 7 jaar opvolgde in een geheugenkliniek. De auteurs kwamen tot het besluit dat men met de SAGE-zelftest minstens 6 maanden vroeger dan met de MMSE-test een conversie van milde cognitieve stoornis (MCI) naar dementie kan detecteren (1,2). Hoe de auteurs tot dit besluit kwamen, doet echter denken aan wat we eerder bespraken in een methodologisch artikel in Minerva. We haalden toen aan dat een statistisch significant verschil in uitkomst tussen de start en de beëindiging van een behandeling in één studie-arm niet gebruikt mag worden om de superioriteit aan te tonen ten opzichte van placebo in een andere studie-arm waar men geen statistisch significant verschil zag tijdens de follow-up (3).

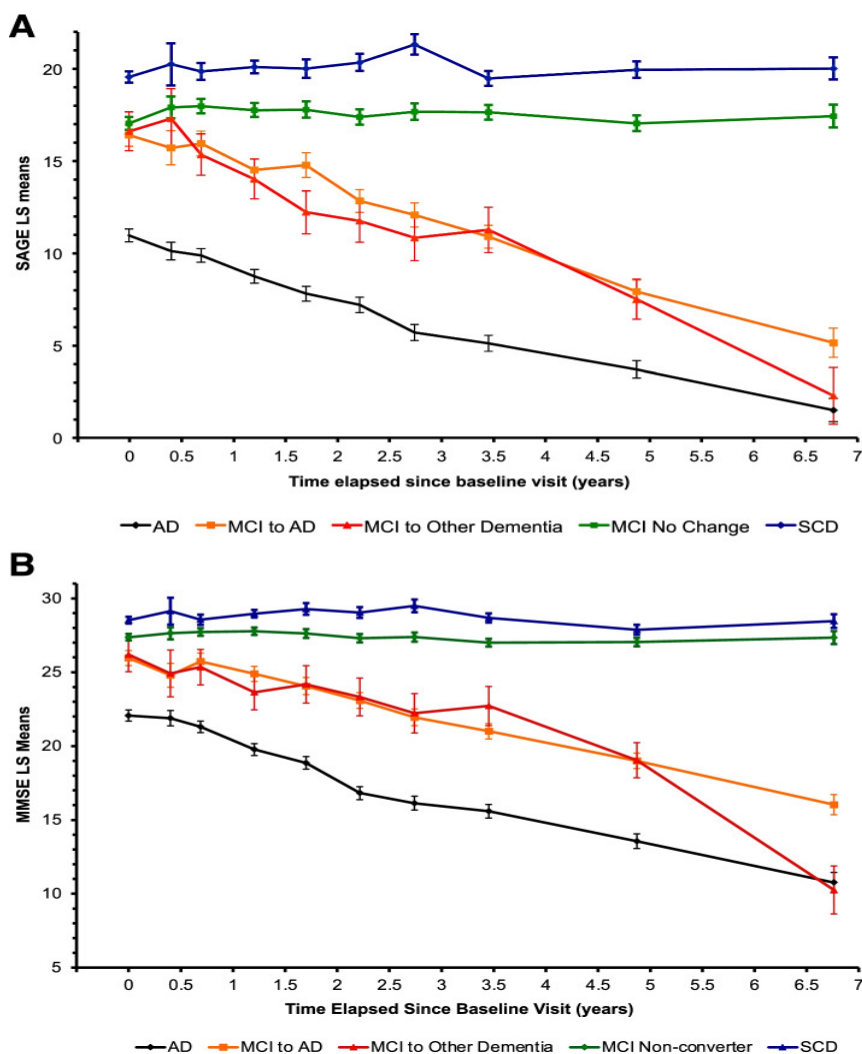
In plaats van de klassieke diagnostische criteria te gebruiken, waarbij een test beoordeeld wordt op zijn aantoonende en uitsluitende kracht op basis van sensitiviteit en specificiteit, wordt in deze studie alleen gekeken naar de snelheid waarmee de score van beide testen afzonderlijk afneemt ten opzichte van de beginwaarden. Er wordt dus geen directe statistische vergelijking gemaakt tussen de SAGE- en de MMSE-scores. Evenmin wordt er gekeken naar het aantal deelnemers bij wie de testresultaten op basis van vooropgestelde criteria veranderd zijn. Men stelt geen absoluut aantal punten voorop als afkapwaarde, zoals bijvoorbeeld 'het aantal deelnemers dat minder scoort dan x punten' of 'het aantal deelnemers met een verschil van y punten tussen twee metingen met een tussentijd van zes maanden of een jaar'. Zonder dit aan een klinisch relevante afname te koppelen en zonder directe vergelijking tussen SAGE en MMSE gebruiken de onderzoekers een regressiecurve om per test de daling in score weer te geven (zie *figuur 1*). We zien dat beide curves vrij gelijklopend dalen voor SAGE en MMSE. De onderzoekers publiceren echter ook een tabel waarin ze, zowel voor de SAGE- als de MMSE-test, de gemiddelde afname per meetmoment ten opzichte van de beginwaarden weergeven (zie *tabel 1*). Voor elke afname berekent men een p-waarde en men spreekt van een statistisch significante afname wanneer de p-waarde kleiner is dan 0,0055. Zo zien we bijvoorbeeld dat het verschil tussen SAGE en MMSE op meetpunt V3 in de groep die evolueert van MCI naar niet-alzheimer dementie ogenschijnlijk klein is, maar de p-waarde voor SAGE ( $p=0,0051$ ) lijkt net statistisch significant en voor MMSE ( $p=0,0061$ ) net niet statistisch significant te zijn. Op basis van dit minimale verschil concluderen de auteurs dat men met de SAGE-test dementie vroeger kan opsporen. Maar is dat wel betrouwbaar? Vooreerst is de statistische precisie erg afhankelijk van het aantal deelnemers. In sommige subgroepen is het aantal deelnemers laag (zoals  $n=49$  in de groep van MCI die naar AD en  $n=21$  die naar niet AD dementie evolueert). Bovendien moeten we er ook rekening mee houden dat een éénmalige significante daling veroorzaakt kan zijn door bijvoorbeeld 'ruis' bij afname van de zelftest. Ook mogen we een eventueel leereffect van herhaaldelijk testen niet uit het oog verliezen. Alhoewel de onderzoekers met dit laatste rekening hielden door de SAGE-test in 4 gevarieerde versies aan te bieden, wat het leereffect van de zelfafname had moeten afremmen.

Op basis van dit alles lijkt het ons dus onmogelijk om op basis van tabel 2 te claimen dat de SAGE-zelftest de conversie van MCI naar dementie minstens 6 maanden eerder detecteert dan MMSE. Evenmin kan men op basis van de studie een valide uitspraak doen over het verschil in aantoonende en uitsluitende kracht van SAGE ten opzichte van de MMSE (4,5). De betrouwbaarheid, validiteit en andere criteria, waaraan een goed bruikbare test moet voldoen, laten we hierbij nog buiten beschouwing (6).

## Besluit

Het indirect vergelijken van consecutieve meetresultaten van twee testen levert onbetrouwbare resultaten op die we hooguit als beschrijvend en hypothesevormend mogen beschouwen.

figuur 1



(uit: Scharre DW, Chang SI, Nagaraja HN, et al. Self-Administered Gerocognitive Examination: longitudinal cohort testing for the early detection of dementia conversion. *Alzheimers Res Ther* 2021;13:192. DOI: 10.1186/s13195-021-00930-4)

tabel 1

| Subgroup <sup>b</sup> | Test              | Visit number (average duration in months since baseline visit V0) |                             |                             |                             |                    |                             |                    |                    |                     |
|-----------------------|-------------------|---|-----------------------------|-----------------------------|-----------------------------|--------------------|-----------------------------|--------------------|--------------------|---------------------|
|                       |                   | V1 (4.8)  | V2 (8.3)                    | V3 (14.4)                   | V4 (20.4)                   | V5 (26.6)          | V6 (32.9)                   | V7 (41.4)          | V8 (58.4)          | V9 (81.1)           |
| MCI to AD dementia    | SAGE              | -0.68 (0.4593)  | -0.44 (0.5043)              | -1.88 (0.0016) <sup>c</sup> | -1.61 (0.0158)              | -3.57 <sup>d</sup> | -4.32 <sup>d</sup>          | -5.51 <sup>d</sup> | -8.47 <sup>d</sup> | -11.25 <sup>d</sup> |
|                       | MMSE              | -1.17 (0.1544)  | -0.24 (0.6862)              | -1.07 (0.0440)              | -1.89 (0.0015) <sup>c</sup> | -2.87 <sup>d</sup> | -4.01 <sup>d</sup>          | -4.96 <sup>d</sup> | -6.96 <sup>d</sup> | -9.92 <sup>d</sup>  |
| MCI to other dementia | SAGE              | 0.70 (0.6427)   | -1.26 (0.1904)              | -2.59 (0.0051) <sup>c</sup> | -4.39 <sup>d</sup>          | -4.85 <sup>d</sup> | -5.77 <sup>d</sup>          | -5.35 <sup>d</sup> | -9.10 <sup>d</sup> | -14.33 <sup>d</sup> |
|                       | MMSE <sup>e</sup> | -1.27 (0.3663)  | -0.85 (0.3749)              | -2.56 (0.0061)              | -2.02 (0.0528)              | -2.87 (0.0072)     | -3.97 (0.0005) <sup>c</sup> | -3.48 (0.0005)     | -7.15 <sup>d</sup> | -15.94 <sup>d</sup> |
| AD dementia           | SAGE              | -0.85 (0.0406)  | -1.09 (0.0001) <sup>c</sup> | -2.22 <sup>d</sup>          | -3.17 <sup>d</sup>          | -3.77 <sup>d</sup> | -5.27 <sup>d</sup>          | -5.85 <sup>d</sup> | -7.26 <sup>d</sup> | -9.46 <sup>d</sup>  |
|                       | MMSE              | -0.17 (0.7150)  | -0.77 (0.0195)              | -2.23 <sup>c,d</sup>        | -3.29 <sup>d</sup>          | -5.34 <sup>d</sup> | -5.94 <sup>d</sup>          | -6.48 <sup>d</sup> | -8.52 <sup>d</sup> | -11.34 <sup>d</sup> |

Abbreviations: AD Alzheimer's disease, MCI Mild cognitive impairment, MMSE Mini-Mental State Examination, SAGE Self-Administered Gerocognitive Examination, SCD Subjective cognitive decline

<sup>a</sup>Visit Numbers V1-V9 are labeled as V1: within 6 months; V2: 6–12 months; V3: 12–18 months, V4: 18–24 months, V5: 24–30 months, V6: 30–36 months, V7: 36–48 months, V8: 48–72 months, V9: over 72 months (max 105 months)

<sup>b</sup>None of the mean differences in the MCI (non-converter) and SCD groups was significant

<sup>c</sup>First significant difference; threshold for significance with Bonferroni correction is a *p* value below 0.05/9 = 0.00555

<sup>d</sup>Represents *p* values under 0.0001

<sup>e</sup>One outlier removed

(uit: Scharre DW, Chang SI, Nagaraja HN, et al. Self-Administered Gerocognitive Examination: longitudinal cohort testing for the early detection of dementia conversion. *Alzheimers Res Ther* 2021;13:192. DOI: 10.1186/s13195-021-00930-4)

## Referenties

1. Tournoy Jos. Kan een cognitieve zelftest de conversie van MCI naar dementie vroeger opsporen? *Minerva Duiding* 15/03/2023.
2. Scharre DW, Chang SI, Nagaraja HN, et al. Self-Administered Gerocognitive Examination: longitudinal cohort testing for the early detection of dementia conversion. *Alzheimers Res Ther* 2021;13:192. DOI: 10.1186/s13195-021-00930-4
3. Chevalier P. Hoe studie-armen correct met elkaar vergelijken? *Minerva* 2011;10(9):116.
4. Schoenmakers B. Korte cognitieve testen om klinische Alzheimer-dementie te onderscheiden van milde cognitieve achteruitgang of normaal cognitief functioneren. *Minerva Duiding* 17/12/2020.
5. Hemmy LS, Linskens EJ, Silverman PC, et al. Brief cognitive tests for distinguishing clinical Alzheimer-type dementia from mild cognitive impairment or normal cognition in older adults with suspected cognitive impairment. *Ann Intern Med* 2020;172:678-87. DOI: 10.7326/M19-3889
6. Michiels B. Waarop letten bij de interpretatie van de resultaten uit diagnostische studies? *Minerva* 2015;14(6):76.