



Herhaalde metingen, hoe analyseren?

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

In dit nummer van Minerva bespreken we een gerandomiseerde gecontroleerde studie over het effect van slaaprestrictie op slaapkwaliteit (1). De onderzoekers gebruikten onder andere twee schalen, de **PSQI-score** en de **ISI-score**, om zowel in de interventiegroep als in de controlegroep de slaapkwaliteit op een objectieve manier te meten (2).

Table 2. Pittsburgh Sleep Quality Index and Insomnia Severity Index for simplified sleep restriction versus control at baseline, 3 months, and 6 months. Data presented as mean (SD)

	Intervention (simplified sleep restriction)			Control			Difference between the groups (95% CI), <i>P</i> -value ^b
	Baseline score ^a <i>n</i> = 46	3-month score ^a <i>n</i> = 45	6-month score ^a <i>n</i> = 43	Baseline score ^a <i>n</i> = 51	3-month score ^a <i>n</i> = 49	6-month score ^a <i>n</i> = 50	
Primary outcomes							
Pittsburgh Sleep Quality Index ²⁵	10.43 (3.08)	6.24 (3.11)	6.49 (3.52)	10.29 (3.02)	8.71 (3.32)	8.10 (3.49)	-2.14 [-3.15 to -1.13] <0.0001
Insomnia Severity Index ²⁶	14.76 (3.85)	8.47 (4.45)	8.28 (4.41)	14.51 (3.36)	11.51 (3.94)	10.50 (3.72)	-2.50 [-3.97 to -1.03] 0.001

Statistically significant results are in bold. ^aObserved (raw) data. ^bMixed model for repeated measures adjusted for age, sex, and baseline insomnia severity (using ISI score).

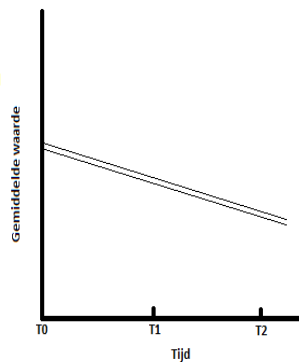
Ref.: Falloon K, Elley CR, Fernando A 3rd, et al. Simplified sleep restriction for insomnia in general practice: a randomized controlled trial. *Br J Gen Pract* 2015;65:e508-15.

De onderzoekers voerden bij elke patiënt drie metingen uit, op het moment van de randomisatie, na drie maanden en na zes maanden follow-up. Omdat de beginscores zowel voor de PSQI- als voor de ISI-score tussen beide groepen weinig verschilden, zouden we in principe het verschil in gemiddelde verschilscore (score na follow-up in de zesde maand verminderd met de beginscore) tussen beide groepen kunnen gebruiken om te bepalen of de behandeling effectief is (3). Voor de PSQI-score zouden we dan uitkomen op een daling van $(6,49 - 10,43) - (8,10 - 10,29) = -1,75$ punten en voor de ISI-score op een daling van $(8,28 - 14,76) - (10,50 - 14,51) = -2,47$ punten. In tabel 2 van het hier besproken artikel vermelden de onderzoekers echter een verschil in PSQI-score en ISI-score tussen beide groepen van respectievelijk -2,14 (met 95% BI van -3,15 tot -1,13) punten en -2,50 (met 95% BI van -3,97 tot -1,03) punten. Vooral voor de PSQI-score zien we dus een verschil van iets minder dan een half punt. Vanwaar komt dit verschil?

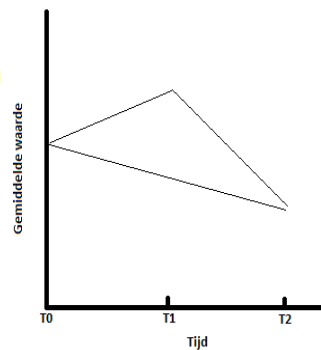
Als we een bepaalde uitkomstmaat op verschillende opeenvolgende tijdstippen meten, zullen twee metingen bij eenzelfde persoon meer op elkaar gelijken dan twee metingen bij verschillende personen. Er bestaat dus een zekere correlatie tussen herhaalde metingen bij eenzelfde persoon (4). Door deze correlatie kan het gemiddelde effect van een behandeling per meetmoment in een behandelgroep verschillen. Bovendien kan ook het relatieve effect tussen een interventie- en een controlegroep per meetmoment verschillen door een interactie tussen de effectgrootte van de behandeling (interventie versus controle) en het tijdstip waarop we meten. In de figuren hieronder worden de gemiddelde

waarden van een uitkomst per tijdstip apart weergegeven voor de interventie- en de controlegroep. Figuur A toont in beide behandelgroepen een gelijkmatige afname van de gemiddelde waarde van de uitkomst op tijdstip T1 en T2. Er bestaat dus geen interactie tussen de effectgrootte en het tijdstip waarop we meten. Omdat de lijnen bijna op elkaar liggen, is er geen verschil in effect tussen beide behandelgroepen. In figuur B bestaat er geen verschil in de gemiddelde waarde van de uitkomst tussen de twee behandelgroepen op tijdstip 0 en 2, maar wel op tijdstip 1. Stel dat we nu enkel rekening zouden houden met tijdstip 0 en 2, dan zouden we besluiten dat er geen verschil in effect bestaat tussen de interventie- en de controlegroep. We zien echter een groot verschil in effectgrootte tussen enerzijds tijdstip 0 en 2 en anderzijds tijdstip 1. Hier is dus wel sprake van een interactie tussen effectgrootte en tijdstip waarop we meten. Door hiermee rekening te houden, zal er waarschijnlijk wel een statistisch significant verschil in effect tussen de twee behandelgroepen gevonden worden.

Figuur A



Figuur B



Om te corrigeren voor de correlatie van opeenvolgende metingen voor elke deelnemer binnen een behandelgroep en om te weten of er een interactie bestaat tussen effectgrootte en tijdstip waarop we meten, kunnen we een repeated measures ANOVA gebruiken. ANOVA (analysis of variance) wordt toegepast om het gemiddelde van een continue uitkomstmaat van drie of meer onafhankelijke (behandel-)groepen met elkaar te vergelijken (5). Een repeated measures ANOVA vergelijkt het gemiddelde van een continue uitkomstmaat (interventie versus controle) op verschillende tijdstippen en gaat na of er een interactie tussen effectgrootte en tijdstip van meting bestaat. Deze methode stelt echter zeer strenge voorwaarden waaraan het werkelijke onderzoek niet altijd kan voldoen. De metingen moeten voor alle deelnemers op hetzelfde moment gebeuren en ontbrekende gegevens worden niet toegelaten. Bovendien is het met deze methode niet mogelijk om de effectgrootte zelf te berekenen. Daarom wordt vaak gebruik gemaakt van een lineair mixed model. Eigenlijk is dit een complexe regressieanalyse die toelaat om het behandel-effect te berekenen, rekening houdend met de variatie van metingen over de tijd bij één deelnemer en de mogelijke interactie tussen deze variatie en de effectgrootte. Het voordeel is dat deze methode, net als ANCOVA (3), ook rekening kan houden met andere confounders. Het enige nadeel is de complexiteit en de vereiste rekenkracht die om geoptimaliseerde software vraagt (6).

In ons voorbeeld werd gekozen voor een mixed model om toe te laten de effectgrootte van slaaprestrictie versus controle te berekenen, rekening houdend met de correlatie van de metingen bij één deelnemer, de interactie van een verschil in effectgrootte tussen verschillende meetmomenten, alsook met confounders zoals leeftijd, geslacht en graad van slapeloosheid bij de start. Het feit dat er alleen voor de PSQI-score een verschil bestond in effectgrootte bij gebruik van een mixed model doet vermoeden dat er een verschillend behandel-effect per meetmoment optrad alleen voor deze PSQI-score maar niet voor de ISI-score.

Besluit

Wanneer de uitkomstmaat voor alle deelnemers van een studie op verschillende tijdstippen gemeten wordt, moeten we een analysemethode gebruiken die rekening houdt met het feit dat metingen niet onafhankelijk van elkaar gebeuren en dat naargelang het meetmoment het effect van de behandeling kan verschillen. Om hiervoor te corrigeren is een mixed model complexer maar algemener te gebruiken dan een repeated measures ANOVA.

Referenties

1. Declercq T. Vereenvoudigde vorm van slaaprestrictie als behandeling van slapeloosheid in de eerste lijn. *Minerva* 2016;15(6):136-9.
2. Falloon K, Elley CR, Fernando A 3rd, et al. Simplified sleep restriction for insomnia in general practice: a randomized controlled trial. *Br J Gen Pract* 2015;65:e508-15.
3. Poelman T. Continue variabelen analyseren met ANCOVA. *Minerva* 2014;13(8):103.
4. Bland JM, Altman DG. Correlation, regression, and repeated data. *BMJ* 1994;308:896.
5. Sedgwick P. One way analysis of variance. *BMJ* 2012;344:e2427.
6. Petrie A, Sabin C. *Medical statistics at a glance*. John Wiley & Sons 2009 (3rd ed.).