



Onderzoeksrapport

EEN EFFECTONDERZOEK NAAR DE ANDER LEVEN INTERVENTIE OP BASIS VAN DE SF-36 VRAGENLIJST

Tim Choy

18-10-2018

DE INHOUD VAN DIT RAPPORT IS VERTROUWELIJK EN ALLEEN BEDOELD VOOR DEGENE AAN WIE DE INFORMATIE GERICHT IS.
HET KAN NIET ZONDER TOESTEMMING VAN EVIDENTO MET DERDEN WORDEN GEDEELD.

Inhoudsopgave

1. Inleiding	2
2. Onderzoeksozet.....	3
3. Steekproef en generaliseerbaarheid	7
4. Kenmerken onderzoeksgroep	9
5. Uitkomsten op de voor- en nametingen	11
6. Effectiviteit van de Ander Leven interventie.....	15
7. Percentage deelnemers dat duurzaam herstelt.....	18
8. Conclusies.....	19
Literatuur.....	21
Bijlage 1: Bedrijfsprofiel	22
Bijlage 2: Tijdsmetingen	23
Bijlage 3: SF/RAND-36 vragenlijst (NL).....	24
Bijlage 4: Rekensystematiek RAND-36 dimensies.....	25
Bijlage 5: Argumentatie voor (non-)parametrische testen.....	28
Bijlage 6: Effect size formules	31
Bijlage 7: Steekproefomvang i.r.t. effectmeting	32
Bijlage 8: Descriptieve statistiek	33
Bijlage 9: Grafieken en tabellen onderzoeksresultaten.....	36
Bijlage 10: Repeated Measures test nader bekeken	53

1. Inleiding

Ander Leven past een interventie toe op mensen met (chronische) lichamelijke en geestelijke klachten. De interventie bestaat uit een driedaagse training en een aantal onderdelen die daar omheen zijn gebouwd, te weten een aanmeldingsprocedure en een persoonlijk consult voor de training en een verdiepingsdag en twee telefonische follow-ups na de training. Daarnaast is er een professioneel ondersteuningsproces bestaande uit intensieve nazorg (bijvoorbeeld tot aan de verdieping altijd kunnen bellen) en reguliere nazorg (bijvoorbeeld een jaar na de training contact kunnen opnemen), een zelfregulerend ondersteuningsproces (o.a. buddy-systeem, Ander Leven CD en dagboek bijhouden) en een zelfregulerend helingsproces: het zelfstandig lopen van AC-processen (dit is een specifieke imaginatieoefening).

Sinds 2011 verzamelt Ander Leven data over de gezondheidstoestand c.q. het welzijn van cliënten voor aanvang en meerdere keren na toepassing van de interventie aan de hand van de SF/RAND-36 vragenlijst. Ander Leven wil namelijk weten in hoeverre de interventie effectief is: *verbetert de lichamelijke en geestelijke gezondheid na de interventie en is het resultaat blijvend?* Door antwoord op deze vraag te krijgen kan Ander Leven uitspraken doen over de effectiviteit van haar interventie en belanghebbenden daarover informeren.

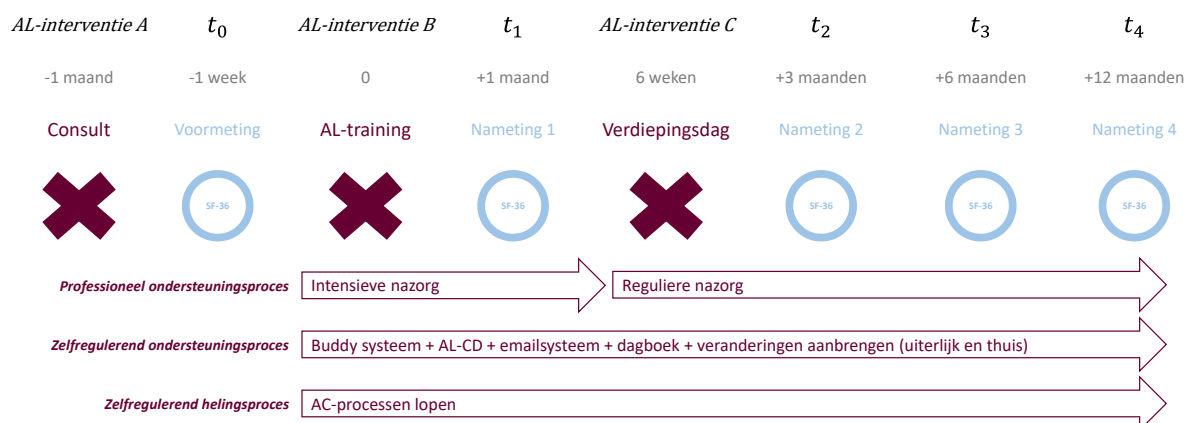
Ander Leven heeft in het voorjaar van 2017 een onafhankelijk wetenschappelijk onderzoeksbureau (Evidento) ingeschakeld om effectonderzoek naar de interventie te doen (zie bijlage 1 voor een bedrijfsprofiel). De uitkomsten van het onderzoek geven meer uitsluitsel over de effectiviteit van de Ander Leven interventie en in hoeverre deze als 'evidence based' kan worden aangemerkt. Deze zijn ook overlegd aan de Universiteit van Gent, met als uitkomst om op basis van dit rapport een wetenschappelijk artikel te publiceren.

2. Onderzoeksopzet

De onderzoeksopzet is gebaseerd op hoe Ander Leven mensen selecteert om deel te nemen aan de interventie, bij wie gegevens zijn verzameld, op welke tijdstippen gegevens zijn verzameld en welke gegevens zijn verzameld. Dit gaf een aantal mogelijkheden om adequaat effectonderzoek te kunnen doen, maar ook een aantal beperkingen.

Deelnemers met lichamelijke en/of geestelijke klachten, waarop de interventie is toegepast, zijn geselecteerd op basis van een aanmeldingsprocedure, bestaande uit het invullen van een aanmeldingsformulier met specifieke vragen en een telefonisch intake gesprek. Dit betekent voor het onderzoek dat deelnemers niet 'aselect' geselecteerd zijn, oftewel, niet elke aanmelding c.q. persoon met lichamelijke en/of geestelijke klachten heeft evenveel kans heeft om de interventie te ondergaan. Deelnemers in dit onderzoek voldoen aan bepaalde (motivatie gedreven) voorwaarden en deze kunnen van invloed zijn op de uitkomsten.

De onderzoekspopulatie bestaat uit alle deelnemers die vanaf 2011 tot en met 2017 de Ander Leven interventie hebben ondergaan. Op basis van deze populatie is de effectiviteit van de interventie onderzocht. Dit is gebeurd door de lichamelijke en geestelijke gezondheidstoestand van de deelnemers *voor* en *na* de training te meten en deze met elkaar te vergelijken. Dat heeft geresulteerd in een voormeting (1 week voor aanvang van de training) en vier nametingen (1, 3, 6 en 12 maanden na de training). Onderstaande figuur geeft een visuele representatie van de meetpunten in relatie tot de toepassing van de Ander Leven interventie (zie ook bijlage 2 voor uitleg en verantwoording over de tijdsmetingen).



De lichamelijke en geestelijke gezondheidstoestand is gemeten aan de hand van de SF/RAND-36 vragenlijst (zie bijlage 3). Dit is een betrouwbare en wetenschappelijk gevalideerde vragenlijst die de volgende variabelen meet:

- Algemene gezondheid
- Lichamelijke gezondheid
- Rolbeperkingen n.a.v. lichamelijke problemen
- Lichaamspijn
- Geestelijke gezondheid
- Rolbeperkingen n.a.v. geestelijke (emotionele) problemen
- Vitaliteit/energie
- Sociaal functioneren

De variabelen worden via een rekensystematiek omgezet naar een meetschaal van 1-100 (zie bijlage 4), waarbij geldt: hoe hoger de score, hoe beter de gezondheidstoestand (Ware et al., 1993). Op basis van een bepaalde rekensystematiek zijn ook twee geaggregeerde variabelen gemeten: lichamelijk welzijn ('Physical Component Score') en geestelijk welzijn ('Mental Component Score'). De eerste is een soort van gemiddelde van de eerste vier variabelen en de tweede een soort van gemiddelde van de laatste vier variabelen. De meetschaal van de welzijnsvariabelen kent een standaardisatiewaarde van 50, welke als normering geldt voor het landelijk gemiddelde van een westerse bevolking (Ware et al., 1993). Een score onder de 50 is lager dan het landelijk gemiddelde en een score boven de 50 is hoger dan het landelijk gemiddelde. Van belang is in hoeverre deelnemers voor en na de training van deze standaardisatiewaarde afwijken (om meer te kunnen zeggen of deelnemers herstellen naar een 'normaal' gezondheidsniveau).

De SF/RAND-36 vragenlijst is gedigitaliseerd in een enquête-softwareprogramma (SurveyMonkey), zodat cliënten de vragenlijst online konden invullen. Deelnemers kregen op gezette tijden een e-mail met de vraag om de enquête in te vullen. De antwoorden van elke respondent zijn opgeslagen in het SurveyMonkey account van Ander Leven. Alle gegevens over de periode 2011-2017 zijn uiteindelijk geëxporteerd naar Excel en daar verder bewerkt en opgeschoond (hierbij zijn ook bepaalde gegevens van respondenten uit het administratiesysteem van Ander Leven toegevoegd, zoals geslacht en leeftijd). De bewerkte Excel data is uiteindelijk naar SPSS geëxporteerd; een statistisch software-programma om de data te analyseren.

Deelnemers die de vragenlijst op alle vijf de meetmomenten volledig hebben ingevuld vormen de onderzoeksgroep c.q. steekproef, op basis waarvan de effectiviteit van de interventie is vastgesteld. In elke meting zitten dus dezelfde deelnemers. Dit maakt het onderzoek krachtiger, omdat een eventueel geconstateerd verschil tussen de voor- en nametingen minder snel toegerekend kan worden aan mogelijke verschillen in de karakteristieken van de deelnemers (denk bijvoorbeeld aan veel meer oudere deelnemers in de voormeting dan in de nameting, waardoor 'ouderdom' het verschil zou kunnen verklaren).

Deelnemers in de steekproef zijn echter niet gerandomiseerd (op toeval) aan de onderzoeksgroep toegewezen, waardoor de kans bestaat dat een bepaald type deelnemer in de steekproef oververtegenwoordigd is (bijvoorbeeld meer tevreden deelnemers of deelnemers die veel tijd hebben om de vragenlijst op alle tijdstippen in te vullen). Dit zou de onderzoeksresultaten mogelijk kunnen verkleuren (mensen met veel tijd hebben wellicht minder stress dan mensen met weinig tijd). Ook was het niet mogelijk om de onderzoeksgroep te vergelijken met een controlegroep (mensen met overeenkomstige lichamelijke en/of geestelijke klachten die de Ander Leven interventie niet volgen en die gelijktijdig met de onderzoeksgroep de vragenlijst invullen). Indien een effect van de interventie zichtbaar zou worden bij de onderzoeksgroep en niet bij een controlegroep, zou er ook met meer zekerheid vastgesteld kunnen worden dat de interventie effect heeft.

In dit onderzoek was het wel mogelijk om het verschil tussen de voor- en nametingen te controleren op bepaalde karakteristieken van de deelnemers, o.a. geslacht, leeftijd, nationaliteit, inkomenssituatie en type klacht. Met andere woorden, als de interventie voor iedereen effectief is, dan moet het verschil tussen voor- en nametingen hetzelfde zijn voor bijvoorbeeld mannen en vrouwen, ouderen en jongeren, Nederlanders en Belgen, armen en rijken en mensen met lichamelijke en geestelijke klachten.

Voorts zijn deelnemers alleen bevraagd op hun gezondheidstoestand en niet op de mate waarin zij gebruik maken van het professionele en zelfregulerend ondersteuningsproces, alsmede in hoeverre zij het zelfregulerend helingsproces (thuis) uitvoeren. Daardoor kon ook niet onderzocht worden of de

mate waarin deze processen zijn toegepast, effect hebben gehad op de gezondheidstoestand. Hier kan alleen verondersteld worden dat deze processen zijn toegepast. Daarom doet dit onderzoek voornamelijk een uitspraak over het effect van de driedaagse training (verschil tussen de voor- en eerste nameting) en hoe lang dit effect stand houdt in de tijd (na 1, 3, 6 en 12 maanden). Daarnaast kan wat gezegd worden over de effectiviteit van de verdiepingsdag (deze ligt tussen de eerste en tweede nameting). Het onderzoek kan geen uitspraak doen over het effect van het consult, omdat de voormeting daarna heeft plaatsgevonden.

Gegeven de (on)mogelijkheden in het onderzoek kan het design in vaktermen worden aangeduid als een *non-gerandomiseerde longitudinale cohort studie* (Katz, 2010). Non-gerandomiseerd wil zeggen dat mensen met lichamelijke en/of geestelijke klachten niet op basis van toeval aan de onderzoeksgroep zijn toegewezen. Een longitudinale cohort studie betekent dat een groep mensen met een bepaald gemeenschappelijk kenmerk (in dit geval lichamelijke en/of geestelijke klachten) in de tijd wordt gemeten.

Om binnen de beperkingen van het onderzoek de effectiviteit van de interventie vast te stellen, moet er een statistisch significant verschil zijn tussen de voormeting en de vier nametingen: de lichamelijke en geestelijke gezondheidstoestand van de deelnemers in de onderzoeksgroep moet na de interventie dusdanig verbeterd zijn dat deze verbetering – statistisch gezien – niet op toeval berust. Daarbij komt dat de verbeterde gezondheidstoestand op de vier nametingen min of meer gelijk moet blijven om aan te kunnen geven dat het effect van de interventie, over een periode van 12 maanden, duurzaam is.

Hierover is meer uitsluitsel verkregen door bepaalde statistische testen uit te voeren die op basis van specifieke procedures gekozen zijn. Daarbij zijn de uitgangspunten van Katz (2010) en Field (2018) gevolgd. Op basis van de karakteristieken van de variabelen en bepaalde statistische voorwaarden, komen twee statistische tests naar voren om de effectiviteit te bepalen: de Friedman test en de 'Repeated Measures' (RM) test. De eerste is een zogenoemde non-parametrische test en de tweede een parametrische test.

Een parametrische test is sterker dan een non-parametrische test, maar deze moet voldoen aan bepaalde statistische voorwaarden. Dit onderzoek voldoet grotendeels aan deze voorwaarden (zie bijlage 5) en daarom is de RM test uitgevoerd. Niettemin blijft er ruimte voor discussie en daarom is in dit onderzoek ervoor gekozen om naast de parametrische test ook de non-parametrische test uit te voeren. Dit biedt niet alleen meer zekerheid, maar geeft ook de mogelijkheid om de tests met elkaar te vergelijken en meer over de effectiviteit van de interventie te zeggen. Indien de twee testen allebei hetzelfde uitwijzen, kan met meer stelligheid een uitspraak over de effectiviteit van de interventie worden gedaan.

De testen geven uitsluitsel of er een significant verschil is tussen de voormeting ten opzichte van de nametingen. Binnen de RM test kan ook getoetst worden of de onderlinge nametingen significant van elkaar verschillen. Bij de Friedman test is dat niet mogelijk. Daarom wordt ook de non-parametrische Wilcoxon test toegepast om deze verschillen met elkaar te vergelijken (zie Field, 2018).

Een significant verschil wil zeggen dat het verschil – statistisch gezien – niet meer op toeval berust en dus dat het verschil aan 'iets' kan worden toegewezen. Zoals eerder aangegeven is de veronderstelling dat de driedaagse training een verklaring geeft voor een significant verschil tussen de voormeting en de eerste nameting. Mogelijk kan er nog een klein verschil optreden tussen de eerste nameting en de twee nameting door de verdiepingsdag, maar over het algemeen wordt verwacht dat de nametingen op hetzelfde niveau zullen zitten. Als de voormeting significant verschilt van de nametingen en de nametingen nauwelijks van elkaar verschillen, geeft dit ook aan dat het effect van de interventie c.q. de driedaagse training stand houdt in de tijd (na 1, 3, 6 en 12 maanden).

Een significant verschil wordt vastgesteld als de nulhypothese wordt verworpen. De nulhypothese geeft de verwachting aan dat de scores tussen de voor- en nametingen gelijk zijn (of weinig van elkaar verschillen). Een aanname of verwerping van de nulhypothese wordt bepaald door het daadwerkelijk gemeten verschil tussen de voor- en nametingen af te zetten tegen de verwachting dat er geen/weinig verschil is. Indien het gemeten resultaat cijfermatig gezien veel afwijkt van de verwachting dat er geen/weinig verschil is, kan de nulhypothese verworpen worden. Dit gebeurt aan de hand van bepaalde rekenformules die binnen de Friedman test en RM test van toepassing zijn. Hierbij speelt kansverdeling een prominente rol: de kans om, als de nulhypothese waar is (in werkelijkheid is er geen verschil), per toeval een verschillscore te meten die duidelijk kleiner of groter is. Dan wordt de nulhypothese ten onrechte verworpen en wordt er een foutieve uitspraak gedaan.

De kans op dit toeval kan via een statistische procedure verkleind worden door de zogenoemde toevalwaarde (ook wel p-waarde genoemd) binnen de testen zo laag mogelijk te stellen. De laagste toevalwaarde binnen de Friedman test is 1%, dat wil zeggen dat in 1 van de 100 gevallen de kans bestaat dat een verschilwaarde wordt gemeten die toevallig veel afwijkt. De laagste toevalwaarde binnen de RM test is 0,1%, hetgeen wil zeggen dat in 1 van de 1000 gevallen de kans bestaat dat een verschil wordt gevonden die toevallig veel afwijkt. Deze toevalwaarden zijn binnen dit onderzoek gebruikt, waardoor de strengste statistische procedures van toepassing zijn. Dus indien een verschilwaarde tussen de voor- en nametingen wordt gevonden die significant afwijkt van de verwachte waarde (nulhypothese: dat er geen/weinig verschil is), kan met zeer grote zekerheid de nulhypothese worden verworpen en het verschil aan de Ander Leven interventie worden toegewezen.

Hoe groot of hoe klein het cijfermatige verschil tussen het daadwerkelijk gemeten verschil en het verwachte verschil moet zijn om een uitspraak te doen (de nulhypothese aan te nemen of te verwerpen), hangt mede af van de statistische formules die binnen de specifieke test gehanteerd worden en de steekproefomvang. De formules houden bijvoorbeeld rekening met een kleine of een grote steekproefomvang (het maakt verschil uit of er 10 of 100 respondenten in de metingen zitten). In dit geval moet de gemeten toevalwaarde bij de verschillscore tussen de voor- en nametingen in de Friedman test in ieder geval lager dan 1% zijn en in de RM test lager dan 0,1%. Daarnaast geven de ontwikkelaars van de SF/RAND-36 vragenlijst (Ware et al., 1993) aan dat een verschil van 5 punten, tussen twee metingen, als een klinisch en sociaal relevant verschil kan worden aangemerkt (zie ook Walters, 2009).

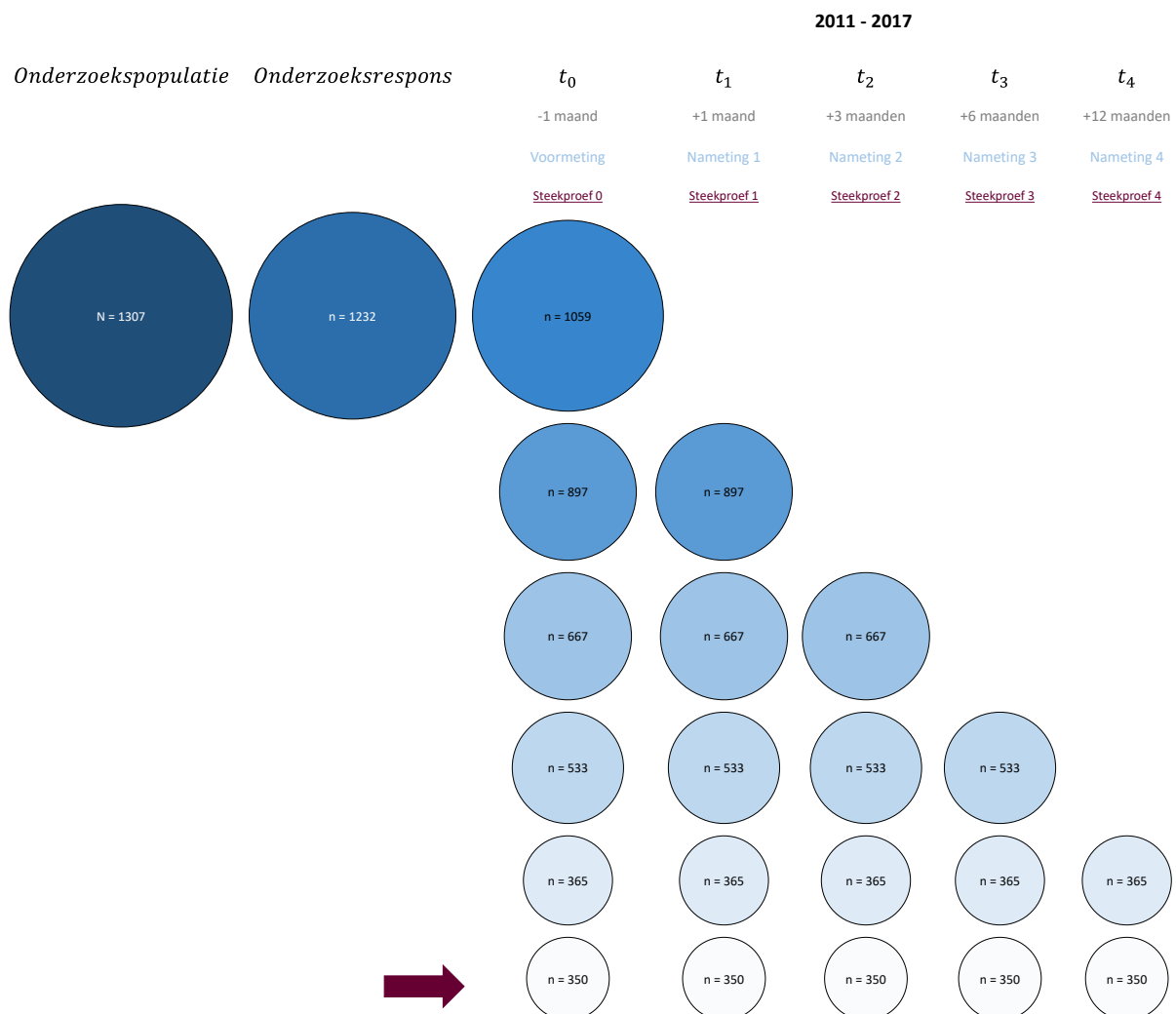
Verder is gekeken naar de zogenaamde 'effect size', oftewel de sterkte van het effect als er een significant verschil tussen de voor- en nametingen wordt vastgesteld. Hierbij hanteren de Friedman en RM test andere formules en normeringen om de effect size te bepalen (zie bijlage 6).

Tenslotte is via de RM test gecontroleerd of verschillen tussen de voor- en nametingen gerelateerd zijn aan bepaalde factoren, zoals geslacht, leeftijd, nationaliteit, type begeleider en type klacht (de toevalswaarde is hier op 5% gezet). Daarmee kan worden aangegeven in hoeverre deze factoren van invloed zijn op het geconstateerde verschil tussen de voor- en nametingen.

3. Steekproef en generaliseerbaarheid

Om uitspraken te kunnen doen op welke populatie van mensen de Ander Leven interventie effect heeft, is het van belang om meer inzicht te krijgen in de doelpopulatie, onderzoekspopulatie en kenmerken van deelnemers in de steekproef. De doelpopulatie van Ander Leven zijn mensen in Nederland en Vlaanderen met (chronische) lichamelijke en/of geestelijke klachten die voldoen aan de selectievoorwaarden van Ander Leven. De onderzoekspopulatie bestaat uit geselecteerde deelnemers uit de doelpopulatie waarop de Ander Leven interventie, in de periode januari 2011 t/m november 2017, is toegepast (in totaal 1307 mensen). In de steekproef zitten mensen uit de onderzoekspopulatie die vijf keer (1 voormeting en 4 nametingen) de SF/RAND-36 vragenlijst volledig hebben ingevuld. In beginsel waren dit 365 mensen, maar na het opschonen van de data en het verwijderen van uitzonderlijke gevallen ('outliers') bestaat de steekproefomvang uit 350 mensen (zie ook bijlage 5).

In onderstaande figuur staat de onderzoekspopulatie en de steekproefomvang bij elke meting visueel weergegeven (de steekproef wordt logischerwijs steeds kleiner, omdat niet iedereen de vragenlijst bij elke meting invult).



Een steekproefomvang van 350 mensen is volgens de uitgangspunten van de ontwikkelaars van de SF/RAND-36 vragenlijst (Ware et al., 1993) groot genoeg om bij een verschil van 5 punten, tussen twee metingen, een significant verschil aan te tonen (zie ook bijlage 7).

Op basis van de steekproef kan een uitspraak worden gedaan over de totale onderzoekspopulatie. Er zijn geen gegevens bekend over de omvang van de doelpopulatie, waardoor er geen uitspraken gedaan kunnen worden over de generaliseerbaarheid van de steekproeven naar de doelpopulatie.

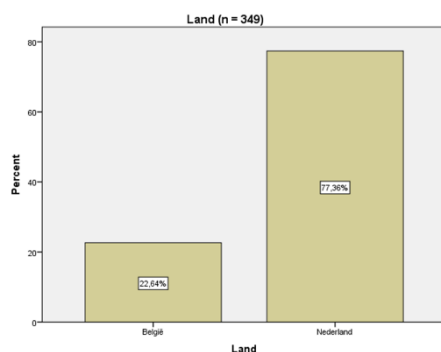
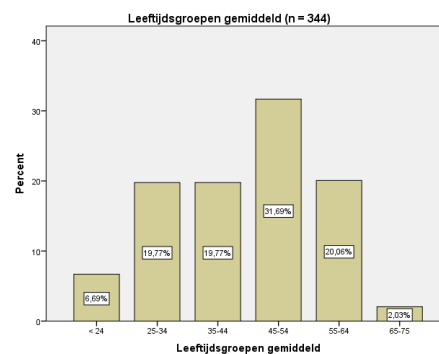
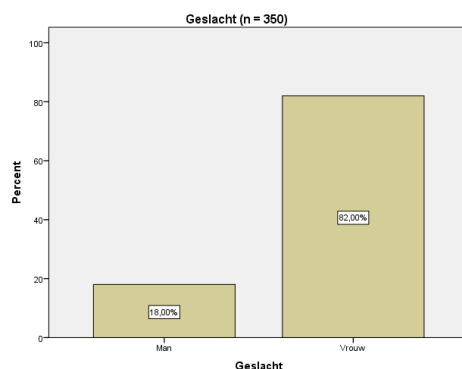
Ruim 23% van de onderzoekspopulatie heeft de vragenlijst op elke meting volledig ingevuld. Maar statistisch zegt dit nog weinig over de representativiteit van de steekproef. De generaliseerbaarheid van de steekproef hangt af van de omvang in relatie tot de omvang van de onderzoekspopulatie en wordt bepaald aan de hand van een *betrouwbaarheidsmarge* en een *foutmarge*.

In de regel zijn de onderzoeksresultaten uit de steekproef te generaliseren naar de onderzoekspopulatie bij een betrouwbaarheidsmarge van 95% en een foutmarge van 5%. Dat wil zeggen dat als uit de onderzoekspopulatie 100 steekproeven worden getrokken met dezelfde steekproefomvang er in 95 gevallen overeenkomstige gemiddelde scores uitkomen (die niet ver van elkaar afwijken) en in 5 gevallen de gemiddelde scores – gegeven het toeval – veel af zullen wijken. Om deze percentages aan te kunnen houden moet de steekproefomvang, ten opzichte van de omvang van de onderzoekspopulatie, een bepaalde grootte hebben. Dit wordt vastgesteld aan de hand van een statistische procedure. Deze wijst uit dat een steekproefomvang van 350 deelnemers ten opzichte van een onderzoekspopulatie van 1307 deelnemers, een foutmarge van 4,48% oplevert als het betrouwbaarheidsmarge op 95% wordt gezet. De foutmarge blijft dus onder de 5%, waardoor de steekproef in dit onderzoek representatief zijn voor de onderzoekspopulatie en de onderzoeksresultaten gelden voor alle deelnemers de Ander Leven training gevolgd hebben.

4. Kenmerken onderzoeksgroep

Om te weten waarop de onderzoeksresultaten betrekking hebben, is het belangrijk om meer inzicht te krijgen in de kenmerken van de onderzoeksgroep (zie ook bijlage 8 voor een totaaloverzicht). De belangrijkste staan hieronder weergegeven:

- De meeste deelnemers zijn vrouw (82%).
- Het merendeel van de deelnemers bevindt zich in de leeftijd van 25 tot 65 jaar, waarbij mensen in de leeftijd van 45 tot 55 het sterkst vertegenwoordigd zijn.
- De meeste deelnemers komen uit Nederland (77%), gevolgd door Vlaanderen (23%).
- De begeleiders van de deelnemers tijdens de training zijn voornamelijk de partners (83%) ten opzichte van niet-partners (17%), zoals een familielid, vriend of kennis.



De type lichamelijke en/of geestelijke klachten van de deelnemers verschillen sterk van elkaar. Om een beter overzicht te maken zijn de specifieke klachten geaggregeerd naar een abstracter niveau (zie onderstaande tabellen). Geaggregeerde lichamelijke klachten die bij meer dan 30 deelnemers geconstateerd zijn, hebben betrekking op:

- het spijsverteringsstelsel (n = 36)
- het zenuwstelsel (n = 41)
- het circulatiestelsel (n = 30)
- onverklaarbare lichamelijke klachten (SOLK) (n = 101)
- een combinatie van diverse klachten (n = 125)

		Statistics																	
	Ademhalingsstelsel	Bloedvatenstelsel	Botstelsel	Celstelsel	Gewrichtenstelsel	Hormoonstelsel	Huidstelsel	Lymfevaten/Immuunstelsel	Sensorischstelsel	Spiersstelsel	Spijsverteringsstelsel	Urinewegstelsel	Voortplantingsstelsel	Zenuwstelsel	Motorischstelsel	Circulatiestelsel	SOLK*	Diversen	
N Valid	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	336	
Missing	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	14	
Aanwezig	3	12	7	7	3	6	8	7	4	9	36	0	1	41	19	30	101	125	
Afwezig	333	324	329	329	333	330	328	329	332	327	300	336	335	295	317	306	235	211	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	15	
15		15	15	15	15	15	15	15	15	15	15</								

* Spasmofilie, CVS, FM, ME, etc.

Geaggregeerde geestelijke klachten die bij meer dan 30 deelnemers geconstateerd zijn, hebben betrekking op:

- Angst- en paniekstoornissen (n = 50)
- Persoonlijkheidsstoornissen (n = 46)
- Slaapstoornissen (n = 37)
- Stemningsstoornissen (n = 36)
- Stress (n = 108)
- Vitaliteit (n = 74)

Statistics													
	Angst- en paniekstoornis	Emotioneel probleem	Gedragsprobleem	Obsessief-compulsieve stoornis	Ontwikkelingsstoornis	Persoonlijkheidsstoornis	Slaapstoornis	Somatische stoornis	Stemmingsstoornis	Stress	Trauma stoornis	Verslavingsstoornis	Vitaliteit
N Valid	336	336	336	336	336	336	336	336	336	336	336	336	336
Missing	14	14	14	14	14	14	14	14	14	14	14	14	14
Aanwezig	50	2	0	1	0	46	37	0	36	108	7	1	74
Afwezig	286	334	336	335	336	290	299	336	300	228	329	335	262

Van de mensen in de steekproef hebben in totaal 130 deelnemers lichamelijke klachten, 190 deelnemers geestelijke klachten en 125 deelnemers zowel lichamelijke als geestelijke klachten.

5. Uitkomsten op de voor- en nametingen

Voormeting

De gemiddelde scores op de SF/RAND-36 variabelen en de geaggregeerde welzijnsvariabelen (schuingedrukt) tijdens de voormeting, op een schaal van 1-100, zijn als volgt:

- Algemene gezondheid ($\bar{x} = 44$)
- Lichamelijke gezondheid ($\bar{x} = 70$)
- Rolbeperkingen n.a.v. lichamelijke problemen ($\bar{x} = 26$)
- Lichaamspijn ($\bar{x} = 62$)
- Geestelijke gezondheid ($\bar{x} = 53$)
- Rolbeperkingen n.a.v. geestelijke (emotionele) problemen ($\bar{x} = 46$)
- Vitaliteit/energie ($\bar{x} = 35$)
- Sociaal functioneren ($\bar{x} = 44$)
- *Lichamelijk welzijn* ($\bar{x} = 43$)
- *Geestelijk welzijn* ($\bar{x} = 33$)

Opvallend genoeg scoort de lichamelijke gezondheid hoog (gemiddeld 70 punten) en rolbeperkingen naar aanleiding van lichamelijke problemen erg laag (gemiddeld 26 punten). Dit verschil is waarschijnlijk te verklaren door de type lichamelijke klachten van de deelnemers en het soort van vragen dat gesteld is om de lichamelijke gezondheid te meten. Veel deelnemers hebben chronische lichamelijke klachten (bijvoorbeeld een prikkelbare darm of continue vermoeidheid), maar kunnen over het algemeen nog goed traplopen, afstanden lopen en zware objecten tillen. De SF/RAND-36 vragenlijst gaat in op dergelijke aspecten om de lichamelijke gezondheid vast te stellen en niet op chronische lichamelijke en/of geestelijke klachten (zie ook bijlage 3).

Nametingen

De gemiddelde scores op de SF/RAND-36 variabelen en de geaggregeerde welzijnsvariabelen (schuingedrukt) op de nametingen tezamen (gemiddelde score op 1, 3, 6 en 12 maanden), zijn als volgt:

- Algemene gezondheid ($\bar{x} = 76$)
- Lichamelijke gezondheid ($\bar{x} = 85$)
- Rolbeperkingen n.a.v. lichamelijke problemen ($\bar{x} = 74$)
- Lichaamspijn ($\bar{x} = 80$)
- Geestelijke gezondheid ($\bar{x} = 77$)
- Rolbeperkingen n.a.v. geestelijke (emotionele) problemen ($\bar{x} = 84$)
- Vitaliteit/energie ($\bar{x} = 70$)
- Sociaal functioneren ($\bar{x} = 82$)
- *Lichamelijk welzijn* ($\bar{x} = 52$)
- *Geestelijk welzijn* ($\bar{x} = 49$)

Uit de onderzoeksresultaten komt naar voren dat de gezondheidstoestand van de deelnemers na de Ander Leven interventie sterk verbetert (zie onderstaande tabel voor een overzicht).

SF/RAND-36	Voormeting (gemid.)	Nametingen (gemid.)	Vershil voor- en nametingen
Algemene gezondheid	44	76	32
Lichamelijke gezondheid	70	85	15
Rolbeperkingen n.a.v. lichamelijke problemen	26	74	48
Lichaamspijn	62	80	18
Geestelijke gezondheid	53	77	24
Rolbeperkingen n.a.v. geestelijke (emotionele) problemen	46	84	38
Vitaliteit/energie	35	70	35
Sociaal functioneren	44	82	38
Lichamelijk welzijn	43	52	9
Geestelijk welzijn	33	49	16

De grootste verbetering is waarneembaar op Rolbeperkingen n.a.v. lichamelijke problemen met maar liefst 48 punten. De gezondheidstoestand op Algemene gezondheid (stijging van 32 punten), Rolbeperkingen n.a.v. geestelijke (emotionele) problemen (stijging van 38 punten), Vitaliteit/energie (stijging van 35 punten) en Sociaal functioneren (stijging van 38 punten) verbetert ook zeer sterk. De Lichamelijke gezondheid (stijging van 15 punten), Lichaamspijn (stijging van 18 punten) en de Geestelijke gezondheid (stijging van 24 punten) verbeteren ook, maar doordat deze wat hoger scoorden op de voormeting is het verschil in toename relatief wat lager. Op geaggregeerd niveau verbetert het lichamelijk welzijn met 9 punten en het geestelijk welzijn met 16 punten (op dit niveau wordt een andere rekensystematiek gebruikt dan bij de acht SF/RAND-36 variabelen, waardoor de stijging relatief lager is). Zoals eerder aangeven wordt een stijging van 5 punten als klinisch en sociaal relevant aangemerkt, mits de steekproef groot genoeg is (hetgeen hier het geval is). De stijgingen in dit onderzoek gaan daar ruimschoots overheen, waardoor veel waarde aan het effect kan worden toebedeeld.

Metingen in perspectief

De gemiddelde scores op de voor- en nametingen krijgen meer betekenis door deze af te zetten tegen normscores van representatieve bevolkingsonderzoeken. In 1991 is een landelijk bevolkingsonderzoek in de Verenigde Staten uitgevoerd met de SF/RAND-36 vragenlijst op basis van een steekproef (2474 respondenten) uit de totale populatie volwassen Amerikanen. Deze steekproef wordt als representatief beschouwd en kan als normering gebruikt worden om scores met elkaar te vergelijken (Ware et al., 1993). Om meer inzicht te krijgen in de lichamelijke en geestelijke gezondheidsproblematiek voor aanvang van de Ander Leven interventie, zijn de gemiddelde scores in de voormeting uitgesplitst naar lichamelijke en geestelijke klachten en vergeleken met normeringen van twee lichamelijke klachten (rugpijn en dermatitis) en een geestelijke klacht (klinische depressie) uit het Amerikaanse bevolkingsonderzoek (zie onderstaande tabel).

SF/RAND-36	Voormeting			Normeringen (VS)		
	Algemeen	Lichamelijk	Geestelijk	Rugpijn	Dermatitis	Clinische depressie
Algemene gezondheid	44	42	48	58	61	53
Lichamelijke gezondheid	70	73	76	66	75	72
Rolbeperkingen n.a.v. lichamelijke problemen	26	21	36	47	57	44
Lichaamspijn	62	59	67	59	67	59
Geestelijke gezondheid	53	54	50	75	76	46
Rolbeperkingen n.a.v. geestelijke problemen	46	52	38	71	79	39
Vitaliteit/energie	35	36	38	52	57	40
Sociaal functioneren	44	47	48	81	84	57

Uit het overzicht komen overeenkomsten en verschillen naar voren, waarbij het lastig is om een duidelijk patroon op te maken. Opvallend is dat deelnemers die alleen lichamelijke klachten hebben op Algemene gezondheid en op Rolbeperkingen n.a.v. lichamelijke problemen duidelijk lager scoren dan mensen met rugpijn of dermatitis. Ook valt het op dat deelnemers die alleen lichamelijke klachten hebben beduidend lager scoren op alle geestelijke variabelen ten opzichte van mensen met rugpijn en

dermatitis. De scores van deelnemers met alleen geestelijke klachten trekken meer naar de scores toe van mensen met een klinische depressie.

Een andere normering zijn de geaggregeerde lichamelijke en geestelijke welzijnscores. Het Amerikaanse bevolkingsonderzoek gaf een gemiddeld lichamenlijk en geestelijk welzijnsniveau van 50 punten aan, met een standaarddeviatie van 10 (gemiddeld wijken mensen 10 punten boven en 10 punten onder deze score af). Zoals eerder aangegeven is deze 50-puntenscore in de praktijk als standaardisatiescore aangenomen waarmee lichamelijke en geestelijke welzijnscores uit andere SF/RAND-36 onderzoeken vergeleken kunnen worden. In de regel geldt hoe lager of hoger de scores van 50 afwijken, hoe slechter of beter het betreffende welzijnsniveau van de onderzochte populatie. De gehanteerde rekensystematiek in dit onderzoek maakt ook dat de 50-puntenscore kan worden toegepast op de Nederlandse bevolking (zie bijlage 4)

Verder geeft een Brits bevolkingsonderzoek van Jenkinson et al. (1999), op een steekproef uit 1991 (n = 3256), aan dat mensen met een langdurige chronische ziekte op lichamenlijk welzijn gemiddeld 48 punten en op geestelijk welzijn gemiddeld 45 punten scoren. Aangezien deze groep overeen komt met de deelnemers aan de Ander Leven interventie, kunnen deze scores als vergelijkende indicatoren worden gebruikt.

De onderzoeksgroep scoorde 43 punten op lichamenlijk welzijn en 33 punten op geestelijk welzijn. Dit geeft aan dat deelnemers aan de Ander Leven interventie voor aanvang laag scoren op lichamenlijk welzijn en met name laag scoren op hun geestelijk welzijn. Dit impliceert dat de deelnemers bij binnenkomst serieuze lichamenlijke en/of geestelijke klachten ervaren.

De gemiddelde scores op de nametingen tezamen kunnen vergeleken worden met normeringen uit bevolkingsonderzoeken onder volwassenen om meer inzicht te krijgen in hoeverre de lichamenlijke en/of geestelijke gezondheidstoestand na toepassing van de Ander leven interventie normaliseert. De scores zijn vergeleken met het eerder genoemde Amerikaanse bevolkingsonderzoek in 1991 (n = 2474) en het Brits bevolkingsonderzoek in 1999 (n = 8889), alsmede met een Canadees bevolkingsonderzoek van Hopman et al. (2000) op een steekproef van 9423 respondenten en een Nederlands bevolkingsonderzoek in de gemeente Emmen van Van der Zee en Sanderma (2012) op een steekproef van 1063 respondenten (zie de tabel op de volgende pagina).

SF/RAND-36	Nameting Interventiegroep	Normeringen			Gemeente Emmen
		Verenigde Staten	Verenigd Koninkrijk	Canada	
Algemene gezondheid	76	72	71	77	73
Lichamenlijke gezondheid	85	84	88	86	87
Rolbeperkingen n.a.v. lichamenlijke problemen	74	81	87	82	79
Lichaamspijn	80	75	79	76	80
Geestelijke gezondheid	77	75	72	78	77
Rolbeperkingen n.a.v. geestelijke (emotionele) problemen	84	81	86	84	84
Vitaliteit/energie	70	61	58	66	67
Sociaal functioneren	82	83	83	86	87
Lichamenlijk welzijn	52	50	50*	51	n.b.
Geestelijk welzijn	49	50	50*	52	n.b.

* gemiddelde score tussen Britse mannen en vrouwen

Uit het overzicht komt naar voren dat de gemiddelde scores op de nametingen min of meer overeenkomen met de bevolkingsonderzoeken, met twee uitzonderingen daargelaten (waarbij een verschil van meer dan 10 punten is geconstateerd).

Een eerste uitzondering is dat deelnemers, na toepassing van de Ander Leven interventie, lager scoren op hun rolbeperkingen n.a.v. lichamenlijke problemen dan de bevolkingsonderzoeken (m.n. ten opzichte van het Verenigd Koninkrijk). Dit kan mogelijk verklaard worden doordat de deelnemers die bij Ander Leven binnenkomen uitzonderlijk sterke klachten ervaren op hun rolbeperkingen n.a.v. lichamenlijke problemen (deze scores zijn erg laag). Aan de andere kant kan het ook betekenen dat

deelnemers, ondanks de Ander Leven interventie, nog steeds wat meer rolbeperkingen n.a.v. hun lichamelijke problemen ervaren dan een gemiddelde westerling.

Een tweede uitzondering is dat deelnemers, na toepassing van de Ander Leven interventie, wat hoger scoren op hun vitaliteit/energie vergeleken met de normscores uit de bevolkingsonderzoeken (m.n. ten opzichte van het Verenigd Koninkrijk, alsook de Verenigde Staten: landen waar de werkdruk en stress over het algemeen hoog kunnen zijn). Een waarschijnlijke verklaring is dat veel deelnemers specifiek bij Ander Leven binnenkomen voor vitaliteits- en energieklaarten, zoals moeheid en weinig puf hebben en dat de interventie ook specifiek voor deze doelgroep is ontworpen. In ieder geval voelen deelnemers zich na de Ander Leven interventie over het algemeen vitaler en energiever dan een 'gemiddeld' Amerikaans of Brits staatsburger (de verschillen met iemand uit Canada of Emmen zijn kleiner; plekken in de wereld waar de werkdruk en stress waarschijnlijk ook wat lager liggen).

De onderzoeksgroep scoort op de geaggregeerde lichamelijke welzijnsscore 2 punten boven de standaardisatiescore van 50 (dus beter dan gemiddeld). De geestelijke welzijnsscore zit slechts 1 punt onder standaardisatiescore van 50, maar kent – zoals eerder aangegeven – een enorme stijging van 16 punten ten opzichte van de voormeting. Dit betekent dat deelnemers na de training op vrijwel hetzelfde welzijnsniveau zitten als een gemiddelde Nederlander.

Concluderend kan gesteld worden dat, gemiddeld genomen, Ander Leven deelnemers voor aanvang van de training serieuze lichamelijke en/of geestelijke gezondheidsklachten hebben en dat deelnemers na de training herstellen naar een normaal gezondheidsniveau.

6. Effectiviteit van de Ander Leven interventie

De onderzoeksresultaten van het effectonderzoek staan weergegeven in onderstaande tabel (en in bijlage 9). De scores in de cellen geven de toevalwaarden (p-waarden) aan en de scores tussen haakjes de sterkte van het effect ('Effect size'), mits er een significant verschil is gevonden (zie ook onderstaande legenda hoe de waarden te interpreteren).

	Non-parametrisch (significant verschil als $p < 0.01$)					Parametrisch (significant verschil als $p < 0.001$)				
	Friedman	Wilcoxon				RM	RM: within-subjects contrasts			
	t0-t4	t0-t1	t1-t2	t2-t3	t3-t4	t0-t4	t0-t1	t1-t2	t2-t3	t3-t4
Algemene gezondheid	0.000	0.000 (0.593)	0.516	0.616	0.235	0.000 (0.324)	0.000 (0.619)	0.549	0.779	0.392
Lichamelijke gezondheid	0.000	0.000 (0.528)	0.018	0.164	0.611	0.000 (0.283)	0.000 (0.434)	0.012	0.249	0.624
Rolbepalingen n.a.v. LG	0.000	0.000 (0.525)	0.068	0.781	0.410	0.000 (0.210)	0.000 (0.523)	0.056	0.867	0.361
Lichaamspijn	0.000	0.000 (0.450)	0.171	0.867	0.461	0.000 (0.139)	0.000 (0.358)	0.171	0.882	0.377
Geestelijke gezondheid	0.000	0.000 (0.573)	0.115	0.772	0.002	0.000 (0.279)	0.000 (0.558)	0.107	0.793	0.005
Rolbepalingen n.a.v. GG	0.000	0.000 (0.454)	0.071	0.848	0.057	0.000 (0.145)	0.000 (0.398)	0.083	0.802	0.057
Vitaliteit/energie	0.000	0.000 (0.607)	0.000 (0.147)	0.624	0.865	0.000 (0.353)	0.000 (0.693)	0.000 (0.113)	0.967	1.000
Sociaal functioneren	0.000	0.000 (0.582)	0.125	0.534	0.180	0.000 (0.310)	0.000 (0.602)	0.084	0.442	0.357
Lichamelijk welzijn (agg.)	0.000	0.000 (0.532)	0.115	0.452	0.388	0.000 (0.231)	0.000 (0.455)	0.351	0.572	0.228
Geestelijk welzijn (agg.)	0.000	0.000 (0.598)	0.033	0.987	0.013	0.000 (0.276)	0.000 (0.580)	0.003	0.853	0.014

In de cellen staan de toevalwaarden (p-waarden) weergegeven: bij een significant verschil staat de Effect size tussen haakjes vermeld

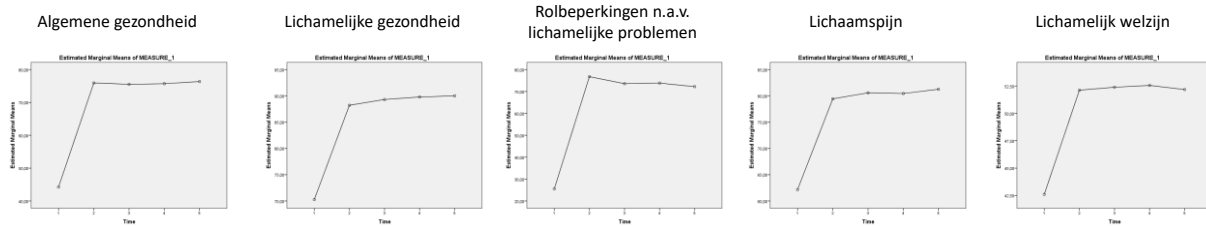
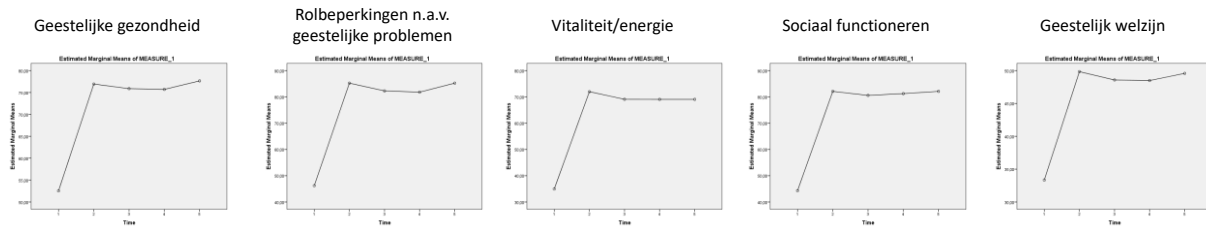
- **Lezen van toevalwaarden:** bij de non-parametrische testen moet de toevalwaarde kleiner zijn dan 1% (0.01) en bij de parametrische testen kleiner dan 0,1% (0.001) om een significant verschil aan te tonen.
- **Lezen van Effect size:** bij de Wilcoxon test is de sterkte van het effect zwak bij scores rond de 10% (0.1), medium sterk bij scores rond de 30% (0.3) en groot bij scores rond de 50% (0.5) of hoger. Deze rangschikking geldt ook voor de 'within-subjects contrast' binnen de RM test. De sterkte van het effect bij de algemene RM test is echter anders: deze is zwak rond de 1% (0.01), medium sterk rond de 6% (0.06) en erg sterk bij 14% (0.14) of hoger.

Zowel de non-parametrische als de parametrische testen geven aan dat er een significant verschil is tussen de voormeting ten opzichte van de nametingen en tussen de voormeting en eerste nameting op alle SF/RAND-36 variabelen, alsmede dat de sterkte van het effect daarvan (zeer) groot is. Op alle testen is de kans op toeval 0% ($p = 0.000$), hetgeen betekent dat het geconstateerde verschil tussen de tijdsmetingen niet op toeval berust. Doordat de non-parametrische en parametrische test beiden een significant verschil aantonen, kan met nog meer zekerheid aangegeven worden dat het verschil ook daadwerkelijk optreedt. Hierbij dient te worden opgemerkt dat bij de RM test de assumptie van sphericiteit is geschonden (zie ook bijlage 4) en dat een Bonferroni correctie is toegepast via de Greenhouse-Geisser test en Huynh-Feldt test, alsmede een multivariate test via Pillai's Trace en Wilks' Lambda (zie bijlage 10 voor meer uitleg). Maar op al deze testen blijft de kans op toeval 0% ($p = 0.000$), hetgeen de uitkomst versterkt.

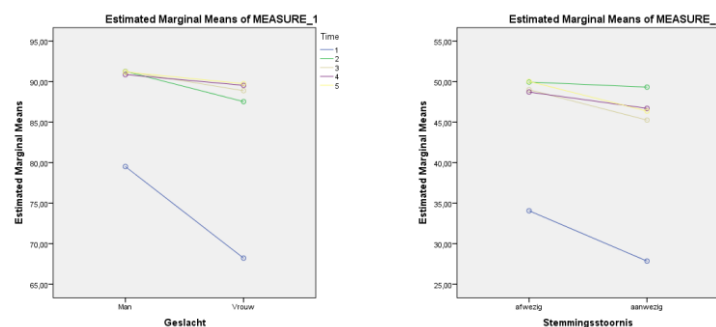
De nametingen op alle SF/RAND-36 variabelen verschillen niet significant van elkaar, behalve bij Vitaliteit/energie. Daar is een significant verschil gevonden tussen de eerste en tweede nameting. Dit impliceert dat de vitaliteit/energie na de Ander Leven training nog kan toenemen (bijvoorbeeld door thuis de AC-processen te lopen) en/of dat het geconstateerde verschil is toe te wijzen aan de verdieping/terugkomdag.

De uitkomsten geven aan dat, binnen de eerder genoemde onderzoekbeperkingen, met grote zekerheid kan worden aangenomen dat het verschil tussen de voor- en nametingen aan de Ander Leven interventie (m.n. de driedaagse training) kan worden toegewezen. Met andere woorden, de Ander Leven interventie heeft een bewezen ('evidence based') effect opgeleverd.

In onderstaande grafieken (z.o.z.) wordt het verschil tussen de voor- en nametingen op de SF/RAND-36 variabelen visueel weergegeven. Er is een duidelijk verschil waarneembaar tussen de voormeting en de eerste nameting (de lijn gaat stijl omhoog) en er is relatief weinig verschil waarneembaar tussen de nametingen (de lijn blijft min of meer op hetzelfde niveau).

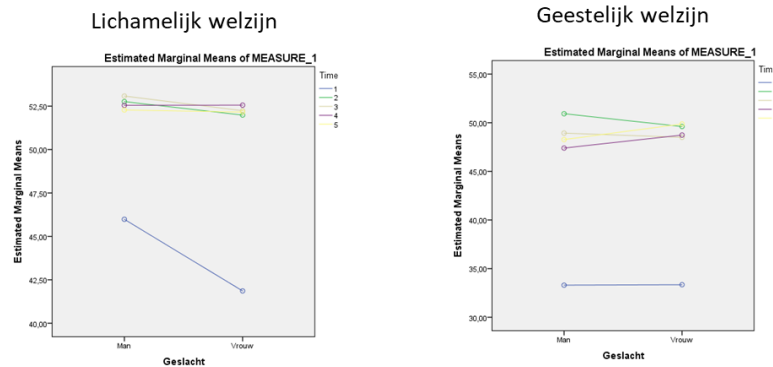
Lichamelijk**Geestelijk****Verskil tussen factoren op de voor- en nametingen**

Een significant verschil tussen de voormeting en eerste nameting is niet toe te wijzen aan een externe factor, zoals geslacht, leeftijd, nationaliteit, inkomenssituatie of type klacht. Er waren enkele uitzonderingen, maar het verschil tussen de factoren (bijvoorbeeld tussen man en vrouw of tussen mensen met of zonder stemmingsstoornis) is te verklaren doordat de scores op de voormeting veel van elkaar afwijken (de lichamelijke gezondheid van vrouwelijke deelnemers voor aanvang van de interventie scoort duidelijk lager dan de lichamelijke gezondheid van mannelijke deelnemers). In onderstaande grafieken komen de verschillen op de voormeting c.q. baseline goed naar voren: het verschil tussen de voormeting (blauwe lijn) en de eerste nameting (groene lijn) is hoger door de baseline verschillen op de voormeting (vrouwen scoren op de voormeting bijvoorbeeld duidelijk lager dan mannen).

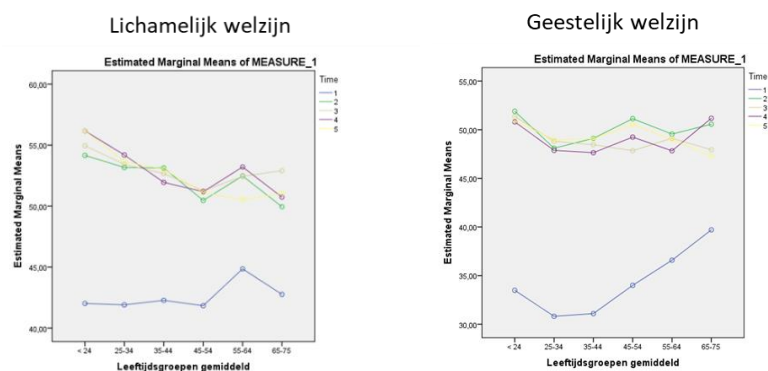


De verschillen op de voormetingen zijn in sommige gevallen opvallend. De belangrijkste staan hieronder weergegeven:

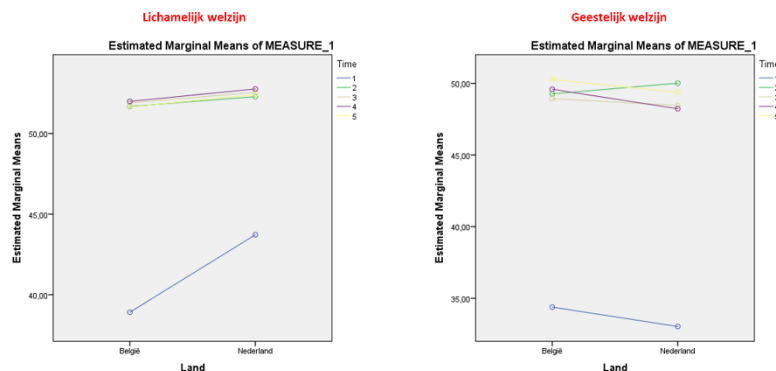
- Vrouwelijke deelnemers scoren voor aanvang van de training beduidend slechter op hun lichamelijk welzijn dan mannelijke deelnemers (het geestelijk welzijn voor mannen en vrouwen is op de voormeting vrijwel gelijk).



- Deelnemers in de leeftijd van 55-75 jaar scoren voor aanvang van de training hoger op hun lichamelijk welzijn en met name hun geestelijk welzijn dan mensen die jonger zijn (z.o.z.).



- Vlaamse deelnemers scoren voor aanvang van de training wat slechter op hun lichamelijk welzijn dan Nederlandse deelnemers (op geestelijk welzijn is dat vrijwel gelijk)

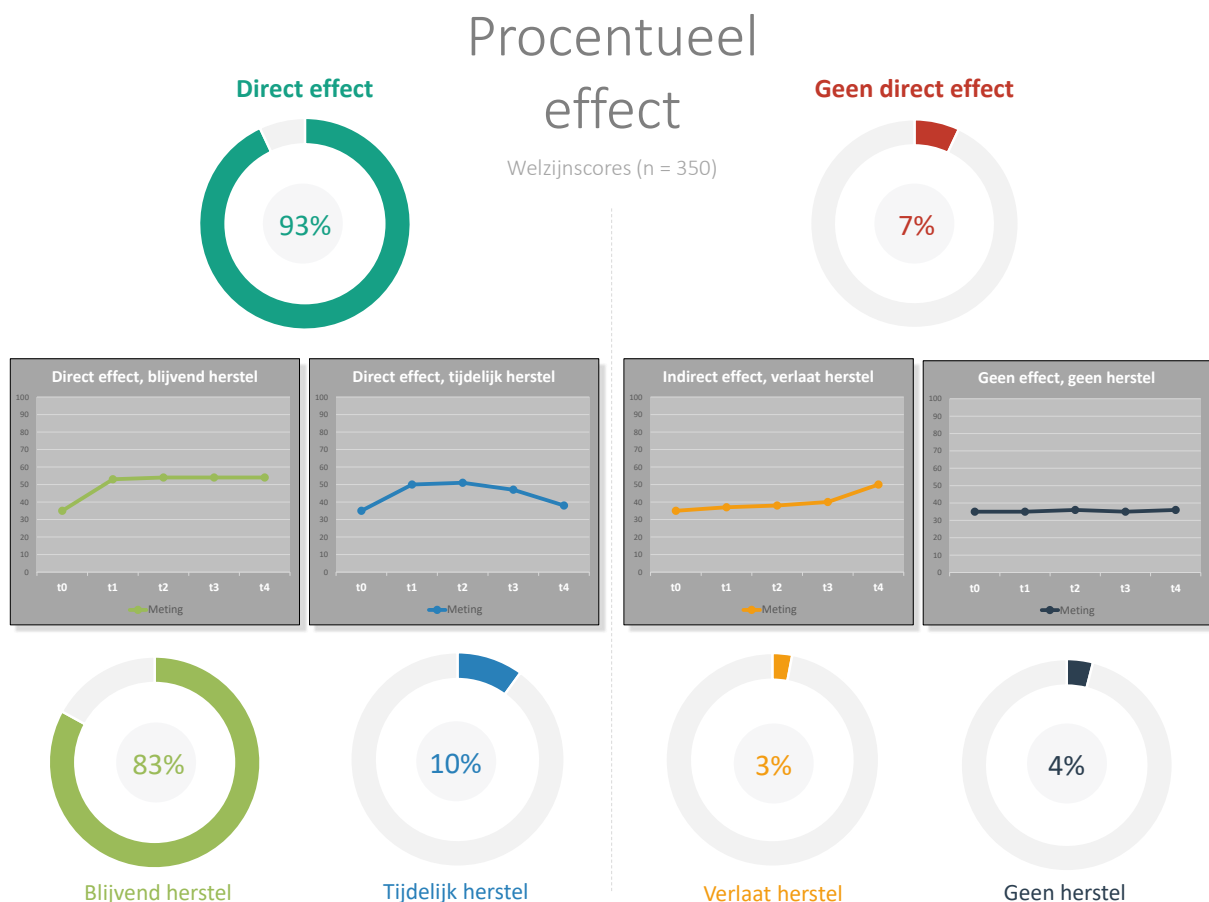


Naast externe factoren is ook gekeken naar karakteristieken van de Ander Leven Interventie zelf (zie ook bijlage 9). De resultaten laten zien dat het niet uitmaakt welke hooftrainer voor de groep staat, hoe groot de groep is (al scoren kleine groepen wat lager dan grote groepen) en welk type begeleider de deelnemer meeneemt (partner of geen-partner). De Ander Leven interventie heeft zich in de tijd doorontwikkeld. Hierbij kan worden aangegeven dat na toevoeging van een buddy kiezen, de dansavond, het begeleiden van partners in AC-processen en het vragenurtje, de lichamelijke welzijnsscores wat hoger scoren.

7. Percentage deelnemers dat duurzaam herstelt

De effectiviteit van de Ander Leven interventie is bepaald op basis van de gemiddelde scores op de voor- en nametingen, maar dit geeft nog niet aan hoeveel procent van de deelnemers een direct effect na de driedaagse training ervaart en duurzaam herstelt. De gegevens zijn nader geanalyseerd om daar meer inzicht in te geven. Hierbij zijn de geaggregeerde lichamelijke en geestelijke welzijnsscores gebruikt, waarbij een deelnemer minimaal 5 punten verschil moet hebben tussen de voor- en nameting en het herstel na een jaar boven de kritische gezondheidsgrens van 40 punten uitkomt. Deze grens is bepaald op basis van de uitgangspunten van Ware et al. (1993). Zij geven aan dat 50 punten een normaal gemiddelde vormt met een standaarddeviatie van 10 punten. Deze standaarddeviatie is als bandbreedte genomen om bij 40 een kritische gezondheidsgrens te trekken en 'herstel' aan te duiden.

Uit de analyse komt naar voren dat bij 93% van de deelnemers een direct trainingseffect optreedt. Dit effect is bij 83% van deze groep ook blijvend (duurzaam herstel). Bij 10% van deze groep treedt een tijdelijk herstel op en vallen na een jaar terug tot onder de kritische gezondheidsgrens. De training heeft op 7% van de deelnemers geen direct trainingseffect, maar bij 3% van deze groep treedt na een jaar alsnog een indirect effect en verlaat herstel op. Op 4% van de deelnemers heeft de training geen effect en vindt er ook geen herstel plaats. In onderstaande figuur worden de procentuele verschillen visueel weergegeven.



8. Conclusies

De Ander Leven interventie bestaat uit een driedaagse training en een aantal onderdelen daaromheen (o.a. een persoonlijk consult, verdiepingsdag en telefonische follow-ups), met als doel om lichamelijke en geestelijke klachten bij (chronisch) zieke mensen te verminderen. De vraag is in hoeverre deze interventie effectief is: verbetert de lichamelijke en geestelijke gezondheid van deelnemers na de interventie en is het resultaat blijvend?

Om antwoord op deze vraag te geven is effectonderzoek uitgevoerd. Voor aanvang van de driedaagse training hebben deelnemers een gevalideerde vragenlijst ingevuld over hun lichamelijke en geestelijke gezondheidstoestand (SF/RAND-36), alsmede 1, 3, 6 en 12 maanden na de training. Door de scores van de deelnemers voor aanvang van de training (voormeting) te vergelijken met de scores van de deelnemers op de verschillende tijdstippen na de training (nametingen), kan meer gezegd worden over de effectiviteit van de Ander Leven interventie.

In de periode 2011-2017 hebben in totaal 1307 mensen de Ander Leven de training gevolgd. Uit deze populatie is een onderzoeksgroep van 350 deelnemers naar voren gekomen die de vragenlijst op alle voor- en nametingen volledig heeft ingevuld. Dit vormt de steekproef om de effectiviteit van de interventie te evalueren. Statistische procedures wijzen uit dat de steekproefomvang groot genoeg is om de resultaten te generaliseren naar de totale populatie van deelnemers.

Op basis van de steekproef kan geconcludeerd worden dat de Ander Leven interventie zeer effectief is. De resultaten tonen aan dat de gemiddelde lichamelijke en geestelijke gezondheidstoestand van deelnemers voor aanvang van de Ander Leven interventie slecht scoort (deze liggen lager dan normeringen voor mensen met langdurige chronische lichamelijke en geestelijke klachten) en dat deze na toepassing van de interventie normaliseert naar een normale gezondheidstoestand (overeenkomstig met een gemiddelde Nederlander en westerling). Verschillende statistische testen in dit onderzoek tonen allemaal aan dat het gemeten gezondheidsverschil tussen de voor- en nametingen niet op toeval berust en dat de sterkte van het effect zeer groot is. Dus de geconstateerde gezondheidsverbetering kan met grote zekerheid aan de Ander Leven interventie worden toegewezen. Ook houdt het effect stand in de tijd: de gemiddelde welzijnsscore van alle deelnemers in de steekproef blijft na 1, 3, 6 en 12 maanden op vrijwel hetzelfde genormaliseerde niveau.

Het geconstateerde effect van de interventie is voornamelijk van toepassing op de driedaagse training. De verdiepingsdag heeft over het algemeen weinig tot geen effect meer op de gezondheidstoestand van deelnemers, behalve voor de vitaliteit/energie van deelnemers. Deze verbetert nog significant na de verdiepingsdag. Ook is het opvallend dat het gemiddeld vitaliteitsniveau van deelnemers na de training hoger ligt dan het landelijk gemiddelde van een Westerse samenleving.

Het effect van de Ander Leven interventie hangt niet af van geslacht, leeftijd, nationaliteit of inkomenssituatie. Het effect lijkt ook niet af te hangen van het type klacht. Dit was moeilijker te onderzoeken, omdat bepaalde klachten beperkt in de steekproef vertegenwoordigd waren. Over het algemeen kan worden aangegeven dat de gezondheidstoestand van deelnemers, onafhankelijk van hun type klacht, significant verbetert na toepassing van de Ander Leven interventie. Ook eventuele variaties in de interventie hebben weinig tot geen verschil uitgemaakt. Het maakt bijvoorbeeld niet uit welke hoofdtrainer er voor de groep staat, of de deelnemer een partner of een andere naaste als begeleider meeneemt en hoe groot de groep is (al scoren kleine groepen wat minder goed).

Procentueel gezien ervaart 93% van de deelnemers een direct trainingseffect, oftewel, een verbetering in hun lichamelijk en/of geestelijk welzijn. Bij 83% van deze groep is het effect blijvend en treedt

duurzaam herstel op. Bij 10% van deze groep is het effect tijdelijk van aard (deze herstellen niet op de lange termijn en vallen terug naar een ongezond welzijnsniveau). Van de 7% die geen direct trainingseffect ervaart, herstelt uiteindelijk 3% na een jaar alsnog na een gezond welzijnsniveau (verlaat effect). Bij 4% van de deelnemers treedt geen effect en geen herstel op.

De onderzoeksresultaten in dit onderzoek wijzen uit dat de Ander Leven interventie effectief is, maar dit onderzoek kent ook een aantal beperkingen. Daardoor moet er met enige nuance naar de uitkomsten gekeken worden. Ten eerste zijn de deelnemers op basis van bepaalde criteria geselecteerd om deel te nemen aan de interventie. Er heeft geen aselecte selectie plaats gevonden: niet elke aanmelding heeft evenveel kans op de training te volgen. Daarom geldt de uitspraak alleen voor mensen met lichamelijke en geestelijke klachten die aan deze criteria voldoen, bijvoorbeeld voldoende gemotiveerd om de training te gaan volgen. Het is onbekend hoe mensen met lichamelijke en/of geestelijke klachten die niet aan deze criteria voldoen op de Ander Leven interventie reageren.

Ten tweede zijn deelnemers niet gerandomiseerd (op basis van toeval) aan de onderzoeksgroep toegewezen. In de steekproef zitten dus mogelijk deelnemers die op basis van een bepaalde motivatie de vragenlijst op alle tijdstippen hebben ingevuld (bijvoorbeeld deelnemers die meer tevreden of meer tijd hebben om de vragenlijst in te vullen). Dit zou de onderzoeksresultaten kunnen verkleuren.

Ten derde zijn deelnemers niet bevraagd op de mate waarin zij de aangeleerde technieken zelfstandig toepassen. Tijdens de training leren deelnemers bijvoorbeeld een imaginatieoefening te doen (AC-proces lopen) en deze thuis ook toe te passen. Het welzijnseffect zou van deze toepassing af kunnen hangen. Nu is bijvoorbeeld niet te controleren of deelnemers waarop de training geen effect heeft of waar geen blijvend herstel optreedt, te maken heeft met het al dan niet uitvoeren van de aangeleerde technieken. Of dat deelnemers die de technieken meer toepassen ook hoger scoren op hun welzijnsniveau.

Tot slot is de onderzoeksgroep niet vergeleken met een controlegroep (mensen met overeenkomstige klachten die de Ander Leven interventie niet ondergaan), waardoor minder uitsluitsel over het effect kan worden gegeven.

Literatuur

- Aaronson, N. K. et al., Muller, M., Cohen, P. D., Essink-Bot, M., Fekkes, M., Sanderman, R., Sprangers, M.A.G., Te Velde, A., Verrips, E. (1998). Translation, Validation, and Norming of the Dutch Language Version of the SF-36 Health Survey in Community and Chronic Disease Populations. *Journal of Clinical Epidemiology*, 51(11), 1055-1068.
- Bjorner, J., Wolden, M., Gundgaard, J., Bayliss, M., & Miller, K. (2013). Benchmarks for interpretation of score differences on the SF-36 health survey. *Value in Health*, 16(3), A35.
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics*. Los Angeles: Sage.
- George, D., & Mallery, M. (2010). *SPSS for Windows Step by Step: A Simple Guide and Reference*, 17.0 update (10a ed.) Boston: Pearson.
- Hopman, W. et al. (2000). Canadian normative data for the SF-36 health survey. *Canadian Medical Association Journal*, 163 (3):265-271
- Jenkinson, C., Stewart-Brown, S., Petersen, S., & Paice, C. (1999). Assessment of the SF-36 version 2 in the United Kingdom. *Journal of Epidemiol Community Health*, 53:46-50
- Kahler, E., Rogausch, A., Brunner, E., & Himmel, W. (2008). A parametric analysis of ordinal quality-of-life data can lead to erroneous results. *Journal of Clinical Epidemiology*, 61(5), 475-480.
- Katz, M. H. (2010). *Evaluating clinical and public health interventions: A practical guide to study design and statistics*. Cambridge: Cambridge University Press.
- Knapp, T. R. (1990). Treating Ordinal Scales as Interval Scales. *Nursing Research*, 39(2), 121-123.
- Orwelius, L., Nilsson, M., Nilsson, E., Wenemark, M., Walfridsson, U., Lundström, M., Taft, C., Palaszewski, B. & Kristenson, M. (2018). The Swedish RAND-36 Health Survey - reliability and responsiveness assessed in patient populations using Svensson's method for paired ordinal data. *Journal of Patient-Reported Outcomes*, 2(1). doi:10.1186/s41687-018-0030-0
- Van der Zee, K.L. & Sanderman, R. (2012). *Het meten van de algemene gezondheidstoestand met de Rand-36*. Groningen: UMCG/Rijksuniversiteit Groningen, Research Institute SHARE.
- Walters, S. J. (2009). *Quality of life outcomes in clinical trials and health-care evaluation: A practical guide to analysis and interpretation*. Chichester, U.K: John Wiley & Sons.
- Ware, J. E., Snow, K. K., Kosinski, M., & Gandek, B. (1993). *SF-36 Health Survey: Manual and Interpretation Guide*. Boston: New England Medical Center. The Health Institute.
- Ware, J.E., Kosinski, M. & Keller, S.D. (1994). *SF-36 physical and mental health summary scales: A user's manual*, 2nd ed. The Health Institute, Boston MA.

Bijlage 1: Bedrijfsprofiel

Evidento is een bureau dat gespecialiseerd is in het opzetten en uitvoeren van wetenschappelijk effectonderzoek en het ontwikkelen van evaluatiesystemen voor organisaties. Het bureau richt zich met name op organisaties in zorg en welzijn.

Organisaties voeren diverse processen, interventies en activiteiten uit, maar weten doorgaans niet wat deze uiteindelijk opleveren en in hoeverre de gewenste resultaten geboekt worden. Evidento helpt om inzicht te geven in de effectiviteit van processen, interventies en activiteiten (bijvoorbeeld in hoeverre cliënten naar aanleiding van een interventie vooruitgang boeken). Dit gebeurt aan de hand van wetenschappelijk effectonderzoek. Evidento verzamelt data, voert geavanceerde statistische analyses uit via SPSS en levert een onderzoeksverslag op. Met het verslag kunnen organisaties op een valide, betrouwbare en representatieve manier aantonen in hoeverre hun processen, interventies en activiteiten de gewenste effecten opleveren ('evidence based' zijn).

Door technologische ontwikkelingen wordt het voor organisaties mogelijk om op een gebruiksvriendelijke, kostenbesparende en doeltreffende manier *big data* te verzamelen en te analyseren. Evidento begeleidt dit proces en helpt organisaties een evaluatiesysteem te ontwikkelen via een Business Intelligence (BI) systeem. In dit systeem wordt het verzamelen en analyseren van data geautomatiseerd en worden de analyseresultaten weergegeven in een online dashboard waarop gebruikers kunnen inloggen. Het dashboard geeft gerichte stuurinformatie over wat precies te optimaliseren om tot betere resultaten te komen. Daarmee kunnen organisaties hun kwaliteit en succes op een systematische manier vergroten.

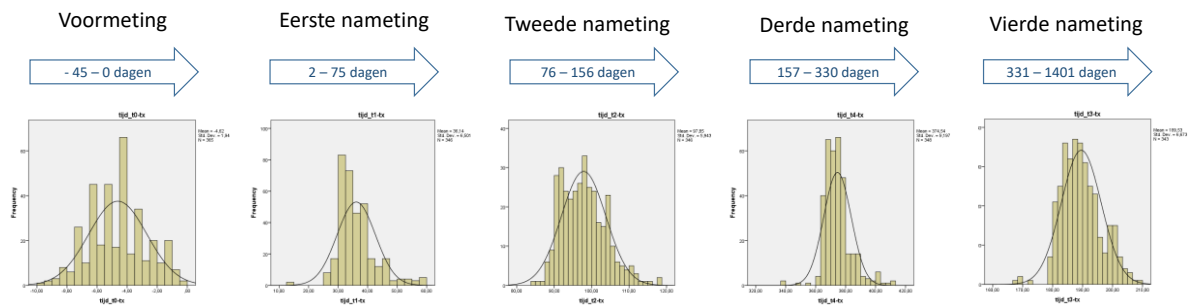
Evidento is een bureau dat sinds 1 juni 2016 staat ingeschreven bij de Kamer van Koophandel in Nederland en is gevestigd aan de Keizersgracht 241 in Amsterdam. Tim Choy (1973) is de oprichter en eigenaar van Evidento. Daarnaast is hij verbonden aan Familiezorg Nederland, een platform om familiezorg ontwikkeling, onderzoek en onderwijs te stimuleren. In het verleden heeft hij Beleid, Communicatie en Organisatie (BCO) aan de Vrije Universiteit gestudeerd en promotieonderzoek aan de Universiteit van Tilburg verricht (onderzoek naar de effecten van evaluatiesystemen op de structuur en processen binnen organisaties en het menselijk gedrag). Hij heeft gewerkt voor verschillende universiteiten (Vrije Universiteit Amsterdam, Universiteit van Tilburg, Radboud Universiteit Nijmegen) en onderzoeksbureaus (TNS/NIPO, fluent Zorgadvies, Prismant).

Tim Choy is sociale wetenschapper en heeft zich verdiept in bepaalde onderwerpen binnen de organisatiekunde, sociologie, psychologie en statistiek. In de beginperiode van zijn carrière heeft hij zich voornamelijk gericht op welke evaluatiesystemen organisaties gebruiken, hoe ze werken en wat het oplevert. Gaandeweg is hij zich steeds meer gaan bezighouden met het zelf opzetten en uitvoeren van evaluatieonderzoek. Hij heeft zich met name gespecialiseerd in het evalueren van zorginterventies op effectiviteit. De afgelopen vijftien jaar heeft hij effectonderzoek verricht bij verschillende soorten zorgaanbieders in de Care sector (instellingen voor Geestelijke gezondheidszorg, Gehandicaptenzorg, Ouderenzorg en Jeugdzorg). In het bijzonder heeft hij veel effectonderzoek gedaan naar de Methode Familiezorg en welke invloed familiecommunicatie heeft op het stressniveau en het welzijn van mantelzorgers. Hij is bekwaam in het opstellen van onderzoeksdesigns voor effectonderzoek en het uitvoeren van geavanceerde statistische analyses in SPSS. Ook heeft hij aan de Radboud Universiteit Nijmegen een certificaat Geavanceerde Statistiek voor promovendi behaald.

Voor meer informatie neem contact op via: tim@evidento.nl of bel 06-13839493.

Bijlage 2: Tijdsmetingen

Deelnemers behoren de vragenlijst 1 week voor aanvang van de driedaagse training en 1, 3, 6 en 12 maanden na de training in te vullen. Maar in de praktijk gebeurt dit niet op gezette tijden. Daarom zijn bepaalde bandbreedtes gebruikt om te bepalen bij welke tijdsmeting de ingevulde vragenlijst behoort. Deze staan hieronder weergegeven, inclusief een histogram wat aantoont dat het aantal dagen binnen de bandbreedte (wanneer de vragenlijst wordt ingevuld) kan variëren.



Gemiddeld wordt de voormeting 4,6 dagen voor de training ingevuld, de eerste nameting 36,1 dagen na de training, de tweede nameting 97,9 dagen na de training, de derde nameting 189,5 dagen na de training en de vierde nameting 374,5 dagen na de training.

De maximale afstand tussen de metingen is bij de voormeting -9,5 dagen (laagste -9,5 dagen, hoogste -0,4 dagen), bij de eerste nameting 46,1 dagen (laagste 12,7 dagen, hoogste 58,8 dagen), bij de tweede nameting 33,7 dagen (laagste 84,8 dagen, hoogste 118,5 dagen), bij de derde nameting 40,9 dagen (laagste 167,7 dagen, hoogste 208,6 dagen) en bij de vierde nameting 75,3 dagen (laagste 337,5 dagen, hoogste 412,8 dagen).

Bijlage 3: SF/RAND-36 vragenlijst (NL)

Wat vindt u, over het algemeen genomen, van uw gezondheid ?

1	uitstekend	zeer goed	goed	matig	slecht
---	------------	-----------	------	-------	--------

In vergelijking met een jaar geleden, hoe zou u NU uw gezondheid in het algemeen beoordelen ?

2	veel beter	iets beter	hetzelfde	iets slechter	veel slechter
---	------------	------------	-----------	---------------	---------------

De volgende vragen gaan over uw dagelijkse bezigheden.

Wordt u door uw gezondheid OP DIT MOMENT beperkt bij deze bezigheden.

Zo ja, in welke mate ?

		ernstig beperkt	beetje beperkt	niet beperkt
3	FORSE INSPANNING (zoals hardlopen, zware voorwerpen tillen, inspannend sporten,...)			
4	MATIGE INSPANNING (zoals het verplaatsen van een tafel, stofzuigen, fietsen,...)			
5	Tillen of boodschappen dragen			
6	EEN PAAR trappen oplopen			
7	EEN trap oplopen			
8	buigen, knielen of bukken			
9	meer dan een kilometer stappen			
10	een halve kilometer stappen			
11	honderd meter stappen			
12	uzelf wassen en aankleden			

Had u, tengevolge van uw lichamelijke gezondheid, DE AFGELOPEN 4 WEKEN een van de volgende problemen bij uw werk of andere dagelijkse bezigheden ?

		ja	nee
13	U heeft minder TIJD kunnen besteden aan werk of andere bezigheden		
14	U heeft MINDER BEREIKT dan u zou willen		
15	U was beperkt in het SOORT WERK of het soort bezigheden		
16	U had moeite met het werk of andere bezigheden (het kostte u bv extra inspanning)		

Had u, tengevolge van emotionele problemen (bijv. doordat u zich depressief of angstig voelde), DE AFGELOPEN 4 WEKEN een van de volgende problemen bij uw werk of andere dagelijkse bezigheden ?

		ja	nee
17	U heeft minder TIJD kunnen besteden aan werk of andere bezigheden		
18	U heeft MINDER BEREIKT dan u zou willen		
19	U heeft werk of andere bezigheden niet zo zorgvuldig gedaan als u gewend bent.		

In hoeverre heeft uw lichamelijke gezondheid of hebben uw emotionele problemen u DE AFGELOPEN 4 WEKEN belemmerd in uw normale sociale bezigheden met gezin, vrienden, burens of anderen ?

20	helemaal niet	enigszins	nogal	veel	heel erg veel
----	---------------	-----------	-------	------	---------------

Hoeveel pijn had u DE AFGELOPEN 4 WEKEN ?

21	geen	heel licht	licht	nogal	ernstig	heel ernstig
----	------	------------	-------	-------	---------	--------------

In welke mate heeft pijn u DE AFGELOPEN 4 WEKEN belemmerd bij uw normale werkzaamheden (zowel werk buitenshuis als huishoudelijk werk) ?

22	helemaal niet	een klein beetje	nogal	veel	heel erg veel
----	---------------	------------------	-------	------	---------------

Deze vragen gaan over hoe u zich de AFGELOPEN 4 WEKEN heeft gevoeld.

Wilt u bij elke vraag het antwoord aankruisen dat het beste aansluit bij hoe u zich heeft gevoeld ?

Hoe vaak gedurende de AFGELOPEN 4 WEKEN ...

	voortdurend	meestal	vaak	soms	zelden	nooit
23	Voelde u zich levenslustig ?					
24	Voelde u zich erg zenuwachtig ?					
25	Zat u zo erg in de put dat niets u kon opvrolijken ?					
26	Voelde u zich kalm en rustig ?					
27	Voelde u zich erg energiek ?					
28	Voelde u zich neerslachtig en somber ?					
29	Voelde u zich uitgeblust ?					
30	Voelde u zich gelukkig ?					
31	Voelde u zich moe ?					

HOE VAAK hebben uw lichamelijke gezondheid of emotionele problemen gedurende DE AFGELOPEN 4 WEKEN uw sociale activiteiten (zoals bezoek aan vrienden of naaste familieleden) belemmerd ?

32	Voortdurend	meestal	soms	zelden	nooit
----	-------------	---------	------	--------	-------

Wilt u het antwoord kiezen dat het beste weergeeft hoe juist of onjuist u elk van de volgende uitspraken voor uzelf vindt.

	volkomen juist	grotendeels juist	weet ik niet	grotendeels onjuist	volkomen onjuist
33	Ik lijk makkelijker ziek te worden dan andere mensen				
34	Ik ben net zo gezond als andere mensen die ik ken				
35	Ik verwacht dat mijn gezondheid achteruit zal gaan				
36	Mijn gezondheid is uitstekend.				

Bijlage 4: Rekensystematiek RAND-36 dimensies

Het omzetten van de scores op de afzonderlijke items c.q. antwoordcategorieën naar de SF/RAND-36 dimensies gebeurt op basis van de rekensystematiek van Ware et al. (1993) en een daarvan afgeleide SPSS syntax van Terwee (2002). Meer specifiek heeft de laatste haar syntax gebaseerd op Van der Zee & Sanderman (1993), Ware et al., (1994) en Aaronson et al. (1998). Deze syntax hoort bij de Nederlandse RAND-36 (deze wijkt iets af van de SF-36). De oorspronkelijke Engelstalige benamingen zijn omgezet naar Nederlandse benamingen:

- Physical functioning (PF) → Lichamelijke gezondheid
- Role – physical (RP) → Rolbeperkingen n.a.v. lichamelijke problemen
- Bodily Pain (BP) → Lichaamspijn
- General Health (GH) → Algemene gezondheid
- Mental Health (MH) → Geestelijke gezondheid
- Role – mental (RM) → Rolbeperkingen n.a.v. geestelijke problemen
- Vitality (VT) → Vitaliteit/energie
- Social functioning(SF) → Sociaal functioneren
- Physical Component Score (PCS) → Lichamelijk welzijn
- Mental Component Score (MCS) → Geestelijk welzijn

De gezondheidsverandering (Health Transformation) is uit dit onderzoek gelaten, omdat de vraag antwoord geeft hoe de gezondheidstoestand in vergelijking met een jaar geleden was. In deze effectmeting is uitgegaan van de huidige gepercipieerde gezondheidsbeleving op een specifiek moment in de tijd: 1 week voor de training en respectievelijk 1, 3, 6 en 12 maanden na de training.

Bij de geaggregeerde welzijnsscores zijn eerst de z-scores berekend ten opzichte van de algemene Nederlandse bevolkingspopulatie. Deze geven de overeenkomst of afwijking ten opzichte van de referentiepoulatie weer in standaard deviaties. De referentiepoulatie is hierbij gezet op 50 en de standaard deviatie op 10. Bij een score boven de 50 scoort de steekproef gemiddeld ‘beter’ dan de referentiepoulatie en bij een score onder de 50 scoort de steekproef gemiddeld ‘slechter’ dan referentiepoulatie. Voor de gebruikte referentiewaarden heeft Terwee (2002) de uitgangspunten van Aaronson et al. (1998) gebruikt. Bij het berekenen van de geaggregeerde welzijnsscores zijn de schalen gewogen opgeteld. De gewichten die hiervoor worden gebruikt, zijn gebaseerd op een regressievergelijking die in een Amerikaanse populatie is gemaakt (Ware et al., 1993), want er zijn geen Nederlandse getallen bekend.

De SPSS-syntax voor het berekenen van de somscores op de RAND-36 variabelen staat hieronder weergegeven.

```
* missing values pf1 eff pf2 pf3 pf4 pf5 pf6 pf7 pf8 pf9 pf10 prf1 prf2 prf3 prf4 erf1 erf2 erf3 sf1 sf2 pn1 pn2 mh1 mh2 mh3
mh4 mh5 en1 en2 en3 en4 ghp1 ghp2 ghp3 ghp4 ghp5 (9).
execute.

* physical functioning PF*.
compute nsom_pf = 10*sum.5(pf1, pf2, pf3, pf4, pf5, pf6, pf7, pf8, pf9, pf10) /Nvalid(pf1, pf2, pf3, pf4, pf5, pf6, pf7, pf8,
pf9, pf10).
compute RAND_PF=((nsom_pf-10)/20)*100.
execute.

* Role functioning physical RP.
compute nsom_rp = 4*sum.2(prf1, prf2, prf3, prf4) /Nvalid(prf1, prf2, prf3, prf4).
compute RAND_RP=((nsom_rp-4)/4)*100.
execute.
```

* Role functioning emotional RE.

Compute nsom_re = 3*sum.2(erf1, erf2, erf3) /Nvalid(erf1, erf2, erf3).

compute RAND_RE=((nsom_re-3)/3)*100.

execute.

* Social functioning SF.

RECODE sf1 (1=5) (2=4) (3=3) (4=2) (5=1) INTO sf1a.

compute nsom_sf = 2*sum.1(sf1a,sf2) /Nvalid(sf1a,sf2).

compute RAND_SF=((nsom_sf-2)/8)*100.

execute.

* Bodily pain BP.

RECODE pn1 (1=6) (2=5) (3=4) (4=3) (5=2) (6=1) into pn1a.

RECODE pn2 (1=5) (2=4) (3=3) (4=2) (5=1) into pn2a.

Compute nsom_bp=(5*pn1a)+(6*pn2a).

compute RAND_BP=((nsom_bp-11)/49)*100.

execute.

*Mental health MH.

Recode mh3 mh5 (1=6) (2=5) (3=4) (4=3) (5=2) (6=1) INTO mh3a mh5a.

compute nsom_mh = 5*sum.3(mh1, mh2, mh3a, mh4, mh5a) /Nvalid(mh1, mh2, mh3a, mh4, mh5a).

compute RAND_MH=((nsom_mh-5)/25)*100.

execute.

* vitality VT.

Recode en1 en2 (1=6) (2=5) (3=4) (4=3) (5=2) (6=1) INTO en1a en2a.

compute nsom_vt = 4*sum.2(en1a, en2a, en3, en4) /Nvalid(en1a, en2a, en3, en4).

compute RAND_VT=((nsom_vt-4)/20)*100.

execute.

* General Health GH.

Recode ghp1 ghp3 ghp5 (1=5) (2=4) (3=3) (4=2) (5=1) INTO ghp1a ghp3a ghp5a.

compute nsom_gh = 5*sum.3(ghp1a, ghp2, ghp3a, ghp4, ghp5a) /Nvalid(ghp1a, ghp2, ghp3a, ghp4, ghp5a).

compute RAND_GH=((nsom_gh-5)/20)*100.

execute.

VARIABLE LABELS RAND_PF "RAND36 physical functioning PF".

VARIABLE LABELS RAND_RP "RAND36 role functioning physical RP".

VARIABLE LABELS RAND_RE "RAND36 role functioning emotional RE".

VARIABLE LABELS RAND_SF "RAND36 social functioning SF".

VARIABLE LABELS RAND_BP "RAND36 bodily pain BP".

VARIABLE LABELS RAND_MH "RAND36 mental health MH".

VARIABLE LABELS RAND_VT "RAND36 vitality VT".

VARIABLE LABELS RAND_GH "RAND36 general health GH".

execute.

** berekenen van de twee geaggregeerde scores PCS en MCS**.

*berekenen Z-scores.

COMPUTE pf_znl = (rand_pf-83.0) / 22.8 .

COMPUTE rp_znl = (rand_rp-76.4) / 36.3.

COMPUTE bp_znl = (rand_bp-74.9) / 23.4 .

COMPUTE gh_znl = (rand_gh-70.7) / 20.7 .

COMPUTE vt_znl = (rand_vt-68.6) / 19.3 .

COMPUTE sf_znl = (rand_sf-84.0) / 22.4 .

COMPUTE re_znl = (rand_re-82.3) / 32.9 .

COMPUTE mh_znl = (rand_mh-76.8) / 17.4 .

COMPUTE agg_ps = 8*sum.4((pf_znl*.42402), (rp_znl*.35119), (bp_znl*.31754), (gh_znl*.24954), (vt_znl*.02877), (sf_znl*-.00753), (re_znl*-.19206), (mh_znl*-.22069)) /Nvalid(pf_znl, rp_znl, bp_znl, gh_znl, vt_znl, sf_znl, re_znl, mh_znl).

execute.

```
COMPUTE agg_mh = 8*sum.4((pf_znl*-.22999), (rp_znl*-.12329), (bp_znl*-.09731), (gh_znl*-.01571), (vt_znl*.23534),  
(sf_znl*.26876), (re_znl*.43407), (mh_znl*.48581)) /Nvalid (pf_znl, rp_znl, bp_znl, gh_znl, vt_znl, sf_znl, re_znl, mh_znl).  
execute.
```

```
COMPUTE PCS = 50 + (agg_ps*10).  
COMPUTE MCS = 50 + (agg_mh*10).  
execute.
```

```
VARIABLE LABELS PCS "RAND 36 physical component summary score".  
VARIABLE LABELS MCS "RAND 36 mental component summary score".  
execute.
```

Bijlage 5: Argumentatie voor (non-)parametrische testen

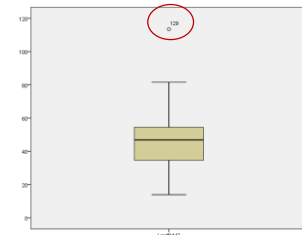
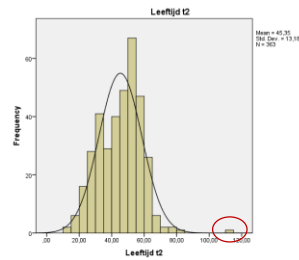
De keuze voor een parametrische of non-parametrische test wordt mede bepaald door de aard van de variabelen. Van nature zijn de variabelen in het SF/RAND-36 vragenlijst categoriaal van aard (geschikt voor een non-parametrische test), maar de geaggregeerde welzijnsvariabelen in de SF/RAND-36 vragenlijst worden in de praktijk vaak als continue variabelen aangemerkt (Bjorner et al, 2013). Dit is ook te doen gebruikelijk bij 'Patient Reported Outcome Measures' (PROM) data, zoals welzijnsmetingen (Kwaliteit van Leven) met de SF/RAND-36. De meeste PROM vragenlijsten hebben ordinale antwoordcategorieën, maar als ze samengevat worden in schalen c.q. dimensies dan worden ze veelal als continue beschouwd (geschikt voor een parametrische test). Dat komt omdat het welzijnsconcept niet als een discrete variabele kan worden aangezien: er zijn geen tussenwaarden op welzijn (bij een dobbelsteen kan je niet 4,5 gooien en dat geldt ook voor welzijn). Het welzijnsconcept loopt op of af op een glijdende schaal en daarom wordt vaak beredeneerd om een categoriale welzijnsschaal om te zetten naar een continue schaal.

Een andere motivatie is dat de rekensystematiek van de SF/RAND-36 (zie bijlage 4) de categoriale meetschalen omzet naar een schaal van 0-100, waarbij hogere scores een betere (fysieke en geestelijke) gezondheid c.q. welzijn indiceren (Ware et al., 1993; Walters, 2009). Normaal geldt dat hoe meer antwoordcategorieën er zijn, hoe beter een ordinale variabele als continue variabele kan worden aangemerkt (hoewel het arbitrair blijft, geven veel onderzoekers aan dat categoriale variabele minimaal 5 categorieën moet hebben, zie ook de adviezen van Grace-Martin (2018): <https://www.theanalysisfactor.com/can-likert-scale-data-ever-be-continuous/>). In de oorspronkelijke vragenlijst van de SF/RAND-36 vragenlijst zitten naast 5- en 6-puntschalen, ook 2-puntschalen (Rolbeperkingen n.a.v. lichamelijke problemen en Rolbeperkingen n.a.v. geestelijke problemen) en een 3-puntschalen (Lichamelijke gezondheid). Maar door alle schalen via de rekensystematiek van de SF/RAND-36 vragenlijst om te zetten naar een schaal van 1-100, ontstaat er in dit onderzoek bij Rolbeperkingen n.a.v. geestelijke problemen een schaal met vier categorieën, op Rolbeperkingen n.a.v. lichamelijke problemen een schaal met vijf categorieën en op Lichamelijke gezondheid een schaal met meer dan tien categorieën. In de regel geldt: hoe meer categorieën in de antwoordschaal, hoe meer categorieën er door de rekensystematiek van de SF/RAND-36 vragenlijst op een schaal van 1-100 er zullen optreden (mits er in de steekproef verschillend geantwoord wordt). Dat betekent ook dat de schaal steeds meer neigt naar een continue variabele, temeer als de respondenten in de steekproef variëren in hun scores van 1-100.

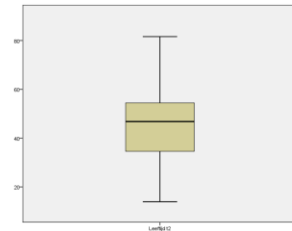
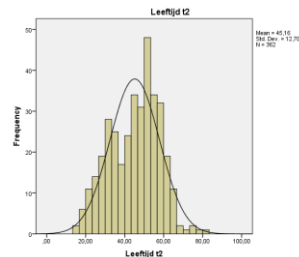
Niettemin blijft er onder statistici en onderzoekers een discussie aan de gang in hoeverre categoriale variabelen in (welzijns)onderzoek als continue variabelen mogen worden aangezien (zie ook Knap, 1990). Daarin is een kamp met voorstanders en een kamp met tegenstanders. Omdat tegenstanders gefundeerde argumentatie hebben (zie ook: Kahler et al., 2008; Orwelius et al., 2018), is in dit onderzoek ook een non-parametrische test uitgevoerd (zie ook het advies van Grace-Martin in de eerder gegeven link). Grace-Martin (2018) geeft aan dat wanneer de non-parametrische test hetzelfde resultaat vertoont als de parametrische test, de conclusies met grotere zekerheid kunnen worden aangenomen. Dit advies is in dit onderzoek dan ook gevolgd.

Daarnaast moeten de scores op de variabelen in het onderzoek aan bepaalde statistische voorwaarden voldoen om een parametrische test (in dit geval een 'Repeated Measures' (RM) test) te mogen uitvoeren. Daartoe zijn eerst 'outliers' c.q. uitzonderlijke gevallen verwijderd op basis van z-scores, histogrammen en boxplots (zie het voorbeeld op de volgende pagina: op de eerste rij tabellen en grafieken indiceren de rode cirkels een outlier en in de tweede rij zijn deze weggehaald).

Zscore: Leeftijd t2				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Extreme (z-score > 3.29)	1	.3	.3	.3
Probable outliers (z > 2.58)	2	.5	.6	.8
Potential outliers (z > 1.96)	3	.8	.8	1.7
Normal range	357	97.8	98.3	100.0
Total	363	99.5	100.0	
Missing	2	.5		
System				
Total	365	100.0		



Zscore: Leeftijd t2				
	Frequency	Percent	Valid Percent	Cumulative Percent
Valid				
Probable outliers (z > 2.58)	2	.5	.6	.6
Potential outliers (z > 1.96)	3	.8	.8	1.4
Normal range	357	97.8	98.6	100.0
Total	362	99.2	100.0	
Missing	3	.8		
System				
Total	365	100.0		



Bij het analyseren van de z-scores wordt verondersteld dat in een normale verdeling circa 5% van de z-scores groter is dan 1.96, circa 1% groter is dan 2.58 hebben en dat (vrijwel) geen score groter is dan 3.29 (Field, 2018). Outliers die buiten deze verdeling vielen, zijn – in combinatie met het vinden van outliers in boxplots – verwijderd. Dit heeft geresulteerd in het verwijderen van 15 van de 365 respondenten in de steekproef.

Vervolgens is de normaalverdeling van elke dimensie in de steekproef beoordeeld. De steekproefomvang is dusdanig groot dat de 'Central Limit Theorem' hier geldt: als de sample grootte toeneemt, dan zal de geobserveerde verdeling meer gaan naar een normale verdeling. Dit geldt voor steekproeven groter dan 100 (waar geen K-S test op wordt toegepast). Er is hier sprake van een grote steekproef ($n = 350$), waardoor de normaal verdeling naar alle waarschijnlijkheid meer overeenkomt met de doelpopulatie (mensen met lichamelijke en/of geestelijke klachten). De steekproefomvang heeft het ook mogelijk gemaakt om op basis van histogrammen en QQ-plots de normaal verdeling te evalueren. Daarnaast is de normaal verdeling beoordeeld aan de hand van Skewness en Kurtosis scores. Hierbij is als uitgangspunt genomen dat de waarden voor Skewness en Kurtosis tussen -2 en +2 als acceptabel worden beschouwd om een normale verdeling aan te tonen (Geoghegan & Mallery, 2010; Field, 2018). Uit de analyse komt naar voren dat vrijwel alle dimensies normaal verdeeld zijn, met een lichte uitzonderling van de dimensie Lichamelijk Gezondheid. Daar scoorde de Kurtosis meestal boven de 2, al kwam deze bij elke meting niet of nauwelijks boven de 3 uit. Dit komt doordat de scores op deze variabele relatief hoog waren (zowel op de voor- als de nametingen).

Voorts is de onafhankelijkheid tussen de tijdsmetingen in ogenschouw genomen (de residuen tussen de metingen moeten niet gecorreleerd zijn). Dit is beoordeeld met de Durbin Watson test. Deze test varieert tussen 0 en 4, waarbij een waarde van (rond de) 2 aangeeft dat de residuen niet met elkaar correleren. Field (2018) geeft aan dat de waarden in de Durbin Watson test tussen de 1 en 3 moeten blijven. Op alle combinaties van tijdsmetingen (waarbij een latere tijdsmeting als afhankelijke variabele en eerdere tijdsmetingen als onafhankelijke variabelen zijn beschouwd) en voor alle gezondheidsdimensies is deze test uitgevoerd. Uit de resultaten komt naar voren dat alle scores rond de 2 uitkomen (en niet onder de 1 of boven de 3). Dit betekent dat voldaan is aan de voorwaarde van onafhankelijkheid tussen de metingen.

Uiteindelijk is de assumptie van sphericiteit beoordeeld. Deze veronderstelling geeft aan dat de variantie in de verschil-scores tussen de tijdsmetingen gelijk is. Variantie wil zeggen een spreiding in een reeks van scores (hoe groter de variantie, hoe meer de scores onderling verschillen en dus ook hoe meer deze van het gemiddelde afwijken). Voordat de resultaten geïnterpreteerd kunnen worden,

is het van belang om eerst te kijken of er is voldaan aan de aanname van sphericiteit. Dit gebeurt in SPSS aan de hand van 'Mauchly's Test of Sphericity'. Indien de p-waarde onder de 5% zit ($p < .05$) is de assumptie van sphericiteit geschonden. In dit onderzoek is dat voor alle SF/RAND-36 variabelen het geval (de variantie in de verschil-scores tussen de condities is niet gelijk). Ondanks dat niet aan deze voorwaarde is voldaan, kan binnen de RM procedure toch het verschil in tijdsmetingen op de gezondheidsdimensies beoordeeld worden (zie bijlage 10).

Bijlage 6: Effect size formules

De sterkte van het effect (Effect size) is berekend aan de hand van de volgende formules:

Effect size non-parametrisch (Wilcoxon test):

$$r_{t_n - t_{...n}} = \frac{Z}{\sqrt{n \times 2}}$$

Effect size parametrisch (RM test):

- Omgega square:

$$\omega^2 = \frac{\left[\frac{k-1}{nk} (MS_M - MS_R) \right]}{MS_R + \frac{MS_B - MS_R}{k} + \left[\frac{k-1}{nk} (MS_M - MS_R) \right]}$$

$$MS_B = \frac{SS_B}{df_B} = \frac{SS_B}{N-1}$$

$$SS_B = SS_T - SS_M - SS_R$$

$$SS_T = S_{grand}^2 (N - 1)$$

- Equation to convert F-values (Field, 2018):

$$r = \sqrt{\frac{F(1, df_R)}{F(1, df_R) + df_R}}$$

- Partial Eta Squared

Effect size	Use	Small	Medium	Large
r	Friedman > Wilcoxon	0.1	0.3	0.5
Omega-squared	Repeated measures (sample)	0.01	0.06	0.14
r	Repeated measures (levels)	0.1	0.3	0.5
Partial Eta squared	Repeated measures (multiv.)	0.01	0.06	0.14

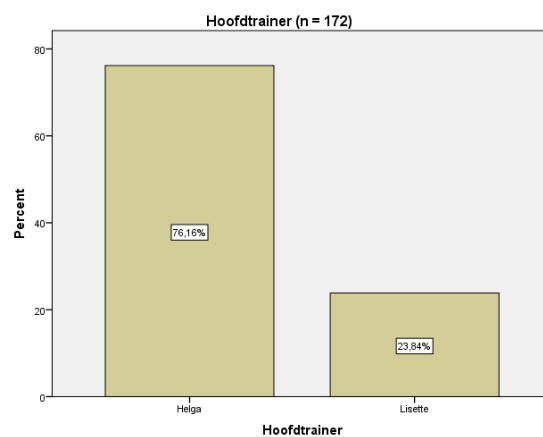
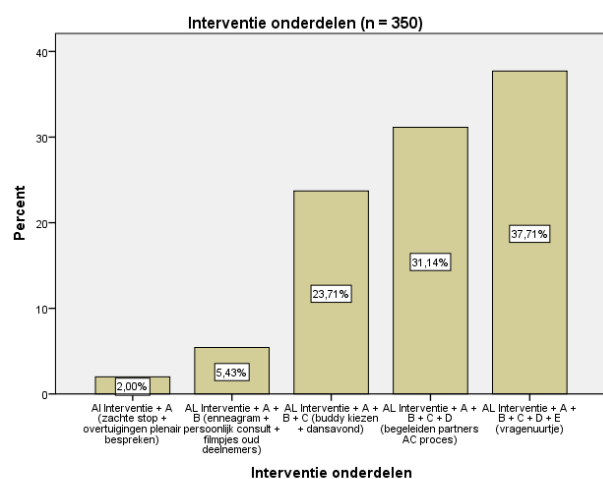
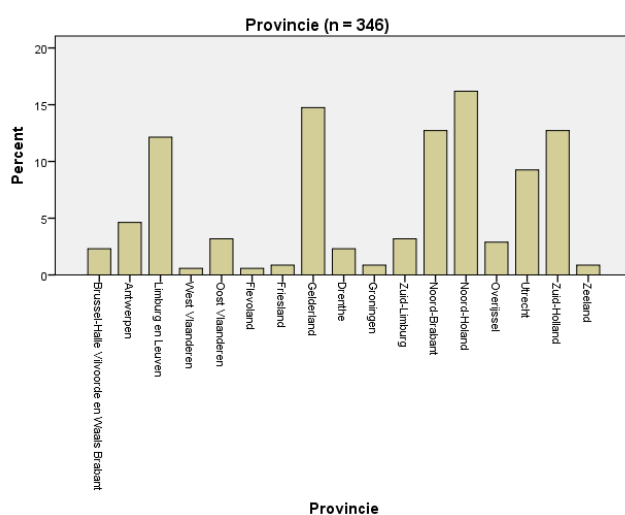
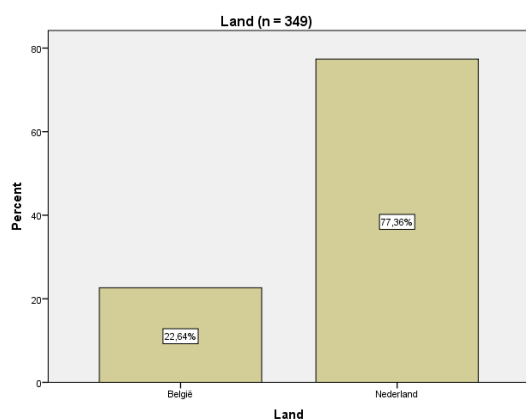
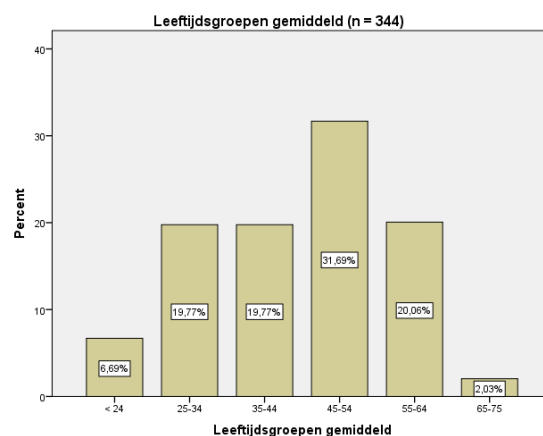
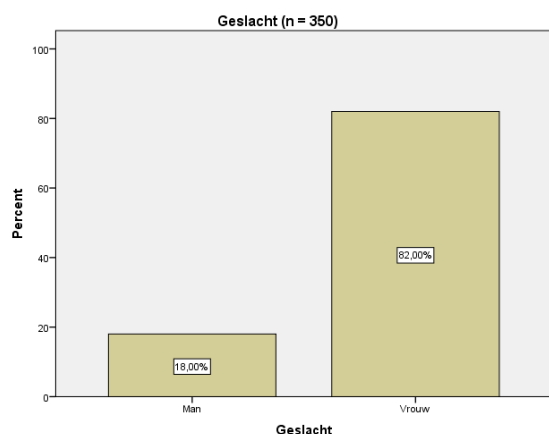
Bijlage 7: Steekproefomvang i.r.t. effectmeting

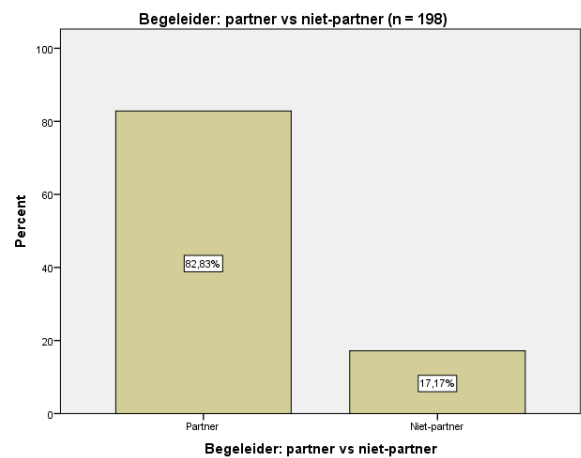
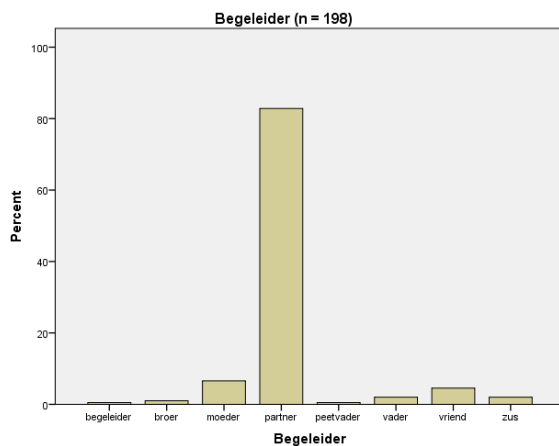
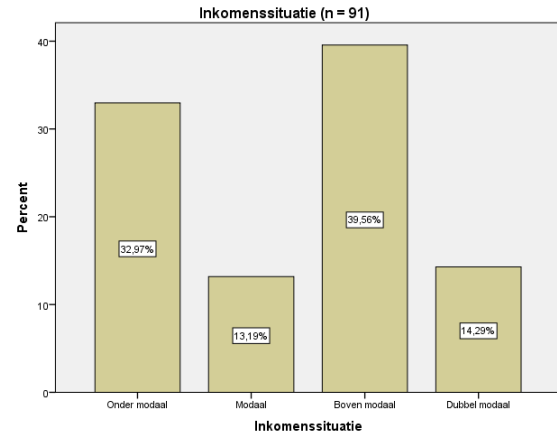
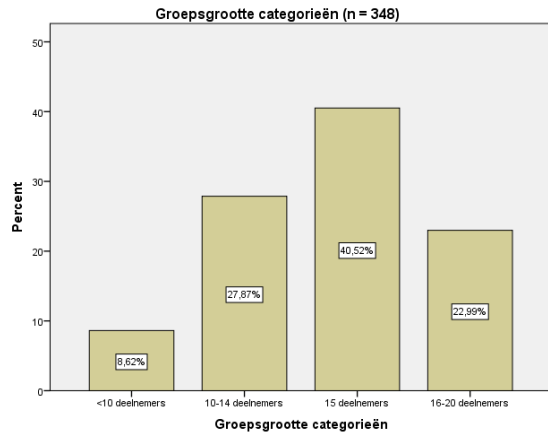
Ware et al. (1993) geven aan hoe groot de steekproefomvang moet zijn om via de SF/RAND-36 vragenlijst een effect te kunnen vaststellen. Dit wordt in onderstaand schema weergegeven (in Engels). De steekproefomvang in dit onderzoek omvat 350 mensen. Dus een verschil van 5 punten is in dit onderzoek al voldoende om een effect te detecteren.

		Number of Points of difference			
Scale		2	5	10	20
Physical functioning	PF	852	137	35	9
Role - Physical	RP	1826	293	74	19
Bodily Pain	BP	883	142	36	10
General Health	GH	654	105	27	7
Vitality	VT	693	111	28	8
Social Functioning	SF	809	130	33	9
Role-Emotional	RE	1721	276	69	18
Mental Health	MH	515	83	21	6

Note. Estimates assume $\alpha = 0.05$, two-tailed t-test, power = 80% and an intertemporal correlation between scores of 0.60 (Cohen, 1988).

Bijlage 8: Descriptieve statistiek





Statistics

	RAND36 algemene gezondheid t0	RAND36 gezondheids- transformatie t0	RAND36 lichamelijke gezondheid t0	RAND36 rol- beperkingen n.a.v. lichamelijke problemen t0	RAND36 lichaamspijn t0	RAND36 geestelijke gezondheid t0	RAND36 rol- beperkingen n.a.v. geestelijke problemen t0	RAND36 vitaliteit t0	RAND36 sociaal functioneren t0	RAND 36 lichamelijk welzijn t0	RAND 36 geestelijk welzijn t0
N Valid	350	350	350	350	350	350	350	350	350	350	350
Missing	0	0	0	0	0	0	0	0	0	0	0
Mean	44,2143	47,8571	70,2571	25,5000	62,1574	52,5257	46,1905	34,9571	44,2857	42,6025	33,3362
Std. Deviation	18,86150	24,79965	22,98479	37,72877	25,63099	16,88383	41,49884	16,70787	25,48647	11,06686	11,01947
Skewness	,425	,013	-,557	1,138	-,324	-,105	,160	,233	,079	,118	,172
Kurtosis	-,262	-,400	-,584	-,355	-,553	,188	-,1602	-,336	-,679	-,741	-,515
Range	95,00	100,00	95,00	100,00	100,00	96,00	100,00	80,00	100,00	53,75	57,22
Minimum	,00	,00	5,00	,00	,00	,00	,00	,00	,00	14,98	8,17
Maximum	95,00	100,00	100,00	100,00	100,00	96,00	100,00	80,00	100,00	68,73	65,39

Statistics

	RAND36 algemene gezondheid t1	RAND36 gezondheids- transformatie t1	RAND36 lichamelijke gezondheid t1	RAND36 rol- beperkingen n.a.v. lichamelijke problemen t1	RAND36 lichaamspijn t1	RAND36 geestelijke gezondheid t1	RAND36 rol- beperkingen n.a.v. geestelijke problemen t1	RAND36 vitaliteit t1	RAND36 sociaal functioneren t1	RAND 36 lichamelijk welzijn t1	RAND 36 geestelijk welzijn t1
N Valid	350	350	350	350	350	350	350	350	350	350	350
Missing	0	0	0	0	0	0	0	0	0	0	0
Mean	75,9571	82,5000	88,2143	76,7857	79,4402	76,9371	85,2381	71,9857	82,1071	52,1248	49,8439
Std. Deviation	17,35392	21,80517	13,17818	34,46309	18,57187	14,18791	28,34483	15,78826	18,99440	7,39507	8,66494
Skewness	-,855	-1,300	-1,476	-1,198	-,851	-,635	-1,937	-,529	-1,048	-,911	-1,260
Kurtosis	,193	1,598	2,135	-,026	,426	,323	2,719	,230	1,049	,341	2,365
Range	75,00	100,00	70,00	100,00	87,76	72,00	100,00	85,00	87,50	39,82	58,22
Minimum	25,00	,00	30,00	,00	12,24	28,00	,00	15,00	12,50	27,94	11,58
Maximum	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	67,76	69,80

Statistics

	RAND36 algemene gezondheid t2	RAND36 gezondheids- transformatie t2	RAND36 lichamelijke gezondheid t2	RAND36 rol- beperkingen n.a.v. lichamelijke problemen t2	RAND36 lichaamspijn t2	RAND36 geestelijke gezondheid t2	RAND36 rol- beperkingen n.a.v. geestelijke problemen t2	RAND36 vitaliteit t2	RAND36 sociaal functioneren t2	RAND 36 lichamelijk welzijn t2	RAND 36 geestelijk welzijn t2
N	Valid	350	350	350	350	350	350	350	350	350	350
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		75,5286	85,0000	89,3143	73,5714	80,5831	75,8857	82,2857	69,0857	80,5714	52,3895
Std. Deviation		18,86289	20,11785	13,02251	38,15350	19,52948	15,69584	31,47414	18,46218	21,22743	7,71494
Skewness		-,916	-,1613	-,1516	-,1051	-,896	-,852	-,1634	-,667	-,1071	-,806
Kurtosis		,548	3,193	2,212	-,551	,067	,797	1,345	,281	,840	-,025
Range		95,00	100,00	65,00	100,00	79,59	80,00	100,00	90,00	100,00	40,11
Minimum		5,00	,00	35,00	,00	20,41	20,00	,00	10,00	,00	26,24
Maximum		100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	66,35	63,77

Statistics

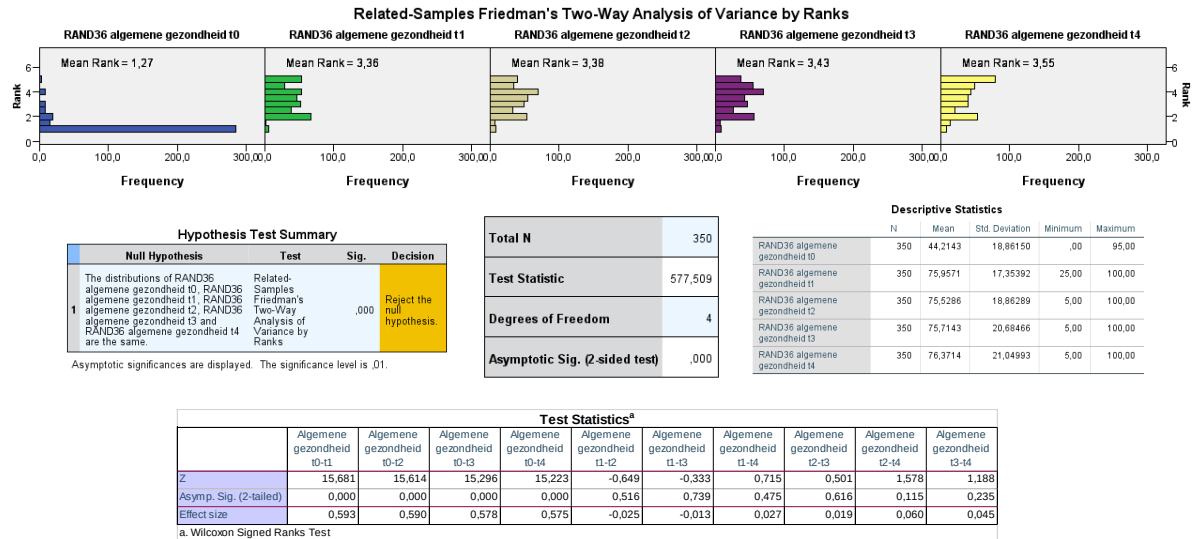
	RAND36 algemene gezondheid t3	RAND36 gezondheids- transformatie t3	RAND36 lichamelijke gezondheid t3	RAND36 rol- beperkingen n.a.v. lichamelijke problemen t3	RAND36 lichaamspijn t3	RAND36 geestelijke gezondheid t3	RAND36 rol- beperkingen n.a.v. geestelijke problemen t3	RAND36 vitaliteit t3	RAND36 sociaal functioneren t3	RAND 36 lichamelijk welzijn t3	RAND 36 geestelijk welzijn t3
N	Valid	350	350	350	350	350	350	350	350	350	350
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		75,7143	85,4286	89,7857	73,8571	80,4490	75,7143	81,8095	69,0571	81,2500	52,5558
Std. Deviation		20,68466	19,98823	13,17275	38,06919	20,47392	16,11024	32,35275	19,35484	21,51060	7,98657
Skewness		-,789	-,1266	-,1662	-,1063	-,961	-,890	-,1620	-,719	-,1138	-,975
Kurtosis		-,166	,878	2,529	-,529	,281	,614	1,234	,195	1,014	,194
Range		95,00	75,00	65,00	100,00	89,80	76,00	100,00	95,00	100,00	38,41
Minimum		5,00	25,00	35,00	,00	10,20	24,00	,00	5,00	,00	28,46
Maximum		100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	66,86	64,62

Statistics

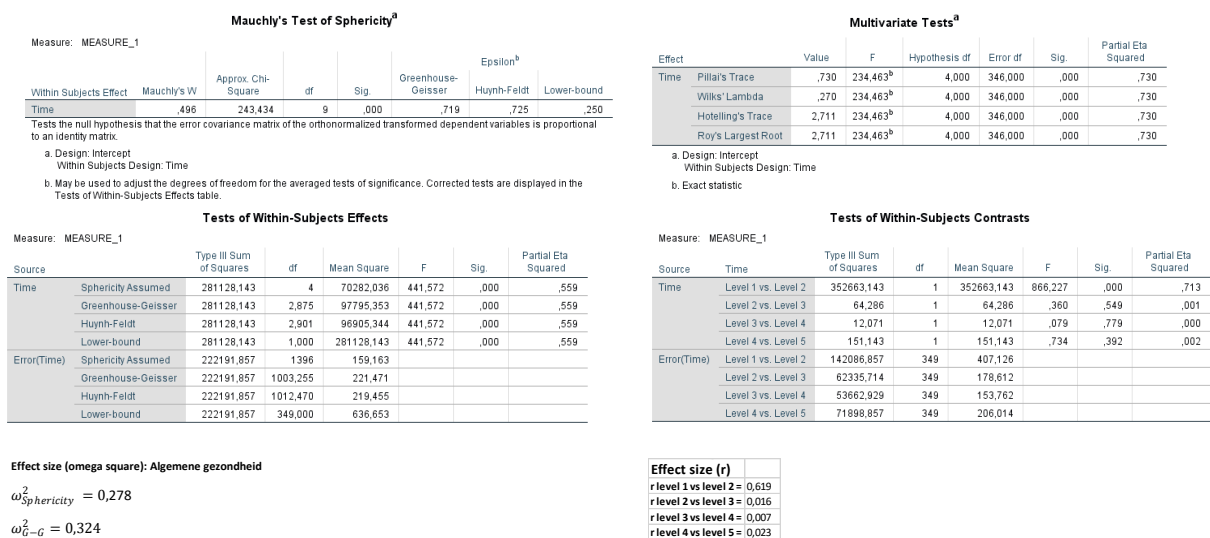
	RAND36 algemene gezondheid t4	RAND36 gezondheids- transformatie t4	RAND36 lichamelijke gezondheid t4	RAND36 rol- beperkingen n.a.v. lichamelijke problemen t4	RAND36 lichaamspijn t4	RAND36 geestelijke gezondheid t4	RAND36 rol- beperkingen n.a.v. geestelijke problemen t4	RAND36 vitaliteit t4	RAND36 sociaal functioneren t4	RAND 36 lichamelijk welzijn t4	RAND 36 geestelijk welzijn t4
N	Valid	350	350	350	350	350	350	350	350	350	350
	Missing	0	0	0	0	0	0	0	0	0	0
Mean		76,3714	84,2857	90,0143	72,2143	81,2653	77,6571	85,2381	69,0571	82,1071	52,1799
Std. Deviation		21,04993	22,13992	13,26660	38,80038	19,69264	15,87765	29,76920	19,69240	21,32618	8,27636
Skewness		-,949	-,1398	-,1788	-,974	-,980	-,961	-,1945	-,700	-,1077	-,1084
Kurtosis		,274	1,359	3,072	-,715	,371	,785	2,502	-,062	,366	,424
Range		95,00	100,00	65,00	100,00	87,76	80,00	100,00	85,00	87,50	39,53
Minimum		5,00	,00	35,00	,00	12,24	20,00	,00	15,00	12,50	25,65
Maximum		100,00	100,00	100,00	100,00	100,00	100,00	100,00	100,00	65,18	64,03

Bijlage 9: Grafieken en tabellen onderzoeksresultaten

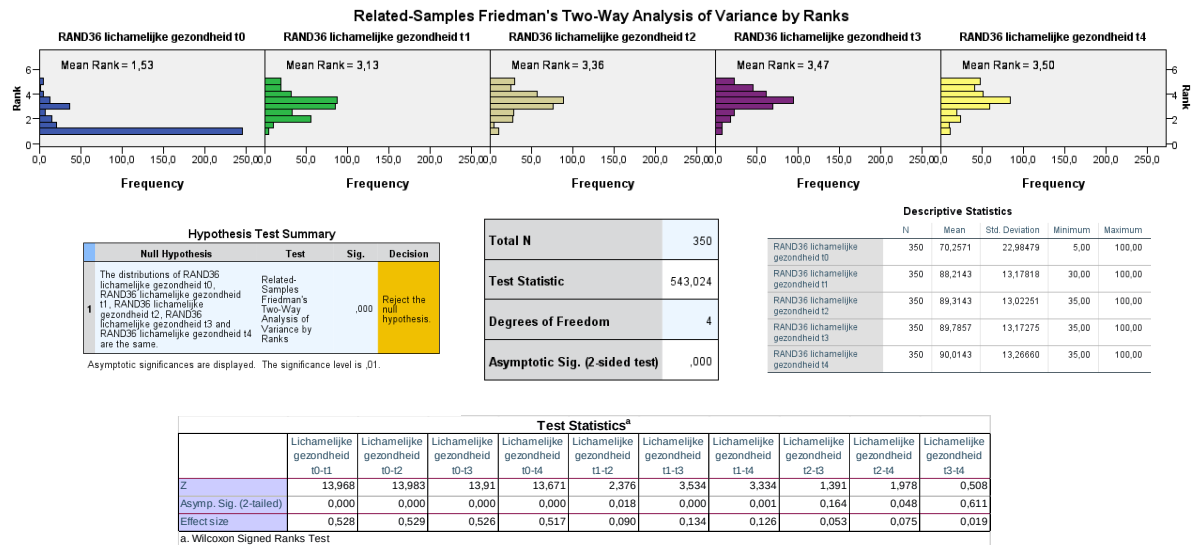
Algemene gezondheid: non-parametrisch



Algemene gezondheid: parametrisch



Lichamelijke gezondheid: non-parametrisch



Lichamelijke gezondheid: parametrisch

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Epsilon ^b	Lower-bound
Time	,139	685,705	9	,000	,465	,468	,250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	
Time	Sphericity Assumed	102551,914	4	25637,979	248,999	,000	,416
	Greenhouse-Geisser	102551,914	1,861	55117,914	248,999	,000	,416
	Huynh-Feldt	102551,914	1,870	54835,842	248,999	,000	,416
	Lower-bound	102551,914	1,000	102551,914	248,999	,000	,416
Error(Time)	Sphericity Assumed	143738,086	1396	102,964			
	Greenhouse-Geisser	143738,086	649,346	221,358			
	Huynh-Feldt	143738,086	652,687	220,225			
	Lower-bound	143738,086	349,000	411,857			

Effect size (omega square): Lichamelijke gezondheid

$$\omega_{Sphericity}^2 = 0,207$$

$$\omega_{G-G}^2 = 0,283$$

Multivariate Tests^a

Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	,502	87,165 ^b	4,000	346,000	,000	,502
	Wilks' Lambda	,498	87,165 ^b	4,000	346,000	,000	,502
	Hotelling's Trace	1,008	87,165 ^b	4,000	346,000	,000	,502
	Roy's Largest Root	1,008	87,165 ^b	4,000	346,000	,000	,502

a. Design: Intercept
Within Subjects Design: Time

b. Exact statistic

Tests of Within-Subjects Contrasts

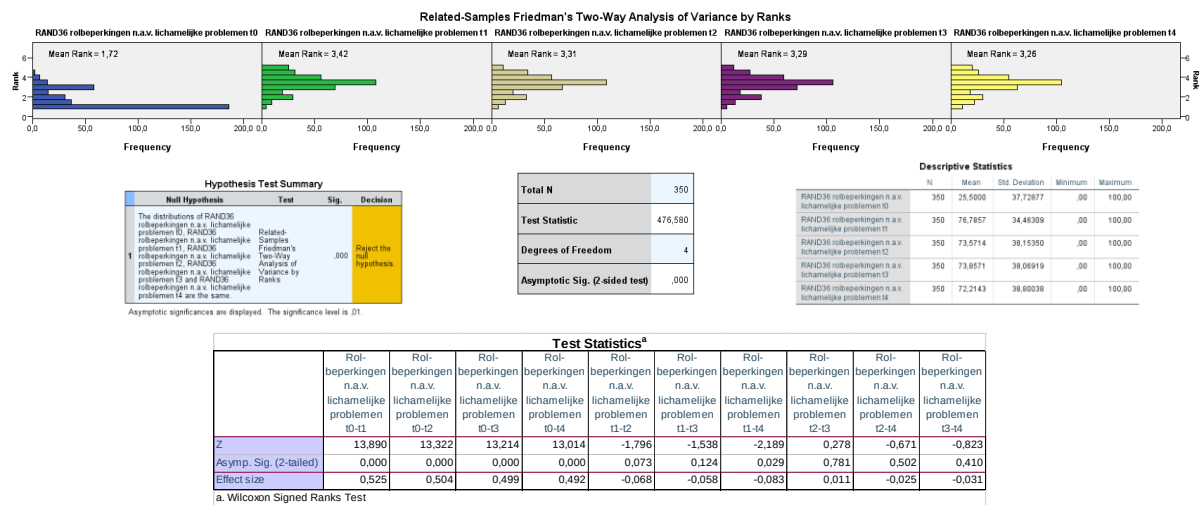
Measure: MEASURE_1

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Level 1 vs. Level 2	112860,643	1	112860,643	323,215	,000	,481
	Level 2 vs. Level 3	423,500	1	423,500	6,412	,012	,018
	Level 3 vs. Level 4	77,786	1	77,786	1,331	,249	,004
	Level 4 vs. Level 5	18,286	1	18,286	,241	,624	,001
Error(Time)	Level 1 vs. Level 2	121864,357	349	349,182			
	Level 2 vs. Level 3	23051,500	349	66,050			
	Level 3 vs. Level 4	20397,214	349	58,445			
	Level 4 vs. Level 5	26531,714	349	76,022			

Effect size (r)

r level 1 vs level 2	= 0,434
r level 2 vs level 3	= 0,068
r level 3 vs level 4	= 0,031
r level 4 vs level 5	= 0,013

Rolbeperkingen n.a.v. lichamelijke problemen: non-parametrisch



Rolbepedingen n.a.v. lichamelijke problemen: parametrisch

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

		Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Epsilon ^b	Lower-bound
Within Subjects Effect	Mauchly's W						
Time		.732	108.563	.000	.854	.863	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	665430.714	4	166357.679	224.856	.000	.392
	Greenhouse-Geisser	665430.714	3.415	194834.813	224.856	.000	.392
	Huynh-Feldt	665430.714	3.453	192699.319	224.856	.000	.392
	Lower-bound	665430.714	1.000	665430.714	224.856	.000	.392
Error(Time)	Sphericity Assumed	1032819.286	1396	739.842			
	Greenhouse-Geisser	1032819.286	1191.960	866.488			
	Huynh-Feldt	1032819.286	1205.169	856.991			
	Lower-bound	1032819.286	349.000	2959.368			

Effect size (omega square): Rolbepedingen n.a.v. lichamelijke problemen

$$\omega_{Sphericity}^2 = 0,195$$

$$\omega_{G-G}^2 = 0,210$$

Multivariate Tests

	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Pillai's trace	,622	142,574 ^a	4,000	346,000	,000	,622
Wilks' lambda	,378	142,574 ^a	4,000	346,000	,000	,622
Hotelling's trace	1,648	142,574 ^a	4,000	346,000	,000	,622
Roy's largest root	1,648	142,574 ^a	4,000	346,000	,000	,622

Each F tests the multivariate effect of Time. These tests are based on the linearly independent pairwise comparisons among the estimated marginal means.

a. Exact statistic

Tests of Within-Subjects Contrasts

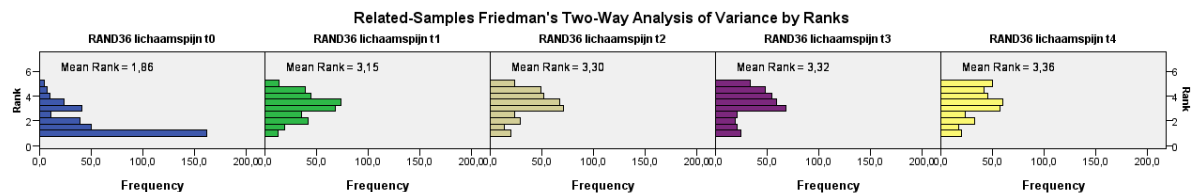
Measure: MEASURE_1

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Level 1 vs. Level 2	920578,571	1	920578,571	525,038	,000	,601
	Level 2 vs. Level 3	3616,071	1	3616,071	3,677	,056	,010
	Level 3 vs. Level 4	28,571	1	28,571	,028	,867	,000
	Level 4 vs. Level 5	944,643	1	944,643	,838	,361	,002
Error(Time)	Level 1 vs. Level 2	611921,429	349	1753,357			
	Level 2 vs. Level 3	343258,929	349	983,550			
	Level 3 vs. Level 4	352471,429	349	1009,947			
	Level 4 vs. Level 5	393430,357	349	1127,308			

Effect size (r)

r level 1 vs level 2 =	0,523
r level 2 vs level 3 =	0,051
r level 3 vs level 4 =	0,004
r level 4 vs level 5 =	0,024

Lichaamspijn: non-parametrisch



Hypothesis Test Summary			
	Null Hypothesis	Test	Sig.
1	The distributions of RAND36 lichaamspijn t0, RAND36 lichaamspijn t1, RAND36 lichaamspijn t2, RAND36 lichaamspijn t3 and RAND36 lichaamspijn t4 are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	,000

Asymptotic significances are displayed. The significance level is .01.

Total N	350
Test Statistic	307,948
Degrees of Freedom	4
Asymptotic Sig. (2-sided test)	,000

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
RAND36 lichaamspijn t0	350	62,1574	25,63099	,00	100,00
RAND36 lichaamspijn t1	350	79,4402	18,57187	12,24	100,00
RAND36 lichaamspijn t2	350	80,5831	19,52948	20,41	100,00
RAND36 lichaamspijn t3	350	80,4490	20,47392	10,20	100,00
RAND36 lichaamspijn t4	350	81,2653	19,69264	12,24	100,00

Test Statistics ^a										
	Lichaams-pijn t0-t1	Lichaams-pijn t0-t2	Lichaams-pijn t0-t3	Lichaams-pijn t0-t4	Lichaams-pijn t1-t2	Lichaams-pijn t1-t3	Lichaams-pijn t1-t4	Lichaams-pijn t2-t3	Lichaams-pijn t2-t4	Lichaams-pijn t3-t4
Z	11,914	12,116	11,396	11,847	1,37	1,207	1,909	-0,168	0,669	0,737
Asymp. Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,171	0,227	0,056	0,867	0,503	0,461
Effect size	0,450	0,458	0,431	0,448	0,052	0,046	0,072	-0,006	0,025	0,028

a. Wilcoxon Signed Ranks Test

Lichaamspijn: parametrisch

Mauchly's Test of Sphericity ^a						
Measure: MEASURE_1						
	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Epsilon ^b
Within Subjects Effect						
Time	,656	146,231	9	,000	,799	,807

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects						
Measure: MEASURE_1						
Source		Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Sphericity Assumed	94128,699	4	23532,175	117,553	,000
	Greenhouse-Geisser	94128,699	3,195	29461,868	117,553	,000
	Huynh-Feldt	94128,699	3,228	29160,908	117,553	,000
	Lower-bound	94128,699	1,000	94128,699	117,553	,000
	Upper-bound	94128,699	1,000	94128,699	117,553	,000
Error(Time)	Sphericity Assumed	279457,307	1396	200,184		
	Greenhouse-Geisser	279457,307	1115,032	250,627		
	Huynh-Feldt	279457,307	1126,540	248,067		
	Lower-bound	279457,307	349,000	800,737		
	Upper-bound	279457,307	349,000	800,737		

Effect size (omega square): Lichaamspijn

$$\omega_{\text{sphericity}}^2 = 0,125$$

$$\omega_{\text{G}}^2 = 0,139$$

Multivariate Tests ^a						
Effect						
		Value	F	Hypothesis df	Error df	Sig.
Time	Pillai's Trace	,425	64,012 ^b	4,000	346,000	,000
	Wilks' Lambda	,575	64,012 ^b	4,000	346,000	,000
	Hotelling's Trace	,740	64,012 ^b	4,000	346,000	,000
	Roy's Largest Root	,740	64,012 ^b	4,000	346,000	,000

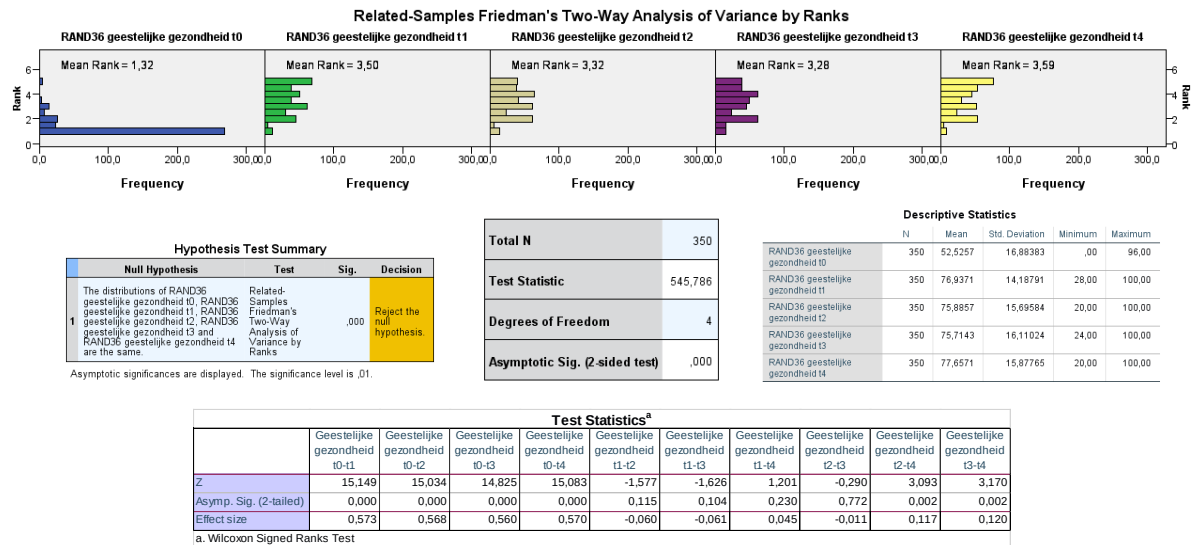
a. Design: Intercept
Within Subjects Design: Time

b. Exact statistic

Tests of Within-Subjects Contrasts						
Measure: MEASURE_1						
Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.
Time	Level 1 vs. Level 2	104543,297	1	104543,297	204,777	,000
	Level 2 vs. Level 3	457,143	1	457,143	1,885	,171
	Level 3 vs. Level 4	6,295	1	6,295	,022	,882
	Level 4 vs. Level 5	233,236	1	233,236	,783	,377
	Level 5 vs. Level 6	233,236	1	233,236	,783	,377
Error(Time)	Level 1 vs. Level 2	178172,238	349	510,522		
	Level 2 vs. Level 3	84615,743	349	242,452		
	Level 3 vs. Level 4	99473,089	349	285,023		
	Level 4 vs. Level 5	103965,015	349	297,894		
	Level 5 vs. Level 6	103965,015	349	297,894		

Effect size (r)	
r level 1 vs level 2 =	0,358
r level 2 vs level 3 =	0,037
r level 3 vs level 4 =	0,013
r level 4 vs level 5 =	0,024

Geestelijke gezondheid: non-parametrisch



Geestelijke gezondheid: parametrisch

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Epsilon ^b	Lower-bound
Time	,589	184,154	9	,000	,757	,765	,250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time						
Sphericity Assumed	162467,712	4	40616,928	338,482	,000	,492
Greenhouse-Geisser	162467,712	3,029	53635,369	338,482	,000	,492
Huynh-Feldt	162467,712	3,059	53118,137	338,482	,000	,492
Lower-bound	162467,712	1,000	162467,712	338,482	,000	,492
Error(Time)						
Sphericity Assumed	167516,288	1396	119,997			
Greenhouse-Geisser	167516,288	1057,161	158,459			
Huynh-Feldt	167516,288	1067,455	156,931			
Lower-bound	167516,288	349,000	479,989			

Effect size (omega square): Geestelijke gezondheid

$$\omega_{Sphericity}^2 = 0,245$$

$$\omega_{G-G}^2 = 0,279$$

Multivariate Tests^a

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time						
Pillai's Trace	,671	176,715 ^b	4,000	346,000	,000	,671
Wilks' Lambda	,329	176,715 ^b	4,000	346,000	,000	,671
Hotelling's Trace	2,043	176,715 ^b	4,000	346,000	,000	,671
Roy's Largest Root	2,043	176,715 ^b	4,000	346,000	,000	,671

a. Design: Intercept
Within Subjects Design: Time

b. Exact statistic

Tests of Within-Subjects Contrasts

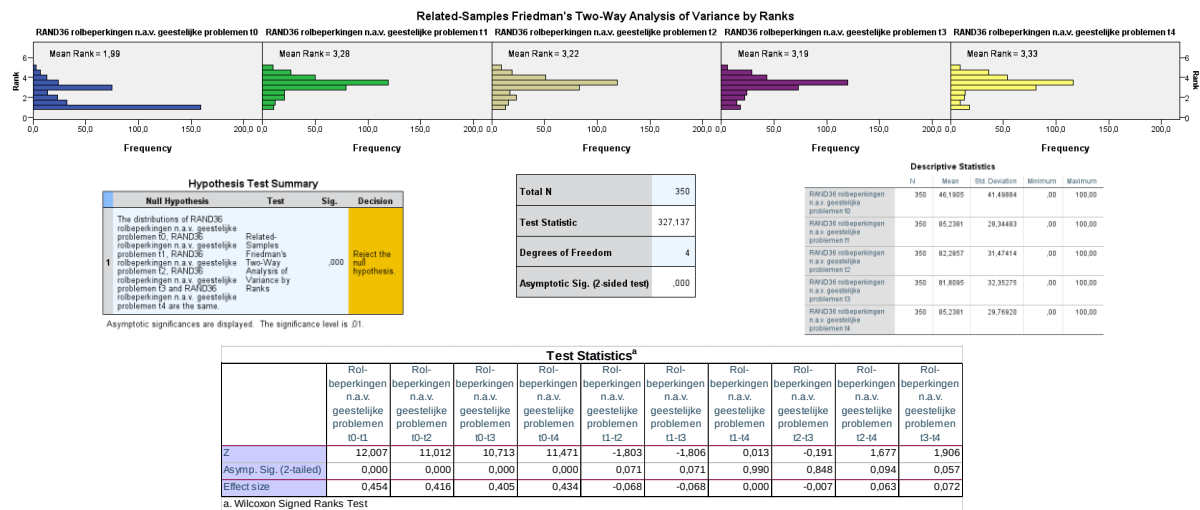
Measure: MEASURE_1

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time							
Level 1 vs. Level 2		208571,246	1	208571,246	631,142	,000	,644
Level 2 vs. Level 3		386,926	1	386,926	2,615	,107	,007
Level 3 vs. Level 4		10,286	1	10,286	,069	,793	,000
Level 4 vs. Level 5		1321,143	1	1321,143	8,078	,005	,023
Error(Time)							
Level 1 vs. Level 2		115332,754	349	330,466			
Level 2 vs. Level 3		51645,074	349	147,980			
Level 3 vs. Level 4		52005,714	349	149,014			
Level 4 vs. Level 5		57078,857	349	163,550			

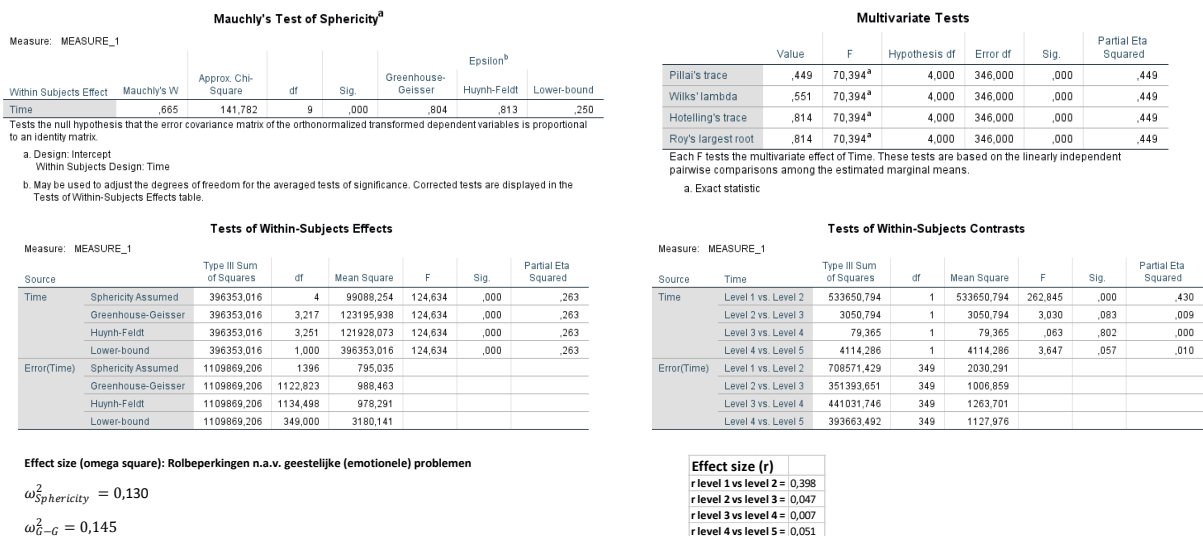
Effect size (r)

r level 1 vs level 2 = 0,558
r level 2 vs level 3 = 0,043
r level 3 vs level 4 = 0,007
r level 4 vs level 5 = 0,076

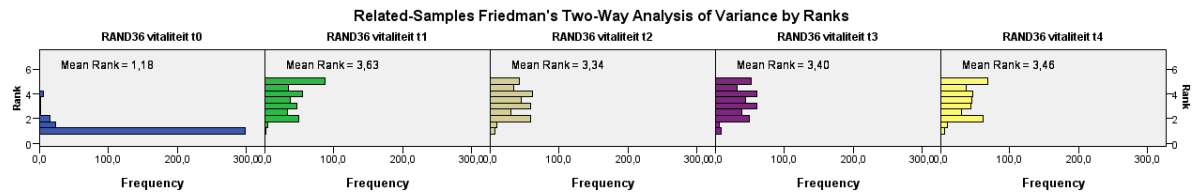
Rolbeperkingen n.a.v. geestelijke problemen: non-parametrisch



Rolbeperkingen n.a.v. geestelijke problemen: parametrisch



Vitaliteit (energie): non-parametrisch



Hypothesis Test Summary			
Null Hypothesis	Test	Sig.	Decision
1 The distributions of RAND36 vitaliteit t0, RAND36 vitaliteit t1, RAND36 vitaliteit t2, RAND36 vitaliteit t3 and RAND36 vitaliteit t4 are the same.	Related-Samples Friedman's Two-Way Analysis of Variance by Ranks	,000	Reject the null hypothesis

Asymptotic significances are displayed. The significance level is .01.

Total N	350
Test Statistic	631,896
Degrees of Freedom	4
Asymptotic Sig. (2-sided test)	,000

Descriptive Statistics					
	N	Mean	Std. Deviation	Minimum	Maximum
RAND36 vitaliteit t0	350	34,9571	16,70787	,00	80,00
RAND36 vitaliteit t1	350	71,9857	15,78826	15,00	100,00
RAND36 vitaliteit t2	350	69,0857	18,46218	10,00	100,00
RAND36 vitaliteit t3	350	69,0571	19,35484	5,00	100,00
RAND36 vitaliteit t4	350	69,0571	19,69240	15,00	100,00

Test Statistics ^a										
	Vitaliteit t0-t1	Vitaliteit t0-t2	Vitaliteit t0-t3	Vitaliteit t0-t4	Vitaliteit t1-t2	Vitaliteit t1-t3	Vitaliteit t1-t4	Vitaliteit t2-t3	Vitaliteit t2-t4	Vitaliteit t3-t4
Z	16,054	15,7	15,721	15,723	-3,9	-3,132	-2,622	0,466	0,371	0,17
Asymp. Sig. (2-tailed)	0,000	0,000	0,000	0,000	0,000	0,002	0,009	0,642	0,711	0,865
Effect size	0,607	0,593	0,594	0,594	-0,147	-0,118	-0,099	0,018	0,014	0,006

a. Wilcoxon Signed Ranks Test

Vitaliteit (energie): parametrisch

Mauchly's Test of Sphericity ^a							
Measure	MEASURE_1						
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Greenhouse-Geisser	Epsilon ^b	Lower-bound
Time	,554	205,079	9	,000	,752	,759	,250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Multivariate Tests ^a							
Effect		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	,795	335,453 ^b	4,000	346,000	,000	,795
	Wilks' Lambda	,205	335,453 ^b	4,000	346,000	,000	,795
	Hotelling's Trace	3,878	335,453 ^b	4,000	346,000	,000	,795
	Roy's Largest Root	3,878	335,453 ^b	4,000	346,000	,000	,795

a. Design: Intercept
Within Subjects Design: Time

b. Exact statistic

Tests of Within-Subjects Effects							
Measure: MEASURE_1							
Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	342094,143	4	85523,536	569,461	,000	,620
	Greenhouse-Geisser	342094,143	3,009	113704,586	569,461	,000	,620
	Huynh-Feldt	342094,143	3,038	112616,196	569,461	,000	,620
	Lower-bound	342094,143	1,000	342094,143	569,461	,000	,620
	Error(Time)	Sphericity Assumed	209655,857	1396	150,183		
	Greenhouse-Geisser	209655,857	1050,009	199,671			
	Huynh-Feldt	209655,857	1060,157	197,759			
	Lower-bound	209655,857	349,000	600,733			

Effect size (omega square): Vitaliteit (energie)

$$\omega_{Sphericity}^2 = 0,309$$

$$\omega_{G-G}^2 = 0,353$$

Tests of Within-Subjects Contrasts							
Measure: MEASURE_1							
Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Level 1 vs. Level 2	479890,286	1	479890,286	1290,209	,000	,787
	Level 2 vs. Level 3	2943,500	1	2943,500	17,903	,000	,049
	Level 3 vs. Level 4	,286	1	,286	,002	,967	,000
	Level 4 vs. Level 5	,000	1	,000	,000	1,000	,000
Error(Time)	Level 1 vs. Level 2	129809,714	349	371,948			
	Level 2 vs. Level 3	57381,500	349	164,417			
	Level 3 vs. Level 4	59949,714	349	171,776			
	Level 4 vs. Level 5	73350,000	349	210,172			

Effect size (r)

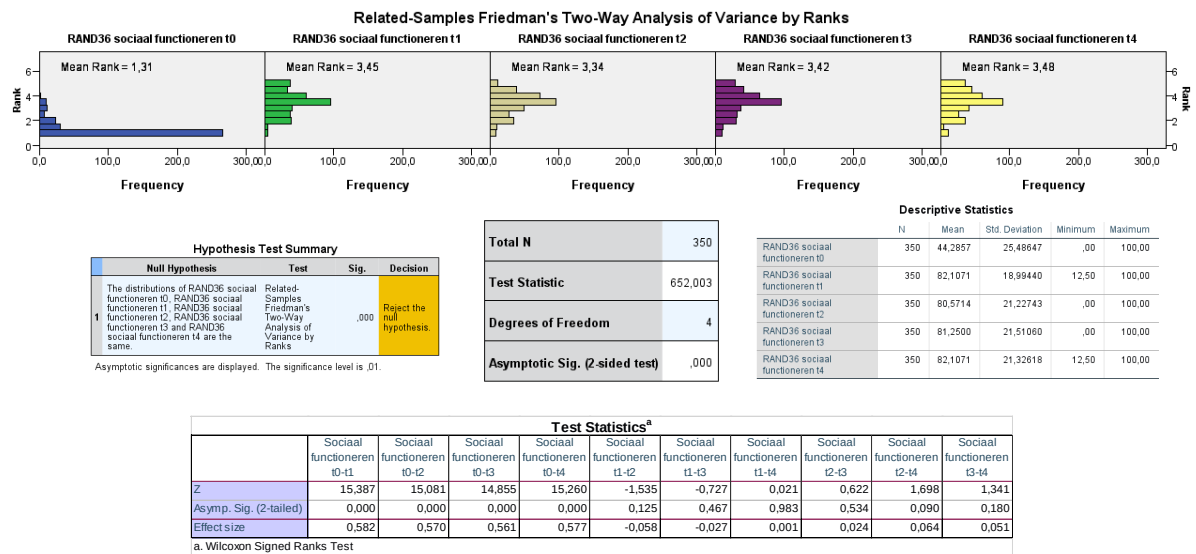
$$r_{\text{level 1 vs level 2}} = 0,693$$

$$r_{\text{level 2 vs level 3}} = 0,113$$

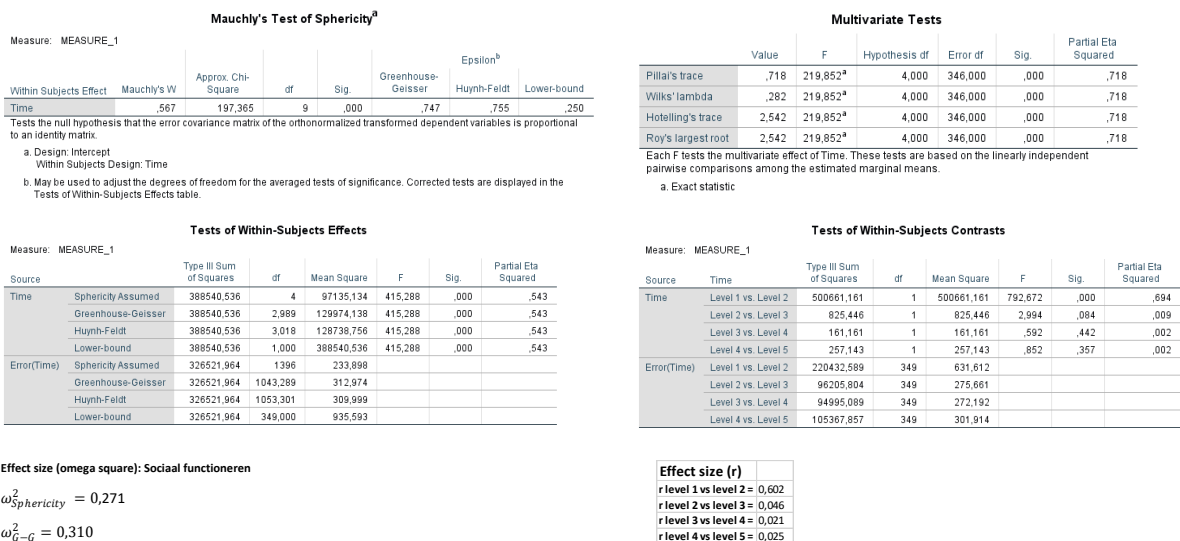
$$r_{\text{level 3 vs level 4}} = 0,001$$

$$r_{\text{level 4 vs level 5}} = 0,000$$

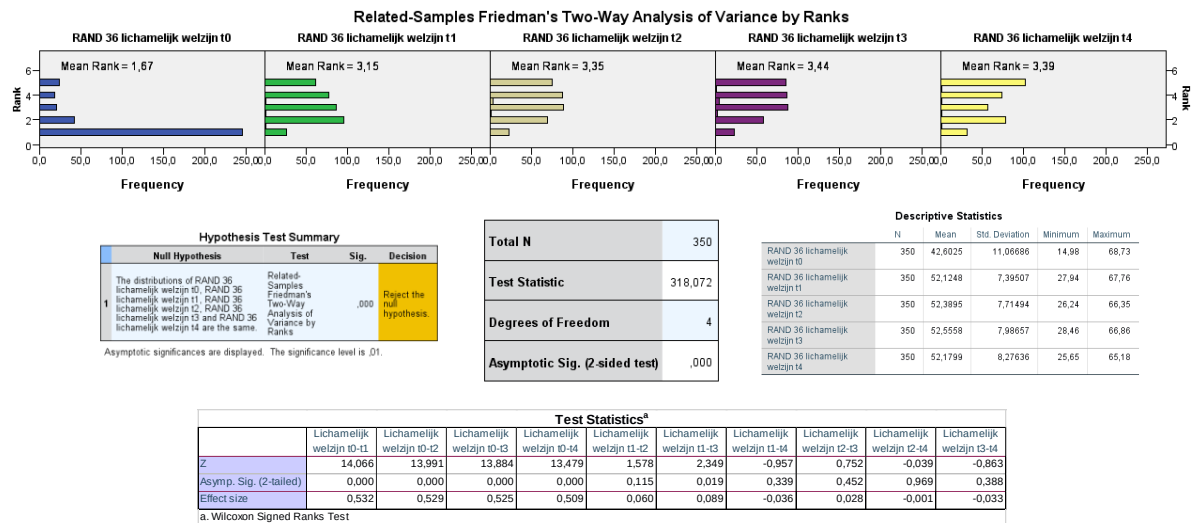
Sociaal functioneren: non-parametrisch



Sociaal functioneren: parametrisch



Lichamelijk welzijn: non-parametrisch



Lichamelijk welzijn: parametrisch

Mauchly's Test of Sphericity^a

Measure: MEASURE_1

		Approx. Chi-Square			Greenhouse-Geisser	Epsilon ^b	
Within Subjects Effect	Mauchly's W	Square	df	Sig.		Huynh-Feldt	Lower-bound
Time	.425	296,958	9	.000	.671	.677	.250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: Time

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Sphericity Assumed	26441,116	4	6610,279	221,913	,000	,389
	Greenhouse-Geisser	26441,116	2,684	9849,657	221,913	,000	,389
	Huynh-Feldt	26441,116	2,707	9766,759	221,913	,000	,389
	Lower-bound	26441,116	1,000	26441,116	221,913	,000	,389
Error(Time)	Sphericity Assumed	41583,702	1396	29,788			
	Greenhouse-Geisser	41583,702	936,880	44,385			
	Huynh-Feldt	41583,702	944,832	44,012			
	Lower-bound	41583,702	349,000	119,151			

Effect size (omega square): Lichamelijk welzijn

$$\omega_{Sphericity}^2 = 0,193$$

$$\omega_{G-G}^2 = 0,231$$

Multivariate Tests^a

Effect	Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time Pillai's Trace	,542	102,445 ^b	4,000	346,000	,000	,542
Wilks' Lambda	,458	102,445 ^b	4,000	346,000	,000	,542
Hotelling's Trace	1,184	102,445 ^b	4,000	346,000	,000	,542
Roy's Largest Root	1,184	102,445 ^b	4,000	346,000	,000	,542

a. Design: Intercept
Within Subjects Design: Time

b. Exact statistic

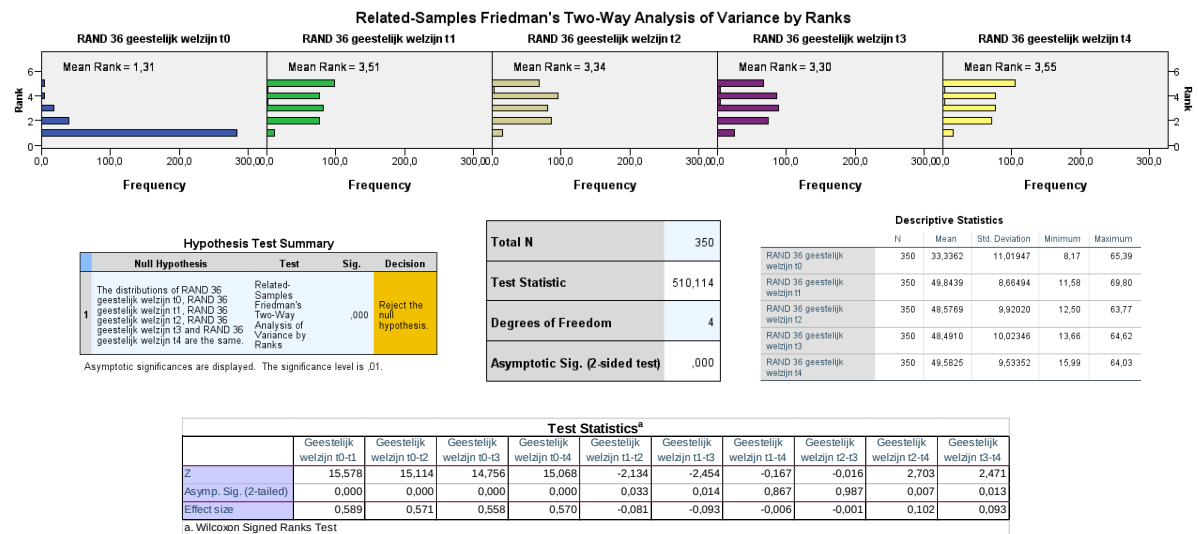
Tests of Within-Subjects Contrasts

Source	Time	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Time	Level 1 vs. Level 2	31736,368	1	31736,368	365,026	,000	,511
	Level 2 vs. Level 3	24,516	1	24,516	,871	,351	,002
	Level 3 vs. Level 4	9,684	1	9,684	,319	,572	,001
	Level 4 vs. Level 5	49,452	1	49,452	1,458	,228	,004
Error(Time)	Level 1 vs. Level 2	30343,017	349	86,943			
	Level 2 vs. Level 3	9824,264	349	28,150			
	Level 3 vs. Level 4	10583,878	349	30,326			
	Level 4 vs. Level 5	11841,219	349	33,929			

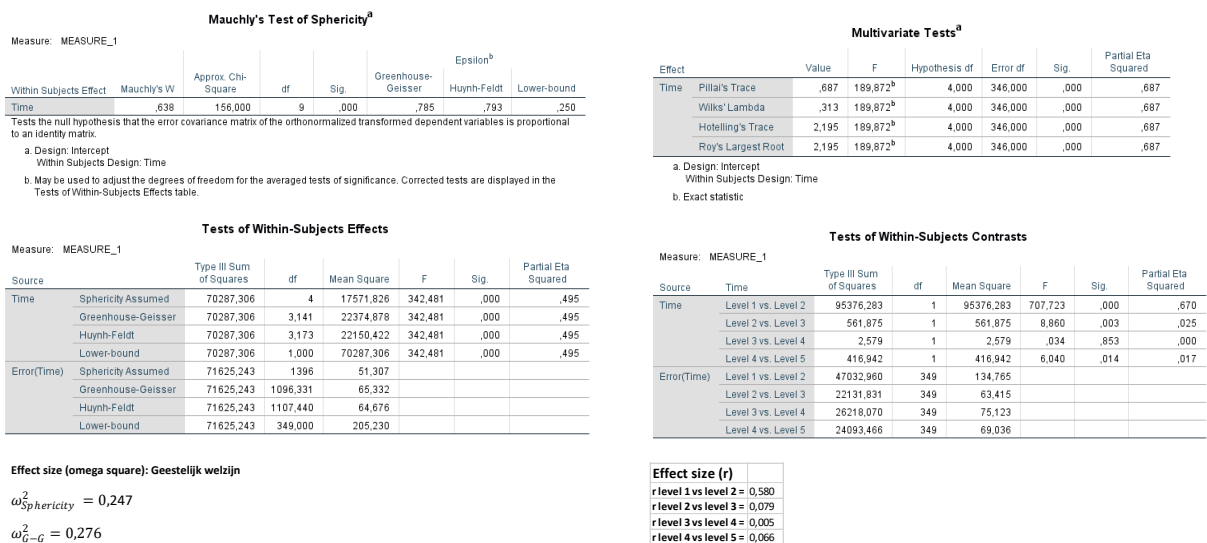
Effect size (r)

r level 1 vs level 2 =	0,455
r level 2 vs level 3 =	0,025
r level 3 vs level 4 =	0,015
r level 4 vs level 5 =	0,032

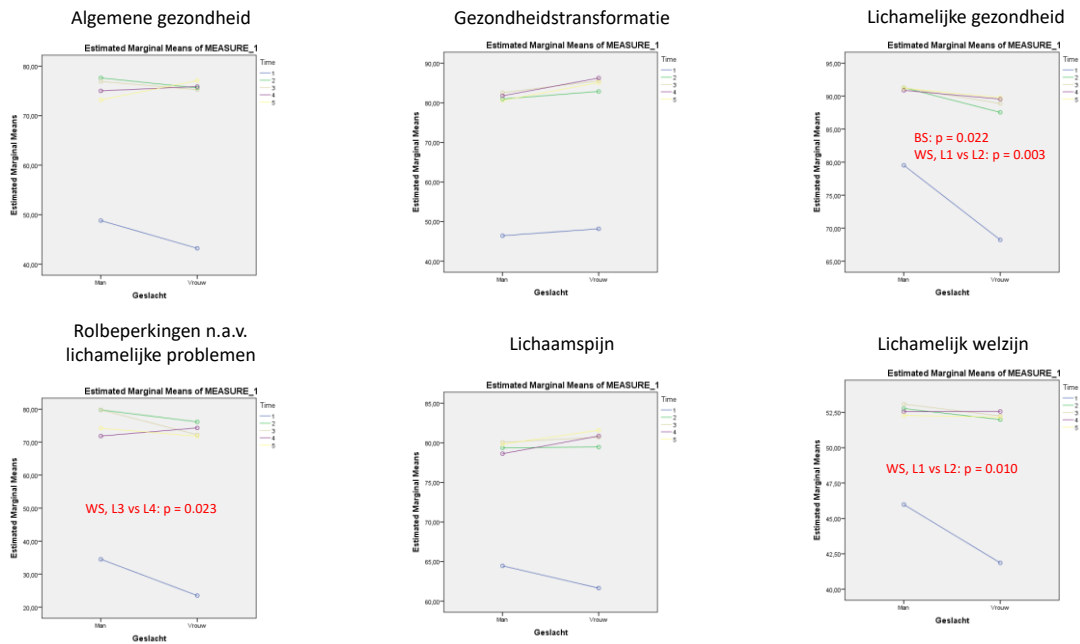
Geestelijk welzijn: non-parametrisch



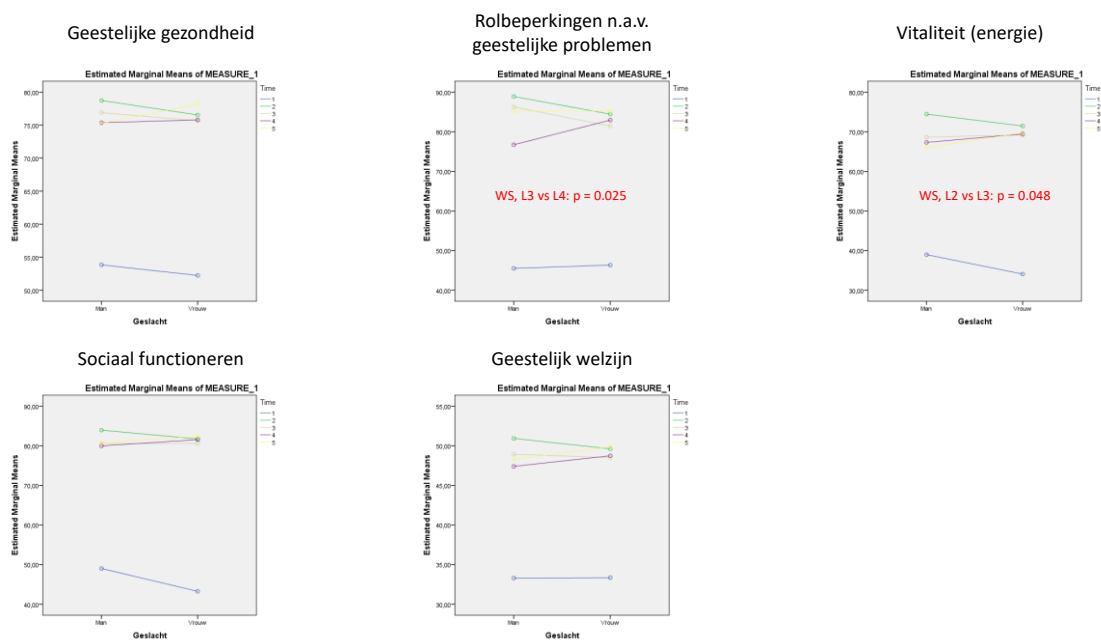
Geestelijk welzijn: parametrisch



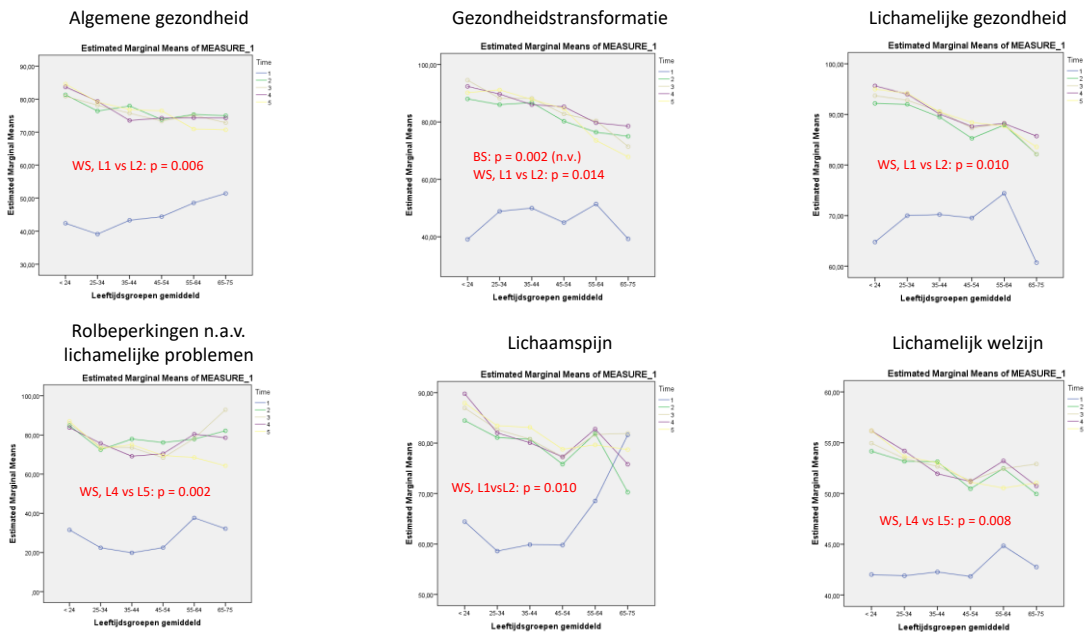
Geslacht x tijdsmetingen: lichamelijk



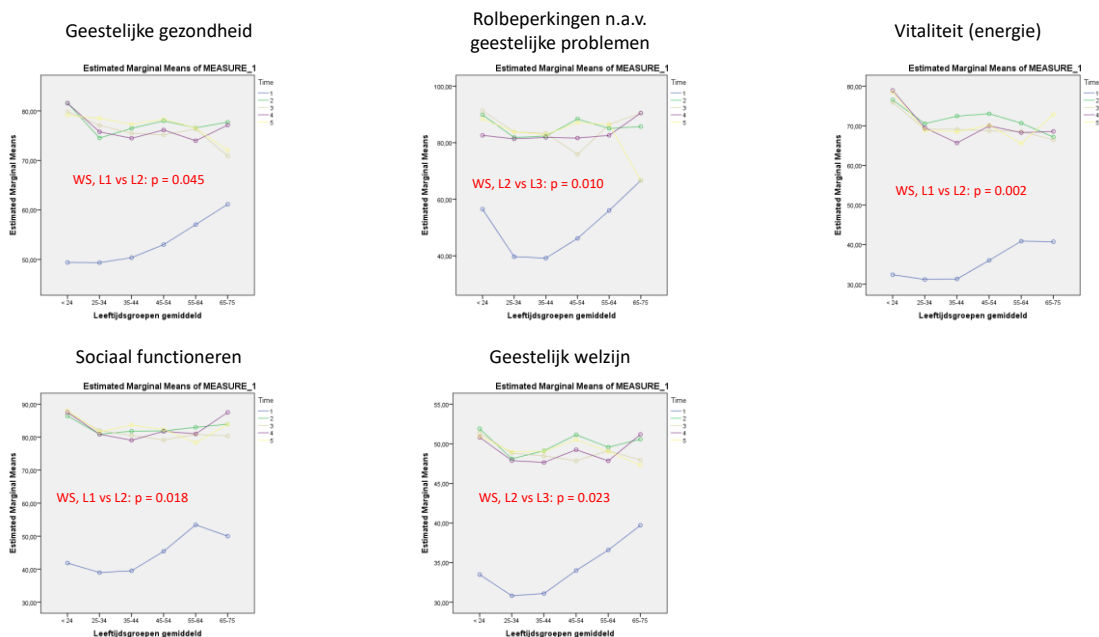
Geslacht x tijdsmetingen: geestelijk



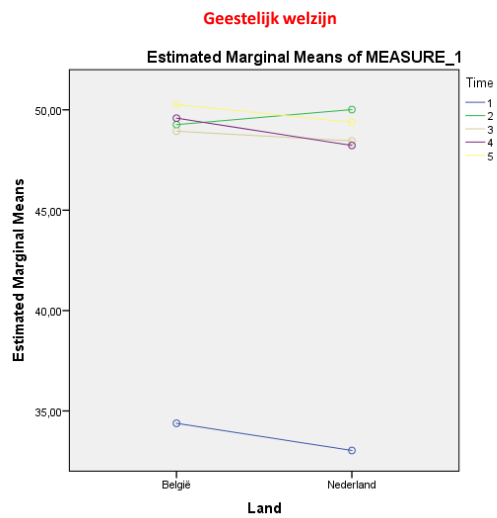
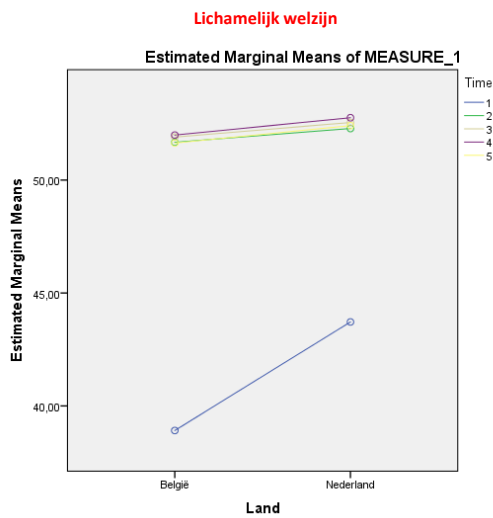
Leeftijd x tijdsmetingen: lichamelijk



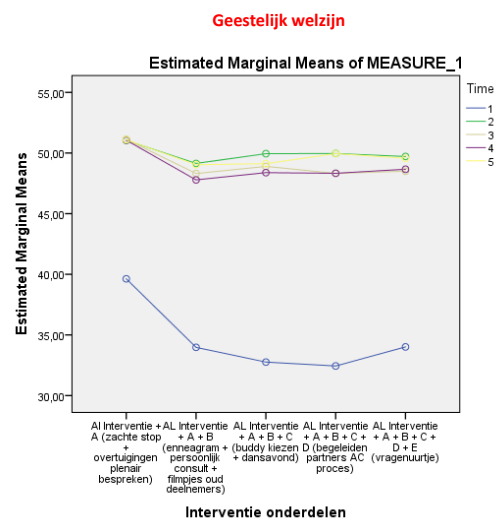
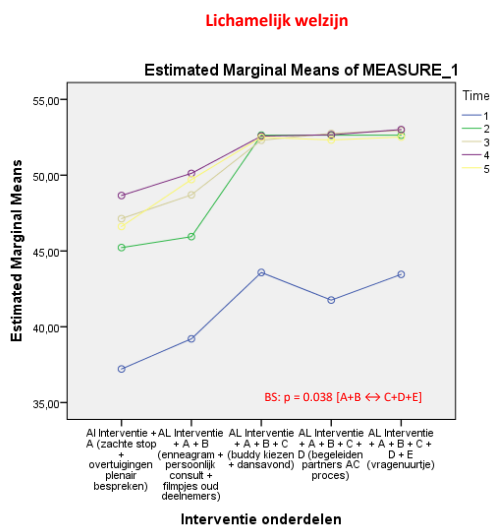
Leeftijd x tijdsmetingen: geestelijk



Land x tijdsmetingen



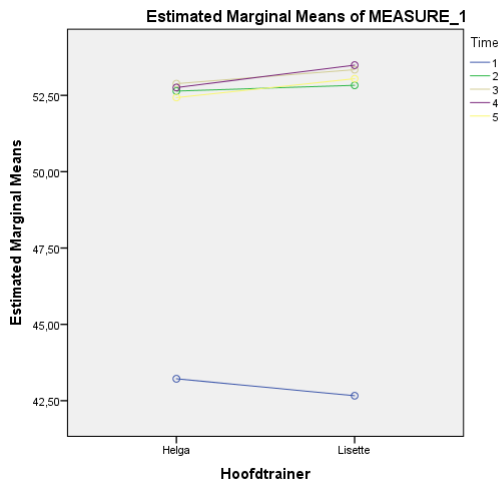
Interventie x tijdsmetingen



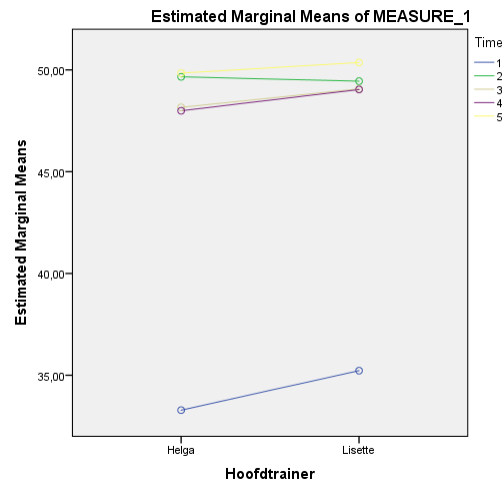
Hoofdtrainer x tijdsmetingen

Between-Subjects Factors		
	Value Label	N
Hoofdtrainer	1,00 Helga	131
	2,00 Lisette	41

Lichamelijk welzijn



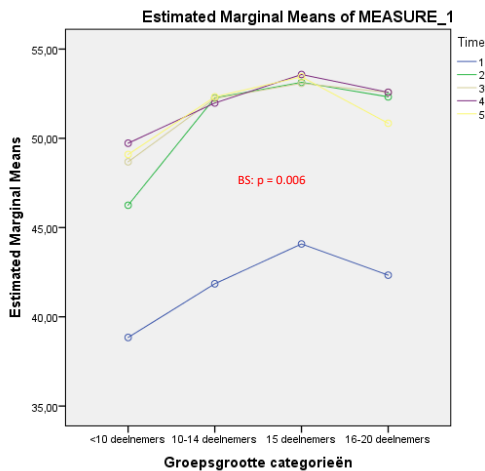
Geestelijk welzijn



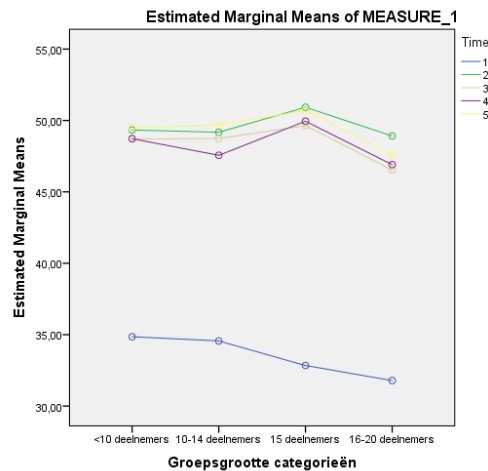
Groepsgrootte x tijdsmetingen

Between-Subjects Factors		
	Value Label	N
Groepsgrootte categorieën	<10 deelnemers	30
	10-14 deelnemers	97
	15 deelnemers	141
	16-20 deelnemers	80

Lichamelijk welzijn



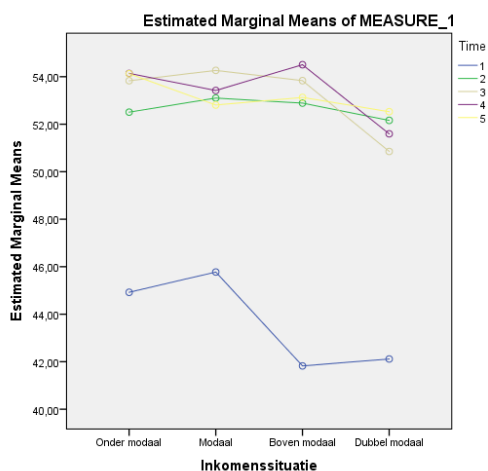
Geestelijk welzijn



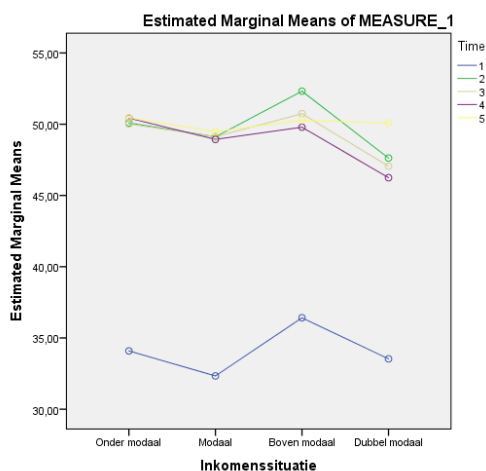
Inkomenssituatie x tijdsmetingen

Between-Subjects Factors		
	Value Label	N
Inkomensituatie	1,00 Onder modaal	39
	2,00 Modaal	12
	3,00 Boven modaal	38
	4,00 Dubbel modaal	13

Lichamelijk welzijn



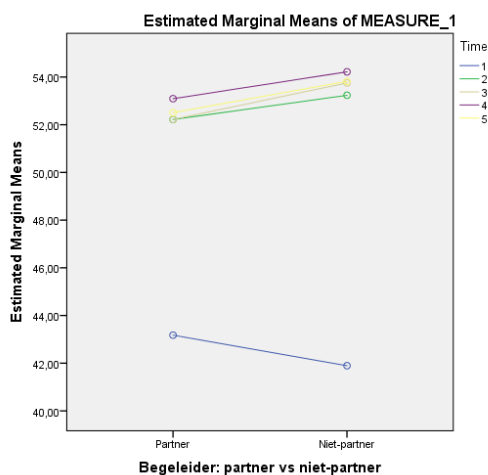
Geestelijk welzijn



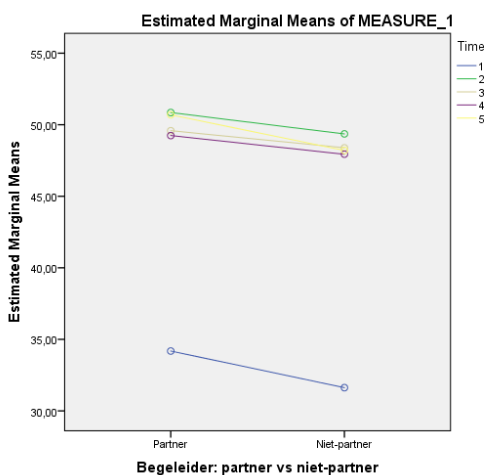
Begeleider x tijdsmetingen

Between-Subjects Factors		
	Value Label	N
Begeleider: partner vs niet-partner	1,00 Partner	164
	2,00 Niet-partner	34

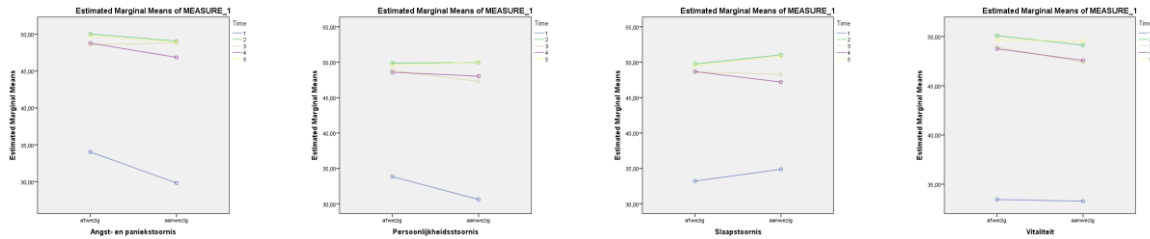
Lichamelijk welzijn



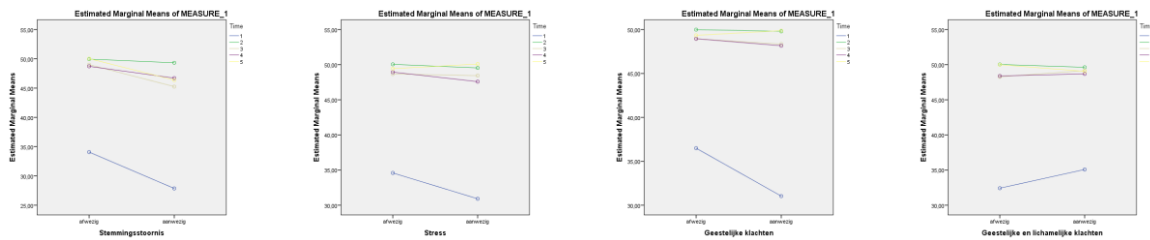
Geestelijk welzijn



Klachten x tijdsmetingen: geestelijk welzijn



Klachten x tijdsmetingen: geestelijk welzijn



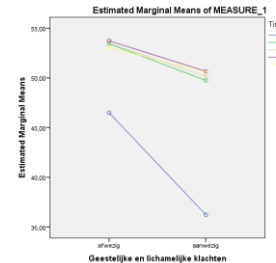
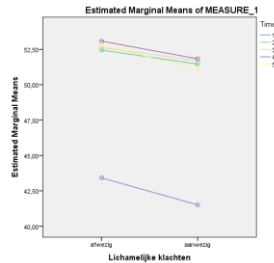
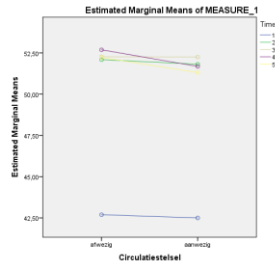
Stemmingsstoornis: n = 36
 Multivariate Test: tijd x stemming, WL: p = 0.027
 Within Subjects Effect: tijd x stemming, H-F: p = 0.037
 Test Within Subjects Contrast: tijd x stemming:
 • L1 vs L2: p = 0.006
 • L2 vs L3: p = 0.028
 Test Between Subjects Effect: p = 0.014

Stress: n = 108
 Multivariate Test: tijd x stress, WL: p = 0.022
 Within Subjects Effect: tijd x stress, H-F: p = 0.007
 Test Within Subjects Contrast: tijd x stress:
 • L1 vs L2: p = 0.019
 • L2 vs L3: p = 0.039

Geestelijke klachten (agg): n = 190
 Multivariate Test: tijd x GK, WL: p = 0.000
 Within Subjects Effect: tijd x GK, H-F: p = 0.000
 Test Within Subjects Contrast: tijd x GK:
 • L1 vs L2: p = 0.000

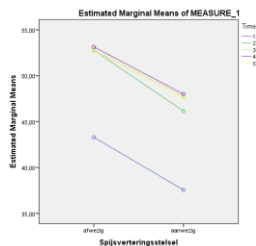
Geestelijke en lichamelijke klachten (agg): n = 125
 Within Subjects Effect: tijd x GLK, H-F: p = 0.027
 Test Within Subjects Contrast: tijd x GLK:
 • L1 vs L2: p = 0.018

Klachten x tijdsmetingen: lichamelijk welzijn

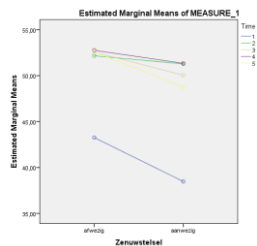


Lichamelijke en geestelijke klachten (agg): n = 125
 Multivariate Test: tijd x LGK, WL: p = 0.000
 Within Subjects Effect: tijd x LGK, G-G: p = 0.027
 Test Within Subjects Contrast: tijd x LGK:
 • L1 vs L2: p = 0.000
 Test Between Subjects Effect: p = 0.024

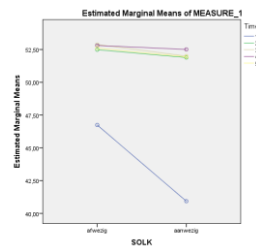
Klachten x tijdsmetingen: lichamelijk welzijn



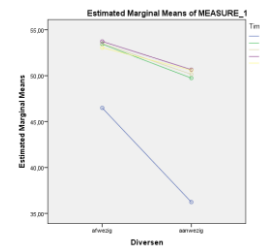
Spijverteringsstelsel: n = 36
 Test Between Subjects Effect: p = 0.000



Zenuwstelsel: n = 41
 Multivariate Test: tijd x zenuw, WL: p = 0.012
 Within Subjects Effect: tijd x zenuw, G-G: p = 0.026
 Test Within Subjects Contrast: tijd x zenuw:
 • L1 vs L2: p = 0.012
 • L4 vs L2: p = 0.012
 Test Between Subjects Effect: p = 0.024



SOLK: n = 235
 Multivariate Test: tijd x SOLK, WL: p = 0.000
 Within Subjects Effect: tijd x SOLK, G-G: p = 0.000
 Test Within Subjects Contrast: tijd x SOLK:
 • L1 vs L2: p = 0.000



Diversen: n = 125
 Multivariate Test: tijd x Divers, WL: p = 0.000
 Within Subjects Effect: tijd x Divers, G-G: p = 0.000
 Test Within Subjects Contrast: tijd x SOLK:
 • L1 vs L2: p = 0.000
 Test Between Subjects Effect: p = 0.000

Bijlage 10: Repeated Measures test nader bekeken

In dit onderzoek is de assumptie van sphericiteit geschonden en dat is een voorwaarde om de parametrische Repeated Measures (MS) test uit te voeren (zie bijlage 5). Het schenden van deze assumptie is bepaald aan de hand van Mauchly's test. Echter, het probleem van deze test is dat in onderzoeken met een grote steekproefomvang ($n > 100$) een kleine afwijking in sphericiteit al significant kan zijn (Field, 2018). Aangezien de steekproefomvang in de onderzoek 350 is, kan deze afwijking dus ook hier optreden. Daarom kan binnen de RM test een zogenaamde Bonferroni correctie worden toegepast (Field, 2018). Hierbij wordt de vrijheidsgraden voor alle F-ratio's aangepast die door de schending beïnvloed zijn (bij 1 is de data 'spherical' en bij minder dan 1 niet). Kleinere vrijheidsgraden maken de p-waarde die geassocieerd is met de F-ratio minder significant. De vrijheidsgraden worden aangepast door:

- Greenhouse-Geisser test. Deze varieert tussen $1/(k - 1)$, waarbij 'k' het aantal tijdsmetingen representeert. In dit onderzoek zijn er 5 tijdsmetingen, waardoor de 'lower bound' 0,25 is en de waarden op deze test daarboven moeten uitkomen.
- Huynh-Feldt test. Deze wordt gebruikt als de Greenhouse-Geisser test groter is dan 0,75.

Binnen de RM procedure zijn deze twee testen gebruikt om te analyseren of er een significant verschil tussen de tijdsmetingen bestaat. Dit is gebeurt binnen de 'Test of Within-Subjects Effects' (waarbij $p < 0.001$). Daarnaast zijn binnen de RM test ook twee multivariate test bekeken: Pillai's Trace en Wilks' Lambda om een significant verschil tussen de tijdsmetingen te evalueren (waarbij $p < 0.001$). Ook deze kunnen worden gebruikt als de assumptie van sphericiteit geschonden is (Field, 2018).

Daarnaast zijn binnen de RM test de onderlinge tijdsmetingen met elkaar vergeleken (in hoeverre deze significant van elkaar verschillen), via de 'Test of Within-Subjects Contrast': een verschil tussen de voormeting en eerste nameting, eerste nameting en tweede nameting, tweede nameting en derde nameting en derde nameting en vierde nameting. Zijdeling is ook gekeken naar de 'Pairwise Comparisons' om te kijken of er nog significante verschillen zijn tussen alle combinaties van tijdsmetingen (zie ook bijlage 9).