

# Samenvatting



zerow.  
11.11.2011

### **Subgroepen bij patiënten met lage rugklachten**

De meeste mensen hebben wel eens last (gehad) van lage rugklachten. Deze klachten kunnen leiden tot een verminderde functionaliteit, kwaliteit van leven, ziekteverzuim en zelfs arbeidsongeschiktheid. De economische gevolgen van lage rugklachten zijn groot. Veel mensen met lage rugklachten worden conservatief behandeld en bezoeken de fysiotherapeut. In de afgelopen decennia zijn vele randomized controlled trials (RCTs) gepubliceerd over conservatieve behandelmethoden bij lage rugpijn. Die studies laten niet altijd dezelfde resultaten zien over de werkzaamheid (effectiviteit) van behandelmethoden en meestal zijn de aangetoonde effecten klein. Een veel geopperde verklaring hiervoor is de heterogeniteit van de populatie van mensen met lage rugklachten. Veel fysiotherapeuten zijn van mening dat patiënten met lage rugklachten kunnen worden onderscheiden in subgroepen die verschillend reageren op verschillende behandelvormen. Omdat het bij mensen met lage rugklachten veelal onmogelijk is een diagnose te stellen op basis van patho-anatomische of pathofysiologische afwijkingen, worden meestal klinische kenmerken en symptomen gebruikt om patiënten in te delen in subgroepen. Er wordt gesuggereerd dat de effectiviteit van fysiotherapie kan worden verbeterd, door aan deze subgroepen specifieke behandelvormen te koppelen. Op dit moment bestaan er veel verschillende classificatiesystemen en is er onduidelijkheid of het gebruik van deze classificatiesystemen ook daadwerkelijk leiden tot betere behandelresultaten. In dit proefschrift staat het onderzoek naar subgroepen bij patiënten met lage rugklachten centraal. Sinds 1996 wordt onderzoek naar klinisch relevante subgroepen beschouwd als topprioriteit in het onderzoeksveld naar lage rugklachten.

**Hoofdstuk 1** beschrijft de achtergrond van het onderzoek, de onderzoeksvraagstellingen en geeft een overzicht van de studies die in dit proefschrift aan de orde komen. Deze studies zijn in twee delen opgedeeld. De hoofdstukken 2-4 zijn gebaseerd op een observationeel cohort onderzoek en gaan over de mogelijkheid om patiënten met psychologische problematiek te identificeren en de hoofdstukken 5-7 gaan over een RCT waarin de kosten- effectiviteit van behandelingen gekoppeld aan een classificatiesysteem wordt onderzocht.

In **hoofdstuk 2** wordt de ontwikkeling van een screeningsinstrument beschreven in een poliklinische revalidatiesetting om patiënten met chronische lage rugpijn (> 13 weken) al of niet te selecteren voor aanvullend psychologisch onderzoek. Bij 229 opeenvolgende patiënten werden, voorafgaande aan de behandeling, 17 variabelen uit de anamnese en het fysiotherapeutisch onderzoek geprotocolleerd vastgelegd. De gegevens van de

---

fysiotherapeut werden vergeleken met het oordeel van de psycholoog. Het oordeel van de psycholoog fungeerde als de 'gouden' standaard, namelijk: wel of geen aanwijzingen voor psychopathologische stoornissen en/of processen die interfereren met de lage rugklachten. De psycholoog vormde zijn oordeel door het voeren van twee gesprekken met de patiënt en de evaluatie van 5 psychologische vragenlijsten. De fysiotherapeut en de psycholoog waren niet op de hoogte van elkaars bevindingen. In de univariate analyse bleken scores op 10 van de 17 variabelen statistisch significant geassocieerd met het oordeel van de psycholoog. Multivariate logistische regressie analyse resulteerde in een model met 4 variabelen: een positieve Waddell score (aanwijzing voor verhoogd pijngedrag), meer locaties op een pijntekening, de afwezigheid van een houdings- en bewegingsvoorkeur en het dagelijks gebruik van pijnmedicatie voor de lage rugklachten (oppervlakte onder de receiver operating characteristic [ROC] curve 0,81 [95% betrouwbaarheidsinterval (BI) 0,75 tot 0,87]). Aan de hand van deze resultaten is een screeningsinstrument ontwikkeld dat de kans op relevante, met de lage rugklachten interfererende, psychopathologische stoornissen en processen voorspelt. Patiënten die hoog scoren op dit screeningsinstrument zouden, naast lichamelijk onderzoek, gediend zijn met aanvullend psychologisch onderzoek. Hoewel de resultaten van dit onderzoek interessant zijn, zal onderzoek in een vergelijkbare onderzoekspopulatie moet uitwijzen of de bevindingen herhaald kunnen worden en in hoeverre het ontwikkelde screeningsinstrument klinisch relevant is in andere settingen dan een poliklinische revalidatieafdeling.

**Hoofdstuk 3** geeft de resultaten weer van onderzoek naar de intra- en inter-beoordelaarsbetrouwbaarheid van de Waddell score. De Waddell score was in hoofdstuk 2 één van de variabelen die opgenomen zijn in het screeningsinstrument. De Waddell score omvat 8 gestandaardiseerde fysieke, niet organische Waddell testen ofwel pijn gerelateerde gedragsmatige testen (de zogeheten 'non-organic signs' of 'Waddell signs') die zijn verdeeld over 5 categorieën. Een categorie is positief als tenminste één test uit die categorie positief is. De Waddell score is positief bij 3 of meer positieve categorieën. Twee fysiotherapeuten van een poliklinische revalidatieafdeling hebben de Waddell score afgenomen bij 126 patiënten met chronische lage rugklachten. Elke patiënt werd door beide therapeuten tweemaal gemeten; eenmaal bij aanvang van een observatieperiode en eenmaal aan het einde ervan (14 dagen later). De twee fysiotherapeuten namen de tests direct na elkaar en onafhankelijk van elkaar af. De interne consistentie (Cronbach's alpha) werd bepaald van de originele Waddell score met 5 categorieën (score 0-5) en van de Waddell score wanneer de 8 testen afzonderlijk

werden opgeteld (score 0-8). De betrouwbaarheid werd bepaald met Cohen's kappa. De prevalentie van een positieve (originele) Waddell score was 36% (n = 45). Voor beide fysiotherapeuten was de interne consistentie bij de originele Waddell score lager (0,65 en 0,72) dan bij de Waddell score waarbij de 8 testen afzonderlijk werden opgeteld (0,71 en 0,78). Een Cronbach's alpha boven de 0,70 betekent dat de items die de Waddell score vormen hetzelfde concept meten. De inter-beoordelaarsbetrouwbaarheid met een gemiddeld tijdsinterval van 14 dagen leverde een kappa op van 0,48 (95% BI 0,30 tot 0,65). Voor de intra-beoordelaarsbetrouwbaarheid met eveneens een gemiddeld tijdsinterval van 14 dagen werden kappa waardes gevonden van 0,65 (95% BI 0,50 tot 0,80) en 0,68 (95% BI 0,53 tot 0,84). Dit leidde tot de conclusie, dat voor getrainde beoordelaars in een populatie van patiënten met chronische lage rugklachten in een poliklinische revalidatiesetting de inter-beoordelaarsbetrouwbaarheid redelijk was en de intra-beoordelaarsbetrouwbaarheid goed. Tot op heden is er weinig onderzoek gepubliceerd over de betrouwbaarheid van de Waddell score. De inter-beoordelaarsbetrouwbaarheid kan mogelijk nog verhoogd worden door een betere standaardisatie van de beoordeling van de Waddell signs.

**Hoofdstuk 4** presenteert de resultaten van een onderzoek naar de construct validiteit van de Waddell score. In deze cross-sectionele studie werden op basis van een gedetailleerd literatuuronderzoek in Medline (periode 1980-2010) *a priori* 20 hypothesen geformuleerd over de grootte en de richting van de associatie tussen de Waddell score met variabelen uit 5 verschillende domeinen (demografisch, pijn, fysiek, pijngedrag en psychologisch). In totaal werden 229 opeenvolgende patiënten met chronische lage rugklachten in een poliklinische revalidatieafdeling onderzocht door psychologen en fysiotherapeuten. De patiënten ondergingen een lichamelijk onderzoek waarbij o.a. de Waddell score werd bepaald. Bovendien werden met behulp van een anamnese en een uitgebreide set vragenlijsten demografische gegevens verzameld en gegevens over de mate van pijn, pijngedrag en de fysieke en de psychologische gesteldheid. Van de *a priori* hypothesen werd 75% bevestigd. Op basis daarvan werd geconcludeerd dat de cross-sectionele construct validiteit van de Waddell score voldoende was. De resultaten van dit onderzoek lieten tevens zien, dat de associaties tussen de Waddell score en de variabelen uit de verschillende domeinen niet erg hoog zijn. Dat betekent dat de aanwezigheid van positieve Waddell testen weinig richting geven aan de specifieke problemen bij de patiënt. De Waddell score kan daarom niet als een zelfstandig psychologisch screeningsinstrument worden beschouwd.

---

In hoofdstuk 5 is het ontwerp beschreven van onderzoek naar de effectiviteit en kosteneffectiviteit van een classificatie-behandelsysteem voor patiënten met lage rugklachten. Dit classificatie-behandelsysteem koppelt patiënten met een bepaalde symptomatologie aan specifieke fysiotherapeutische behandelinterventies. Het systeem werd ontwikkeld door Delitto en zijn collega's en in 1995 gepubliceerd. Twee studies uit de VS hebben dit systeem vergeleken met andere fysiotherapeutische behandelstrategieën bij patiënten met acute (0-6 weken) en sub-acute (7-12 weken) klachten en milde tot ernstige beperkingen (> 25% beperkt op de Oswestry Disability Questionnaire [ODI]). Beide studies vonden licht positieve resultaten ten gunste van Delitto's classificatie-behandelsysteem. De vraagstelling van het hier beschreven onderzoek was of het gebruik van Delitto's systeem effectiever is dan reguliere fysiotherapeutische zorg in ons Nederlandse zorgstelsel. Voor het onderzoek werd het classificatie-behandelsysteem geactualiseerd naar de huidige stand der wetenschap en aangepast aan de Nederlandse situatie. Patiënten met acute lage rugpijn werden uitgesloten van deelname, omdat fysiotherapeutische behandeling in de acute fase wordt afgeraden in de Nederlandse richtlijnen. Fysiotherapeuten in de eerste lijn wierven de patiënten. Vier fysiotherapeut-onderzoekers, die niet in het behandelproces betrokken waren, classificeerden alle patiënten volgens het gemodificeerde classificatie-behandelsysteem in de volgende drie categorieën: 1. richting-specifieke oefeningen, 2. manipulatie en 3. stabiliserende oefeningen. De patiënten werden door middel van loting toegewezen aan classificatiebehandeling of aan reguliere fysiotherapie volgens Nederlandse richtlijnen. Loting werd uitgevoerd door een onafhankelijk persoon met een vooraf opgesteld randomisatieschema dat was gegenereerd door een computer. Patiënten die waren toegewezen aan classificatiebehandeling ontvingen gedurende minimaal 4 weken de specifieke behandeling die bij hun classificatie hoorde. De patiënten die waren toegewezen aan reguliere fysiotherapie ontvingen fysiotherapeutische behandeling volgens de Nederlandse richtlijn. Vragenlijsten werden afgenomen op het moment van insluiting en vervolgens 8, 26 en 52 weken na aanvang van de behandeling. De patiënten en de onderzoeksassistenten die de vragenlijsten verzamelden waren geblindeerd voor het resultaat van de lotingprocedure. De behandelende fysiotherapeuten en de statisticus waren wel geïnformeerd. De primaire uitkomstmaten waren ervaren herstel (7-punts-likert-schaal), functionele status (ODI) en pijnintensiteit (Numerical Rating Scale [NRS]). De secundaire uitkomstmaten waren ervaren gezondheid (Short Form 36), kwaliteit van leven (EuroQol), pijn gerelateerde vrees (Fear-Avoidance Beliefs Questionnaire) en psychosociale status (Örebro Musculoskeletal Pain Screening Questionnaire).

Directe en indirecte kosten werden gemeten met kostendagboekjes en de Productivity and Disease Questionnaire (PRODISQ).

**Hoofdstuk 6** presenteert de effectiviteitsresultaten van de RCT. Een totaal van 156 patiënten werd geïnccludeerd (classificatiegroep n = 74; reguliere fysiotherapie groep n = 82). De hoofdanalyse werd uitgevoerd volgens het 'intention to treat'-principe. Volgens dit principe worden alle patiënten geanalyseerd in de groep waaraan zij door loting zijn toegewezen, ongeacht of zij de toegewezen behandeling gevolgd of voltooid hebben. De resultaten van multilevel analyses toonden aan dat beide groepen gedurende de onderzoeksperiode verbeterden, maar dat er geen verschillen waren tussen de groepen op alle uitkomstmaten en voor alle meetmomenten. Na één jaar scoorde de classificatiegroep weliswaar hoger op ervaren herstel (gecorrigeerde odds ratio 2,75 [95% BI 0,84 tot 9,07]), lager op functionaliteitstoornissen (gemiddeld gecorrigeerde verschil 0,86 punten op de ODI [95% BI -3,61 tot 5,33]) en lager op pijnintensiteit (gemiddeld gecorrigeerde verschil 0,13 punten op de NRS (95% BI -0,83 tot 1,10)), maar deze verschillen waren niet statistisch significant. Onze resultaten komen niet geheel overeen met eerder gepubliceerd onderzoek. Dit heeft mogelijk te maken met verschillen in de selectiecriteria, de behandelvorm waarmee classificatiebehandeling werd vergeleken en de voor dit onderzoek doorgevoerde modificatie van het classificatie-behandelsysteem. De conclusie was dat het classificatie-behandelsysteem, zoals toegepast in deze studie, geen betere behandelresultaten oplevert dan reguliere fysiotherapie in een populatie van patiënten met sub-acute en chronische lage rugklachten.

**Hoofdstuk 7** is gebaseerd op dezelfde RCT en beschrijft de kosteneffectiviteit en de kostenutiliteit van het classificatie-behandelsysteem gemeten na 1 jaar. De resultaten van t-tests lieten zien dat de classificatiegroep significant hoger scoorde op ervaren herstel, maar er waren geen verschillen in pijnintensiteit, mate van beperkingen en kwaliteit van leven. De gemiddelde totale kosten voor de maatschappij voor de classificatiegroep was iets hoger dan voor de reguliere fysiotherapie, respectievelijk €2287 en €2020. Het verschil was €266 (95% BI -€720 tot €1612) en niet statistisch significant. Kosteneffectiviteit en -utiliteitsanalyses toonden aan dat het classificatie-behandelsysteem in vergelijking met reguliere fysiotherapie op alle uitkomstmaten niet kosteneffectief was. Op dit moment wordt invoering van het classificatie-behandelsysteem op landelijk niveau niet aanbevolen.

**Hoofdstuk 8** bevat een samenvatting van de belangrijkste bevindingen van dit proefschrift, een algemene discussie en aanbevelingen voor toekomstig onderzoek.

Onderzoek naar klinisch relevante subgroepen wordt al bijna 2 decennia aangegeven/ genoemd als onderzoeksprioriteit. Ook na ons onderzoek is verder onderzoek naar de effectiviteit en kosteneffectiviteit van Delitto's en andere classificatie-behandelsystemen gewenst. We doen daartoe de volgende aanbevelingen:

Degelijk onderzoek naar de betrouwbaarheid van Delitto's classificatie-behandelsysteem is wenselijk. Slechts enkele studies hebben de inter-beoordelaarsbetrouwbaarheid van de Delitto's classificatie-behandelsysteem onderzocht en de kappa waarden varieerden aanzienlijk (0,15 tot 0,65). De resultaten zijn moeilijk te interpreteren, omdat de algoritmes verschilden en de populaties klein waren. Alleen classificaties met een hoge betrouwbaarheid kunnen in de praktijk een meerwaarde opleveren.

Onderzoek is nodig om het classificatie-behandelsysteem te verfijnen en mogelijk uit te breiden zodat andere potentieel relevante subgroepen kunnen worden onderscheiden. Met een ideaal classificatiesysteem is het mogelijk elke patiënt te classificeren in één enkele subgroep. Uit ons onderzoek bleek, dat met het eerste (en belangrijkste) deel van het classificatiesysteem 24% van de patiënten niet te classificeren was en dat 16% voldeed aan de selectiecriteria van meer dan één subgroep.

Verder onderzoek moet aantonen of het opnemen van belangrijke psychologische factoren in het classificatie-behandelsysteem leidt tot betere behandelresultaten. Onderzoek heeft aangetoond dat belangrijke psychologische factoren als depressie en catastrofen een negatieve invloed hebben op behandelresultaten van de fysiotherapeut. Ook aandacht voor de psychologische component is dan gewenst.