

## Systematische review

# Weinig veelbelovende prognostische modellen bij specifieke nekpijn

**Prognostische modellen kunnen zorgverleners helpen bij het behandelen en adviseren van patiënten met specifieke nekpijn. Er zijn diverse prognostische modellen voor nekpijn ontwikkeld, maar deze zijn niet systematisch geëvalueerd met een specifiek beoordelingsinstrument. Na een systematische review naar de validiteit en toepasbaarheid in de eerste lijn van 99 multivariabele prognostische modellen voor herstel van specifieke nekpijn, bleken er slechts 3 beloftevol voor gebruik in de eerste lijn.**

Tekst: Roel Wingbermühle, Emiel van Trijffel, Paul Nelissen, Bart Koes en Arianne Verhagen

### Achtergrond

Nekpijn is na lage rugpijn de belangrijkste musculoskeletale oorzaak van ziekte-last, en de prevalentie van chronische nekpijn neemt wereldwijd toe.<sup>1,2</sup> Manuele therapie is effectief bij volwassenen met specifieke nekpijn. Ze kan pijn verminderen en het functioneren in ADL verbeteren.<sup>3,4</sup> Manuele therapie wordt dan ook in diverse richtlijnen geadviseerd.<sup>5,6</sup>

Multivariabele prognostische modellen combineren meerdere prognostische factoren tegelijkertijd en kwantificeren de *individuele* kans van een patiënt op een bepaalde uitkomst over een bepaalde tijd. Daarmee kunnen ze zorgverleners ondersteunen in klinische beslissingen en adviezen aan de patiënt.<sup>7,8</sup> Regelmatig wordt in de praktijk een van het model afgeleide scorekaart gebruikt, zoals bijvoorbeeld de STarT back tool voor het voorspellen van risico op chroniciteit bij lage rugpijn.<sup>9</sup> Een ander voorbeeld is de scorekaart voor vertraagd herstel van nekpijn, zoals geadviseerd wordt in de KNGF-richtlijn nekpijn.<sup>5</sup>

Het ontwikkelingsproces van een model doorloopt drie opvolgende fases van wetenschappelijk onderzoek: de ontwikkeling zelf (derivatie), validering bij een andere groep nekpijnpatiënten (externe validatie) en de daadwerkelijke impact op de patiëntenuitkomst, gedrag

van zorgverleners en kosteneffectiviteit (impactanalyse).<sup>7,10</sup> Voordat een model met vertrouwen in de kliniek kan worden gebruikt, dient het in ieder geval het proces van externe validatie goed te hebben doorlopen. Er worden specifieke prestatie-maten gebruikt om dit te bepalen (overallperformance, discriminatievermogen en kalibratie).<sup>7,8,11</sup>

Nu zijn er diverse prognostische modellen voor nekpijn ontwikkeld, maar deze waren voor zover ons bekend nog niet systematisch geëvalueerd met een specifiek beoordelingsinstrument. Daarom hebben wij een systematische review uitgevoerd naar de validiteit en toepasbaarheid in de eerste lijn van multivariabele prognostische modellen voor herstel van specifieke nekpijn. Daarbij hebben we gebruikgemaakt van de PROBAST (Prediction model study Risk Of Bias Assessment Tool).<sup>12</sup> De review is onlangs gepubliceerd in het *Journal of Physiotherapy* en hier volgt een samenvatting. Voor detailgegevens en tabellen verwijzen we naar de originele publicatie.<sup>13</sup>

### Methode

De databases van MEDLINE, EMBASE en CINAHL werden tot 3 mei 2016 doorzocht met een gevalideerde zoekstrategie.<sup>14-16</sup> Twee auteurs (RW, PN) screenen de titels en abstracts onafhankelijk van elkaar op in- en exclusiecriteria en bij discrepanties met behulp van een derde auteur (AV). Alle modellen met een uitkomstmaat op korte of lange termijn voor herstel van pijn, functie en ervaren herstel werden meegenomen. De kwaliteit van de geselecteerde studies werd door diezelfde onderzoekers onafhankelijk beoordeeld met behulp van de PROBAST. Dit is een recent gepubliceerd instrument om het risico op vertekende resultaten (risk of bias), toepasbaarheid en bruikbaarheid van de geïncludeerde prognostische modellen te beoordelen. Data-extractie is door de twee auteurs onafhankelijk van elkaar uitgevoerd met de CHARMS-lijst (Checklist for critical

Appraisal and data extraction for systematic Reviews of prediction Modelling Studies) en werd door de derde auteur random gecontroleerd.

Om te evalueren of een model geschikt was voor klinisch gebruik werd een kwalitatieve analyse uitgevoerd, waarbij rekening is gehouden met het risico op bias, toepasbaarheid en bruikbaarheid gerelateerd aan het discriminatievermogen van het model. Een model werd beoordeeld als gereed voor klinisch gebruik indien dit extern gevalideerd was in een studie met laag risico op bias, bruikbaar was in de praktijk en een acceptabel onderscheidend vermogen had. Het onderscheidend vermogen van een model wordt uitgedrukt in de area under the curve (AUC) van de receiver operating characteristic (ROC); een AUC  $\geq 0,7$  (op een schaal van 0-1) werd als acceptabel beschouwd.<sup>17</sup> Aanvullend zochten we nog naar prognostische factoren die consistent werden verkregen in modellen met een laag risico op bias.

### Resultaten

De zoektocht leverde 2398 records op waarvan 98 full-text artikelen potentieel geschikt leken op basis van titel en abstract. Uiteindelijk bleken na doorlezen van de gehele artikelen er 53 geschikt (46 derivatie-, 4 validatie- en 3 combinatiestudies). Deze studies bevatten in totaal 99 modellen voor het voorspellen van het herstel van nekpijn en in 58 daarvan betrof het modellen specifiek voor patiënten met Whiplash-Associated Disorders (WAD). De modellen bevatten tussen de 1 en 10 predictoren, variërend van diverse anamnestiche factoren (bv. leeftijd, geslacht, klachtenduur, psychosociale en contextuele factoren) tot fysieke factoren (bv. range of motion, pijnintensiteit of temperatuurdrempel).

### Derivatiestudies

*Prestaties en kwaliteit van modellen* – De studies rapporteerden voor zeven modellen een discriminatievermogen (AUC) tussen de 0,66 en 0,93.

Alle 49 derivatiestudies werden beoordeeld als studies met een hoog risico op bias. Zeven studies beschreven modellen die minder geschikt waren voor gebruik in de eerste lijn. De algemene performance werd in 34 studies vermeld, het discriminatievermogen en/of kalibratie in 11 studies. Twee studies voerden interne validatie (validering in dezelfde groep patiënten) uit.

### Validatiestudies

**Prestaties en kwaliteit van modellen** – Er werden 7 externe validatie studies uitgevoerd, al of niet direct beschreven in de derivatiestudie. In vier studies werden modellen voor herstel van nekpijn gevalideerd, in drie studies modellen voor herstel van WAD.

Drie studies werden beoordeeld als overall laagrisico op bias. Deze beschrijven we hier verder. Van deze studies was er één bedoeld voor gebruik bij mensen met specifieke nekpijn<sup>18</sup> en twee voor gebruik bij WAD.<sup>19,20</sup> Het discriminatievermogen van het specifieke nekpijnmodel was AUC 0,65 (95% betrouwbaarheidsinterval [BI] 0,59-0,71). Er waren weinig bedenkingen m.b.t. toepasbaarheid en het model werd als bruikbaar beoordeeld. Het discriminatievermogen van het eerste WAD-model was AUC 0,85 (95% BI 0,79-0,91) en bij kalibratie rapporteerde de studie een overschatting. Het model is bruikbaar, maar de toepasbaarheid in de eerste lijn is onduidelijk. De tweede WAD-studie onderzocht de validiteit van twee paden uit een model (volledig herstel; matige tot ernstige problemen met functioneren in ADL) en presenteerde geen overallperformance of kalibratiematen, maar wel de sensitiviteit en specificiteit voor functie als uitkomstmaat. Er waren weinig bedenkingen m.b.t. toepasbaarheid en het model is bruikbaar. Voor een precieze beschrijving van deze modellen verwijzen we naar onze originele publicatie.

### Prognostische factoren in de modellen

Het specifieke nekpijnmodel bevatte de initiële pijnintensiteit, leeftijd, hoofdpijn, uitstralende pijn naar de arm, eerdere nekpijn, traumatische oorzaak van de klacht, lage rugpijn, werkstatus, score op de EuroQol (kwaliteit van leven). Het eerste WAD-model bevatte de initiële NDI, leeftijd, bewegingsuitslag van de linksrotatie, koude-pijdrempel, quotiënt

voor vasoconstrictie en de totaalscore op de impact- of eventsschaal. Het tweede WAD-model bevatte in beide paden de initiële NDI en leeftijd, met de hyperarousal subschaal van de posttraumatische diagnostische schaal in het pad voor matige tot ernstige problemen met functioneren in ADL.

Prognostische factoren die consistent in alle drie modellen voorkwamen, waren de leeftijd en de initiële NDI-score bij WAD-modellen.

### Discussie

Deze review includeerde 53 studies die 99 modellen voor herstel van nekpijn bevatten, waarbij zeven modellen werden onderzocht in validatiestudies. De meeste modellen werden niet verder onderzocht door middel van extern validatieonderzoek en er waren nog geen impactstudies gedaan. Deze bevindingen komen overeen met andere systematische reviews.<sup>21-24</sup>

De studies in deze review werden geëvalueerd met up-to-datecriteria. Prestatiematen voor modellen werden lang niet altijd gerapporteerd en de methodologische kwaliteit van de derivatiestudies was over het algemeen laag. Een van de grote problemen bij derivatieonderzoek is overschatting (overfitting), waarbij het gevonden model accuraat voorspelt in de derivatiedata, maar slecht(er) voorspelt in een nieuwe, vergelijkbare populatie. Dit gebeurt vooral wanneer er veel prognostische factoren worden geselecteerd voor analyse in relatie tot kleine datasets, wat bij veel van deze studies het geval bleek. Het is van belang om het aantal factoren in het analyseproces te bepalen in relatie tot de groepsgrootte op het moment van follow-up. Een recente publicatie beschrijft nu een methode hiervoor,<sup>25,26</sup> maar het overgrote deel van de geïncludeerde studies benaderde niet eens de bekende vuistregel van 1 variabele op een aantal van 10 uitkomsten.<sup>27</sup> Voor overfitting is te corrigeren door de waarden van predictoren in de verkregen modellen aan te passen (shrinkage) naar aanleiding van de bevindingen uit interne validatie. Slechts twee studies bleken deze interne validatie toe te passen.

Tijdens deze review was er nog geen handleiding voor het gebruik van PROBAST bekend, hetgeen een beperking is van ons onderzoek. Enkele WAD-

Gepubliceerd in: Journal of Physiotherapy 2018;64:16-23. Few promising multivariable prognostic models exist for recovery of people with non-specific neck pain in musculoskeletal primary care: a systematic review.

studies bleken ook mensen buiten de eerste lijn geïnccludeerd te hebben, wat de modellen kan hebben beïnvloed en nader validatieonderzoek in de eerste lijn nodig maakt.

### Conclusie

Eén prognostisch model dat de kans op herstel voorspelt bij mensen met specifieke nekpijn en twee modellen bij mensen met WAD bleken beloftevol voor gebruik in de eerste lijn. We adviseren verder extern validatieonderzoek met deze modellen in een eerstelijnssetting om na te gaan of zij robuust zijn voor gebruik, voordat breed gebruik kan worden geadviseerd.

www Literatuur: [www.kngf.nl/fysiopraxis](http://www.kngf.nl/fysiopraxis)



**Roel W. Wingbermühle**, SOMT Universiteit voor Fysiotherapie, Amersfoort; Erasmus MC, Rotterdam.

**Emiel van Trijffel**, SOMT Universiteit voor Fysiotherapie, Amersfoort; Vrije Universiteit Brussel.

**Paul M. Nelissen**, SOMT Universiteit voor Fysiotherapie, Amersfoort.

**Bart Koes**, Erasmus MC, Rotterdam.

**Arianne P. Verhagen**, Erasmus MC, Rotterdam; Technologische Universiteit Sydney, Australië.



[roelwingbermuehle@me.com](mailto:roelwingbermuehle@me.com)