

# VU Research Portal

## Methodological issues of clinical prediction models for shoulder pain in general practice

Vergouw, D.

2013

### **document version**

Publisher's PDF, also known as Version of record

[Link to publication in VU Research Portal](#)

### **citation for published version (APA)**

Vergouw, D. (2013). *Methodological issues of clinical prediction models for shoulder pain in general practice*. [PhD-Thesis - Research and graduation internal, Vrije Universiteit Amsterdam].

### **General rights**

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

### **Take down policy**

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

### **E-mail address:**

[vuresearchportal.ub@vu.nl](mailto:vuresearchportal.ub@vu.nl)

## **SAMENVATTING**

In de eerstelijns geneeskunde en de huisartsgeneeskunde in het bijzonder, worden medici vaak geconfronteerd met patiënten met bewegingsapparaat gerelateerde klachten. De behandeling van deze klachten blijkt echter vaak lastig, klachten houden lang aan of zijn wederkerig en duidelijke diagnostische criteria ontbreken. Om medici te helpen bij de inschatting van de ernst van de klachten heeft de wetenschap zich de laatste tijd gericht op predictie onderzoek. Hierin wordt getracht om factoren te identificeren welke een ongunstige uitkomst of beloop van de aanwezige klachten kunnen voorspellen. Deze factoren kunnen vervolgens gecombineerd worden tot een rekenregel ofwel een predictie model. Echter, het vinden van een eenvoudig predictie model dat voor alle patiënten consistente voorspellingen geeft blijkt erg lastig. Enkele methodologische kwesties zoals missnede waarden en model stabiliteit bemoeilijken het vinden van een dergelijk model. Met het in dit proefschrift beschreven onderzoek proberen we inzicht in deze versturende factoren te krijgen en proberen we uit te zoeken hoe we hun invloed kunnen minimaliseren. Hiertoe zullen verschillende analyse methoden in verschillende bewegingsapparaat gerelateerde data sets toegepast worden om zo meer kennis te vergaren over de optimale methode om tot een valide predictie model te komen.

In *hoofdstuk 1* beschrijven we de doelstellingen en het kader van dit proefschrift. Daarnaast bespreken we de epidemiologie van bewegingsapparaat gerelateerde klachten in het algemeen, geven we aan hoe predictie regels van toegevoegde waarde kunnen zijn bij de behandeling van bewegingsapparaat gerelateerde klachten in de eerstelijns geneeskunde en benadrukken we hoe methodologische kwesties zoals missende waarden en model stabiliteit de ontwikkeling van klinisch bruikbare predictie regels in de weg kan staan.

In *hoofdstuk 2* bieden we onderzoekers een empirische illustratie van het omgaan met missende waarden door deze te negeren en zodoende alle respondenten met missende waarden uit de data te laten (Complete Case Analyse, CCA) of door de waarde van de missende informatie te schatten (Multiple Imputatie, MI). Hiertoe gebruikten we data van het BeBack (Beliefs about Backpain) cohort, een studie naar de psychologische invloeden op het aanhouden van rugpijn onder 1591 eerstelijns patiënten met rugpijn in het Verenigd Koninkrijk. Deelnemers dienden voor elk meetpunt (voor- en nameting) apart toestemming te verlenen. Naast missende waarden in de voormeting (14%) leverde dit ook een behoorlijke uitval (51%) ten tijde van de nameting op. Wanneer de deelnemers met incomplete data vergeleken werden met de deelnemers waarvan wel alle informatie bekend was, viel het op dat er data selectief ontbrak. Deelnemers met complete data verschilden wat betreft de verdeling van voorspellers en uitkomst ten opzichten van deelnemers met incomplete data. Het gevolg hiervan was dat de predictie modellen ontwikkeld met alleen complete data of alle data (met invulling van de missende waarden) van elkaar verschilden wat betreft model samenstelling, voorspellend vermogen en overoptimisme. Hiermee zijn onze resultaten illustratief voor hoe CCA in het geval van missende waarden de ontwikkeling van een predictie model

en hoe MI deze verstoring minimaliseert. Daarom adviseren we in het geval van missende waarden multiple imputatie te gebruiken bij het ontwikkelen van een predictie model.

Bij het gebruik van multiple imputatie moet echter wel rekening gehouden worden met twee versturende factoren; 1) het tot één model combineren van de meerdere geïmputeerde data sets en 2) de extra instabiliteit geïntroduceerd door het herhaald schatten van de missende waarden. Met het oog op deze versturende factoren onderzochten we in *hoofdstuk 3* of een bootstrap model selectie procedure een waardevolle toevoeging aan multiple imputatie is. Hiertoe werd deze techniek toegepast in een data set afkomstig uit het Nederlands Schouder Onderzoek (NSO), een cohort bestaand uit 587 schouderpijn patiënten afkomstig uit de Nederlandse huisartsenpraktijk. Door in afzonderlijke stappen de sterke van de zwakke voorspellers te scheiden en vervolgens in meerdere bootstrap samples alle mogelijke combinaties van voorspellers te onderzoeken verkregen we een beeld van de robuustheid van het geselecteerde predictie model.

In *hoofdstuk 4* vergelijken we een van de meest gebruikte alternatieve technieken om tot een predictie model te komen; Classification And Ggression Tree (CART) analyse, met de doorgaans gebruikte logistische regressie methode. Om te bepalen welke van de twee beter geschikt is om een model te ontwikkelen pasten we beide technieken toe in de data van het Nederlands Schouder Onderzoek (NSO), een cohort bestaand uit 587 schouderpijn patiënten afkomstig uit de Nederlandse huisartsenpraktijk. We vergeleken de resultaten van beide methoden op verschillende belangrijke punten in de ontwikkeling van een predictie model. Het totale aantal geïnccludeerde voorspellers was voor beide modellen gelijk, echter de exacte model

samenstelling verschilde onderling. Het voorspellend vermogen in de derivatie data was vrijwel gelijk voor beide modellen (gelijke  $R^2_N$  van 19% en een AUC van 0.72 voor logistische regressie en 0.70 voor CART). Echter, bij toepassing in vergelijkbare patiënten (interne validatie) bleek het CART model slechter te presteren dan het logistisch regressie model. Op basis van deze bevindingen concluderen we dat ons logistisch regressie model beter geschikt is voor het voorspellen van aanhoudende schouderklachten dan het door ons ontwikkeld CART model.

In *hoofdstuk 5* trachtten we middels een groep van experts op het gebied van de behandeling van schouderklachten tot consensus te komen wat betreft de belangrijkste voorspellers van aanhoudende schouderpijn. Vervolgens testte we het voorspellend vermogen van deze klinische voorspellers. We stelden een internationaal (Verenigd Koninkrijk en Nederland) en multidisciplinair panel samen van 41 experts op het gebied van schouderklachten. In drie opeenvolgende Delphi rondes selecteerde het expert panel de volgende voorspellers voor aanhoudende schouderpijn; klachtenduur, catastroferen, klachten in het verleden, angst vermijding, bijkomende nekpijn, de ernst van bewegingshinder in de schouder, pijn op meerdere lichaamslocaties, leeftijd, de intensiteit van de schouderpijn en ziekteperceptie. Met deze voorspellers werden er, met gebruik van de NSO data, twee klinisch gebaseerde predictie modellen ontworpen. Eén waarin de informatie zoveel mogelijk als continue voorspellers werd opgenomen en één waarin het grootste gedeelte van de informatie als dichotome voorspellers werd opgenomen. Wanneer vergeleken met een eerder ontwikkeld statistisch model (*hoofdstuk 3*) bleek de prestatie van de klinisch gebaseerde modellen iets slechter (statistisch model  $AUC=0.702$ , klinisch dichotoom model  $AUC=0.656$  en klinisch continu model  $AUC=0.679$ ). Al verschilden de klinische modellen van het statistische model wat

betreft model samenstelling, we concluderen dat beide modellen bevestigen dat klachtenduur, de ernst van bewegingsbeperkingen van de schouder tijdens de voormeting en de intensiteit van de schouderpijn tijdens de voormeting belangrijke voorspellers voor aanhoudende schouderpijn zijn. Daarnaast concluderen we dat de verschillen in interne validiteit tussen de klinisch gebaseerde modellen en het statistische model bevestigd dienen te worden door beide modellen toe te passen op nieuwe schouderpatiënten.

In *hoofdstuk 6* onderzochten we de externe validiteit van de twee meest belovende door ons ontwikkelde modellen, het klinisch gebaseerde continue model uit *hoofdstuk 5* en het statistisch ontwikkelde MI-5 model uit *hoofdstuk 3*. Daarnaast vergeleken we het voorspellend vermogens van beide modellen met de prognose van de huisarts zelf welke gesteld werd tijdens het consult. In 23 huisartspraktijken rekruteerden we 203 patiënten met een nieuwe episode van schouderpijn. Deze patiënten vormden het SchouderPijn Prognose (SPP) cohort. In dit cohort bleken het statistische en klinisch gebaseerde model beiden suboptimaal te voorspellen met AUC waarden van respectievelijk 0.580 en 0.625. De inschatting van de toekomstige staat van schouderpijn van de 35 huisartsen in het onderzoek bleek met een AUC van 0.506 echter nog lager. Hiermee geven onze resultaten aan hoe moeilijk het is om het aanhouden van schouderklachten na drie maanden bij patiënten in de huisartsenpraktijk te voorspellen. Onze resultaten suggereren dat predictie modellen wellicht van toegevoegde waarde zouden kunnen zijn in de huisartsenpraktijk, echter de door ons ontwikkelde modellen bleken niet valide genoeg om een echte toevoeging te bieden.

In *hoofdstuk 7* plaatsen we de resultaten van dit onderzoek in een breder perspectief, sommen we de voornaamste conclusies van dit

proefschrift op en doen we aanbevelingen voor predictief onderzoek. In grote lijnen zouden onderzoekers op het gebied van predictie onderzoek altijd op de hoogte moeten zijn van de versturende invloed van missende waarden en model stabiliteit. De ontwikkeling van een predictie model is geen recht toe recht aan exercitie, in het proces dienen belangrijke keuzes gemaakt te worden en het effect van deze keuzes op de resultaten dienen geëvalueerd en gerapporteerd te worden.

