

## Nederlandse Samenvatting

Patiënten met laaggradige chronische niet-overdraagbare ziektes (zoals gelokaliseerde prostaatkanker en laaggradige dysplasie) krijgen vaak herhaalde invasieve testen (biopsieën, endoscopieën, enz.) om de progressie van ziekte te kunnen detecteren. Progressie gaat vaak gepaard met zware behandelingen met ernstige bijwerkingen, zoals chirurgie of radiotherapie. Om progressie te kunnen vaststellen worden invasieve testen vaak routinematig uitgevoerd volgens een vast (bijv. jaarlijks) schema dat voor alle patiënten gelijk is. Deze invasieve testen zijn belastend, maar tegelijkertijd noodzakelijk voor het tijdig detecteren van ziekteprogressie. Het frequent uitvoeren van invasieve testen kan leiden tot een versnelde detectie van ziekteprogressie, maar het is onnodig en belastend voor patiënten die slechts langzaam achteruitgaan. Daarentegen hebben teststrategieën waarbij niet frequent wordt getest, het nadeel dat de ziekteprogressie vertraagd wordt gedetecteerd. Het doel van dit proefschrift was om met geïndividualiseerde schema's een beter evenwicht te vinden tussen het aantal testen enerzijds en de vertraging in de detectie van progressie anderzijds, dan mogelijk zou zijn met vaste schema's. Hiervoor hebben we statistische methoden ontwikkeld en toegepast met als doel om invasieve diagnostische testen (bijvoorbeeld biopsieën, endoscopieën) op een gepersonaliseerde manier te kunnen plannen.

De eerste stap in het creëren van geïndividualiseerde testschema's was het ontwikkelen van een statistisch model. Dit model hebben we gebruikt om op basis van de verzamelde klinische data van een patiënt, het cumulatieve risico op progressie over de gehele follow-up periode te voorspellen. Zo'n risicoprofiel voorspelt de progressie van de ziekte van een patiënt over de tijd, van laaggradig naar gevorderd. Dit risicoprofiel gebruikten we vervolgens als leidraad om de timing van toekomstige invasieve testen te bepalen. De tweede stap was dus het creëren van een geïndividualiseerd testschema voor een patiënt op basis van zijn geschatte risicoprofiel. Voor dit doel hebben we 'doelfuncties' (bijvoorbeeld een kwadratische functie) van relevante klinische parameters (zoals de vertraging in detectie van ziekteprogressie) geoptimaliseerd. De parameters die we hiervoor hebben gebruikt

staan beschreven in Hoofdstukken 2, 3 en 4. In elk hoofdstuk hebben we de doelfunctie geoptimaliseerd ten opzichte van het geschatte cumulatieve risico op progressie, geïndividualiseerd voor de patiënt.

De gekozen doelfunctie en parameters heeft invloed op het testschema. In Hoofdstuk 2 bijvoorbeeld, hebben we gebruik gemaakt van drie standaard doelfuncties uit de Bayesiaanse besliskunde, namelijk 'squared loss', 'abslote loss', en 'multilinear loss'. Op basis van deze doelfuncties hebben we het tijdsverschil tussen de toekomstige test en het echte moment van progressie geoptimaliseerd. Het gebruik van 'squared' en 'absolute loss' resulteerde in testen op respectievelijk het gemiddelde en het mediane tijdstip van progressie, maar 'multilinear loss' resulteerde in het tijdstip waarop het voorspelde risico van een patiënt een bepaalde drempelwaarde heeft bereikt. Het gebruik van 'squared' en 'abslote loss' resulteert in het testen op het precieze moment van progressie, en leidt dus tot geen vertraging in het detecteren hiervan. Echter, door gebruik te maken van deze functies wordt geen rekening gehouden met de variantie van de 'posterior predictive distribution' (posterior voorspelverdeling) van het tijdstip van progressie van een patiënt. Het herhaaldelijk toepassen van deze functies tot het moment van progressie, betekent dat minder testen uitgevoerd worden maar dit kan ook leiden tot een vertraging in het detecteren van progressie. Daarentegen biedt het plannen van een test op het moment dat het risico van progressie een bepaalde drempelwaarde heeft bereikt, artsen de mogelijkheid om een afweging te maken tussen het onnodig uitvoeren van testen en een vertraging in het detecteren van progressie. Het kiezen van een lage drempelwaarde betekent bijvoorbeeld dat een patiënt ervoor kiest om veel testen te ondergaan en zo weinig mogelijk risico te lopen dat progressie te laat wordt gedetecteerd.

Als gebruik wordt gemaakt van een drempelwaarde, is het de vraag hoe een geschikte drempelwaarde gekozen dient te worden. Meestal worden drempelwaardes gekozen op basis van 'receiver operating characteristic curve' (ROC analyse) of op basis van hoe voor patiënten de last van het uitvoeren van een onnodige testen opweegt tegen het te laat detecteren van progressie. Om de keuze van een geschikte drempelwaarde verder te vergemakkelijken, hebben we in Hoofdstuk 3, een realistische gerandomiseerde, gecontroleerde

simulatiestudie uitgevoerd met verschillende drempelwaardes voor het actieve surveillancescenario voor prostaatkanker. Hoewel de resultaten van dit simulatieonderzoek alleen van toepassing zijn op het onderzoekscohort van onze dataset, zijn bepaalde resultaten generaliseerbaar naar alle ziektes. De belangrijkste uitkomst was dat drempelwaardes niet gekozen moeten worden op basis van maatstaven voor diagnostische nauwkeurigheid, zoals Youden's J of de F1 score, aangezien het voor dergelijke maatstaven niet mogelijk is de gevoeligheid en specificiteit van een drempelwaarde te controleren.

In Hoofdstuk 2 en Hoofdstuk 3 hebben we maar één toekomstige test tegelijk gepland. In Hoofdstuk 4 hebben we onze methode uitgebreid zodat een volledig testschema in één keer gepland kan worden. Om patiënten en artsen te assisteren bij het vinden van een optimaal schema, hebben we het verwachte aantal testen en de verwachte vertraging in het detecteren van progressie voor bepaalde testschema's berekend, beide geïndividualiseerd voor de specifieke patiënt. Onze beweegredenen voor de keuze voor deze criteria zijn als volgt. Ten eerste stellen we dat vertraging in het detecteren van progressie een makkelijk te kwantificeren surrogaatuitkomst is voor belangrijke klinische aspecten zoals de kans op curatieve behandeling, het risico op nadelige 'downstream' uitkomsten, de voor kwaliteit gecorrigeerde resterende levensduur en aanvullende complicaties bij de behandeling van een progressie die te laat wordt gedetecteerd. Ten tweede zijn de criteria (het aantal testen en de vertraging in het detecteren van progressie) voor zowel patiënten als artsen gemakkelijk te begrijpen, en ze kunnen een tussen arts en patiënt gedeelde besluitvorming van testschema's faciliteren.

Een geïndividualiseerd schema is slechts zo goed als de voorspellende waarde van het onderliggende statistische model. Om deze voorspellende waarde te bepalen, hebben we het model dat we hebben voorgesteld voor actieve surveillance van prostaatkanker extern gevalideerd in de grootste zes cohorten van de Movember Foundations Global Action Plan (GAP3) database (Hoofdstuk 5). We hebben voor elk cohort de tijdsafhankelijke gemiddelde absolute voorspellingsfout en de tijdsafhankelijke oppervlakte onder de receiver operating characteristic (AUC) curve berekend. Ons model had een redelijk lage voorspellingsfout en AUC. We hebben ook geïndividualiseerde

biopsieschema's geïmplementeerd voor echte patiënten van de gevalideerde cohorten in een webapplicatie ([https://emcbiostatistics.shinyapps.io/prias\\_biopsy\\_recommender/](https://emcbiostatistics.shinyapps.io/prias_biopsy_recommender/)). Het gebruik van geïndividualiseerde schema's is echter niet beperkt tot invasieve testen. We hebben dit aangetoond door het plannen van metingen van de biomarker NT-proBNP (een eenvoudige bloedtest) voor patiënten met chronisch hartfalen op basis van een bestaande geïndividualiseerde methode (Hoofdstuk 6).