



Universiteit
Leiden

The Netherlands

The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling

Luijken, K.

Citation

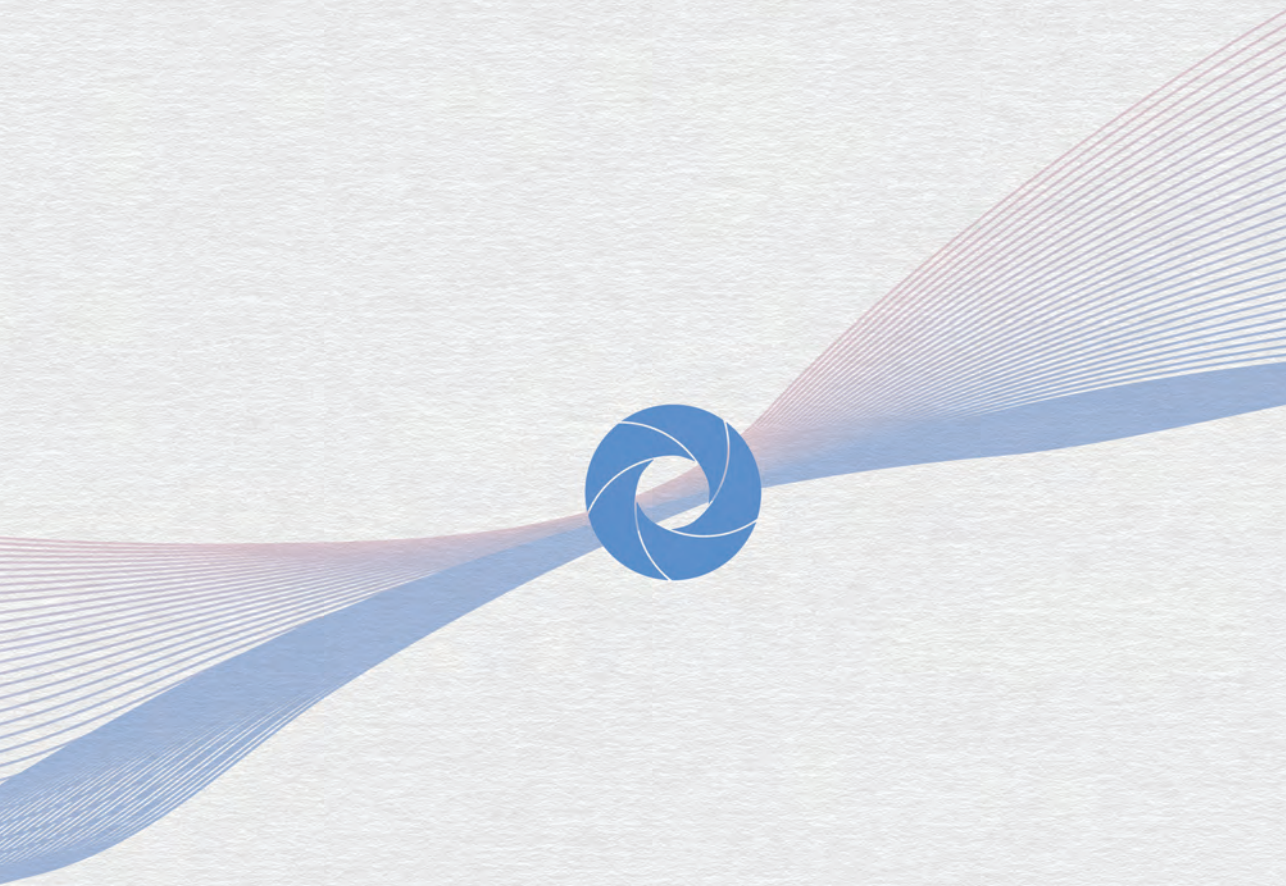
Luijken, K. (2022, May 19). *The impact of epidemiologic methods on findings in studies of causal effects and prediction modelling*. Retrieved from <https://hdl.handle.net/1887/3304345>

Version: Publisher's Version

License: [Licence agreement concerning inclusion of doctoral thesis in the Institutional Repository of the University of Leiden](#)

Downloaded from: <https://hdl.handle.net/1887/3304345>

Note: To cite this publication please use the final published version (if applicable).



Samenvatting in het Nederlands

In de afgelopen decennia, met name in de laatste jaren, zijn epidemiologische methoden steeds verder verfijnd, waardoor het een uitdaging is om op de hoogte te blijven van alle methodologische ontwikkelingen. De keuze voor een data-analysemethode beïnvloedt de interpretatie en klinische betekenis van resultaten direct en het is ongewenst dat het onderwerp van een onderzoek wordt bepaald door technische overwegingen. Een beter begrip van de invloed die beslissingen ten aanzien van de data-analyse kunnen hebben op de interpretatie van numerieke onderzoeksresultaten kan helpen om te komen tot de inzet van analytische instrumentaria die zowel passend als valide zijn om klinische vragen te beantwoorden. Het doel van dit proefschrift is om de invloed van keuzes omtrent het ontwerp en de statistische analyse van een studie op de betekenis van de numerieke resultaten te onderzoeken, met specifiek aandacht voor onderzoek naar causale effecten (Deel I) en voorspellingsonderzoek (Deel II).

Deel I: Invloed van toegepaste methoden op de betekenis van numerieke schattingen in onderzoek naar causale effecten

De studies beschreven in Deel I van dit proefschrift zijn voorbeelden van de invloed van keuzes omtrent het studie ontwerp en de statistische analyse op numerieke resultaten in onderzoeken naar causale effecten. Dit deel schetst de vele data-analytische beslissingen die in zulke onderzoeken dienen te worden genomen. In de hoofdstukken 2 en 4 worden twee specifieke beslissingen uitgelicht en is de conclusie dat de interpretatie van effectschattingen in onderzoeken naar causale effecten in belangrijke mate afhangt van de keuze voor respectievelijk de studietijdoorsprong en de strategie ten aanzien van selectie van covariabelen. De resultaten van deze onderzoeken impliceren dat het belangrijk is om een tegengesteld proces te vermijden, waarbij de toegepaste onderzoeksopzet en statistische analyse impliciet de betekenis van de onderzoeksresultaten bepalen. In plaats daarvan zou juist eerst het doel van het onderzoek duidelijk moeten zijn, worden vastgesteld met welke onderzoeksresultaten die doelen kunnen worden bereikt om pas daarna te bepalen welke data-analyse daarvoor geschikt is. Dit wordt benadrukt door de onderzoeken beschreven in de hoofdstukken 3 en 5, waarin wordt besproken hoe een gericht geformuleerd onderzoeksdoel kan verduidelijken of de klinische doelen, de onderzoeksopzet en de

interpretatie van resultaten op de juiste manier op elkaar zijn afgestemd. Het vaststellen van het onderzoeksdoel is echter een uitdaging in klinische onderzoeken. In hoofdstuk 2 wordt deze uitdaging weerspiegeld in de lage frequentie van expliciet gerapporteerde *estimands*, ofwel de 'te schatten waarde die de klinische onderzoeksvraag waarin men geïnteresseerd is beantwoordt (of het beste benadert)', in onderzoeken naar effecten van farmacologische behandelingen.

Hoofdstuk 2 beschrijft dat de invloed van de gekozen studietijdsoorsprong op numerieke resultaten moeilijk te beoordelen was in farmaco-epidemiologische studies vanwege onvolledige rapportage. De rapportage over keuzes met betrekking tot het operationaliseren van de tijdsoorsprong van een studieontwerp werd onderzocht in een systematische review van 89 vergelijkende cohortstudies voor effectiviteit en veiligheid, gepubliceerd in zes hoog aangeschreven farmaco-epidemiologische tijdschriften in 2018 en 2019. Veertig procent van de onderzoeken rapporteerde de studie inclusie te beperken tot alleen nieuwe gebruikers (d.w.z., een 'incidentgebruikersontwerp') en 13% rapporteerde ook bestaande gebruikers te includeren (d.w.z., een 'prevalentgebruikersontwerp'). De start van opvolging binnen de studie, het moment waarop aan de toelatingscriteria werd voldaan en de start van de behandeling bleken overeen te komen in 53% van de onderzoeken met een incidentgebruikersontwerp (19 van de 36 onderzoeken) en werd in 42% van de onderzoeken met een incidentgebruikersontwerp onvoldoende beschreven. De validiteit van de operationalisering van de studietijdsoorsprong kon alleen worden beoordeeld met betrekking tot de *estimand*. De *estimand* werd echter in slechts 22% van de onderzoeken met een incidentgebruikersontwerp expliciet gerapporteerd.

In **Hoofdstuk 3** wordt de nauwkeurigheid waarmee het onderzoeksdoel wordt gedefinieerd in explorerende etiologische studies gekoppeld aan de interpretatie van de bevindingen van dergelijke studies. Er wordt een continuüm beschreven van kritische evaluatie voor het uitvoeren van studies, variërend van *ad-hoc* tot *doordacht*. De positie waarop een explorerende etiologische analyse zich op dit continuüm bevindt, beïnvloedt direct de interpreteerbaarheid van bevindingen. Er wordt beargumenteerd dat het handelen naar resultaten van *ad-hoc* analyses alsof ze voortkomen uit *doordachte* analyses, bijvoorbeeld door verdere bevestigende studies uit te voeren of door ze in de klinische praktijk te implementeren, kan bijdragen aan verspilling van onderzoeksmiddelen en patiënten mogelijk kan schaden. Er worden praktische

aanwijzingen gegeven voor een goede uitvoering van explorerend etiologisch onderzoek, zoals het gebruik van rigoureuze methodologische en statistische benaderingen en het nemen van verantwoordelijkheid voor exploratieve bevindingen door een duidelijke agenda voor toekomstig onderzoek te rapporteren.

De studie beschreven in **Hoofdstuk 4** illustreert dat het toepassen van achterwaartse eliminatie om het aantal covariabelen waarmee voor vertroebeling (*confounding*) gecorrigeerd wordt te verminderen zelden efficiënter is dan selectie van covariabelen op basis van causale kennis. Er wordt een uitdrukking afgeleid die kwantificeert hoe het weglaten van een variabele zich verhoudt tot vertekening (*bias*) en variantie van effectschatters. Er zijn simulaties uitgevoerd om te onderzoeken of en onder welke omstandigheden de schatting van causale effecten in observationele onderzoeken kan worden verbeterd door achterwaartse eliminatie te gebruiken op een vooraf gespecificeerde set van potentiële *confounders*. Het toepassen van achterwaartse eliminatie verminderde de gemiddelde fout van effectschatters niet in vergelijking met een volledig model inclusief alle vooraf gespecificeerde covariabelen, maar de *bias* was toegenomen. In minder dan 3% van de 3960 beschouwde scenario's was de gemiddelde fout van effectschatters lager wanneer achterwaartse eliminatie werd gebruikt in vergelijking met het volledige model. Deze bevindingen geven daarmee aan dat wanneer een eerste set van potentiële *confounders* kan worden gespecificeerd op basis van achtergrondkennis, er een minimale toegevoegde waarde is van achterwaartse eliminatie.

In **Hoofdstuk 5** wordt een methode voorgesteld om studies die in systematische reviews van chirurgische interventies worden geïncludeerd te beoordelen met betrekking tot keuzes omtrent het ontwerp en de statistische analyse. Dit is bedoeld als een eerste stap richting het samenvatten van de belangrijkste informatie die nodig is om de toepasbaarheid en methodologische kwaliteit van studies te beoordelen. Daartoe is – op basis van bestaande instrumenten die het risico op bias beoordelen – een beknopte set van items gekozen voor de initiële evaluatie van studies van chirurgische interventies. De set bevat negen items: populatie, interventie, comparator, uitkomst, *confounding*, ontbrekende gegevens en selectiebias, interventiestatus, uitkomstbeoordeling en pre-specificatie van de analyse. De beoordeling van toepasbaarheid en methodologische kwaliteit kan worden gedaan als onderdeel van een systematische review om een eerste schifting te maken waarin studies van lage kwaliteit relatief gemakkelijk kunnen worden uitgesloten. Studies van hogere kwaliteit kunnen aan verdere beoordeling

van methodologische kwaliteit worden onderworpen met behulp van bestaande beoordelingsinstrumenten.

Deel II: Invloed van toegepaste methoden op de betekenis van numerieke schattingen in voorspellingsonderzoek

De studies beschreven in Deel II van dit proefschrift zijn gericht op voorspellingsonderzoek en bestuderen de invloed van veranderingen in strategieën voor het meten van voorspellers tussen verschillende modelleringsfasen op de mate waarin voorspelmodellen correcte voorspellingen doen (d.w.z., de ‘modelprestaties’). Dergelijke veranderingen worden *voorspeller-meetheterogeniteit* genoemd. Het fenomeen *voorspeller-meetheterogeniteit* werd in dit proefschrift formeel gedefinieerd met behulp van meetfoutmodellen. Hierdoor was het mogelijk de implicaties van *voorspeller-meetheterogeniteit* te onderzoeken door middel van analytische benaderingen, simulaties, analyse van empirische datasets en een voorgestelde kwantitatieve voorspellingsanalyse. Al deze analyses geven aan dat zelfs wanneer alle andere factoren, zoals de modelleringsstrategie, uitkomstprevalentie, voorspellers en patiëntkenmerken, constant worden gehouden in alle modelleringsfasen, een verandering in de meetprocedure de modelprestaties beïnvloedt. Dit nodigt uit tot een heroverweging van de manier waarop voorspellingsmodellen worden gespecificeerd, en met name of voorspellermeetprocedures een integraal onderdeel van de modelspecificatie zouden moeten zijn.

Hoofdstuk 6 beschrijft hoe strategieën voor het meten van voorspellers relateren aan klinische toepasbaarheid van voorspellingen van binaire logistische voorspellingsmodellen met behulp van analytische- en simulatiebenaderingen. Een gevestigde taxonomie van meetfoutmodellen wordt gebruikt om het fenomeen *voorspeller-meetheterogeniteit* te definiëren en te verduidelijken. Voorspeller-meetheterogeniteit refereert naar variatie in de meetprocedure van voorspeller(s) tussen de data waarin een voorspellingsmodel wordt afgeleid en waarin een model wordt gevalideerd of geïmplementeerd. Met behulp van analytische en simulatiebenaderingen wordt aangetoond dat modelprestaties buiten de steekproef van binaire logistische voorspellingsmodellen kunnen worden belemmerd wanneer voorspellers anders worden gemeten bij afleiding en externe validatie. Deze bevindingen benadrukken dat het onvoldoende is om een voorspellingsdoel in algemene termen te beschrijven zonder de procedures te specificeren waarmee voorspellers (moeten) worden gemeten.

In **Hoofdstuk 7** wordt getoond hoe strategieën voor het meten van voorspellers relateren aan klinische toepasbaarheid van voorspellingen van binaire logistische diagnostische modellen met behulp van empirische illustraties in drie klinische datasets. Negen scenario's van *voorspeller-meetheterogeniteit* worden geëvalueerd in eerder ontwikkelde voorspellings-modellen voor de diagnose van eierstokkanker, mutatiedragers voor Lynch-syndroom en intra-uteriene zwangerschap. Het wijzigen van de meetprocedure van een voorspeller had invloed op de modelprestaties bij validatie van de diagnostische modellen, met name modelkalibratie, waarbij de coëfficiënt voor gemiddelde kalibratie bij validatie varieerde van -0,70 tot 1,43 en de kalibratiehelling van 0,50 tot 1,67.

Hoofdstuk 8 beschrijft een kwantitatieve voorspellingsanalyse om te anticiperen op de invloed van veranderingen in voorspellermeetstrategieën voor prognostische tijd-tot-gebeurtenis uitkomstmodellen. Met behulp van simulaties met verschillende scenario's van *voorspeller-meetheterogeniteit*, hebben we aangetoond dat modelprestaties buiten de steekproef kunnen worden belemmerd wanneer voorspellers anders worden gemeten bij validatie en implementatie voor tijd-tot-gebeurtenis uitkomstmodellen. Er wordt een kwantitatieve voorspellingsanalyse voorgesteld om de invloed te kwantificeren van de verwachte *voorspeller-meetheterogeniteit* over de validatie- en implementatiesetting heen.

Tot slot

Dit proefschrift beschrijft de invloed van verschillende keuzes omtrent het ontwerp en de statistische analyse van een studie op de betekenis van de numerieke resultaten. Het duidelijk definiëren van een klinisch relevante *estimand* zorgt ervoor dat data-analytische beslissingen zinvolle resultaten opleveren. Echter, het vereist een diepgaand begrip van statistische theorie om een *estimand* te definiëren en hiervoor zijn weinig aanbevelingen beschikbaar, met name niet voor niet-therapeutisch onderzoek.

In **Hoofdstuk 9**, de algemene discussie van dit proefschrift, stellen we voor hoe dergelijke aanwijzingen zouden kunnen worden ontwikkeld om onderzoekers te helpen tot een voldoende goed gedefinieerde *estimand* te komen, te beginnen vanuit een klinisch perspectief, dat wil zeggen door het definiëren van een '*doordachte onderzoeksvraag*'. Vijf onderwerpen voor toekomstig onderzoek worden voorgesteld die kunnen bijdragen aan het centraal stellen van *doordachte onderzoeksvragen* in klinisch onderzoek, te weten:

- (1) Onderscheid een klinisch perspectief van een statistisch perspectief met als doel de input van beide op elkaar af te stemmen.
- (2) Identificeer de optimale balans tussen beknoptheid en volledigheid van *doordachte onderzoeksvragen*.
- (3) Specificeer hoe en wanneer *doordachte onderzoeksvragen* verschillen tussen onderzoeken naar causale effecten en voorspellingsonderzoek.
- (4) Voer empirische voorbeeld-studies uit om te begrijpen hoe gerichte onderzoeksvragen gesteld kunnen worden.
- (5) Geef les in het stellen van *doordachte onderzoeksvragen*.

Door *doordachte onderzoeksvragen* centraal te stellen in kwantitatief klinisch onderzoek kan voorbarig vertrouwen in (complexe) methoden worden verminderd en kan het begrip van de waarde en betekenis van onderzoeksresultaten worden vergroot.