



Universiteit
Leiden
The Netherlands

Safe anytime-valid inference: from theory to implementation in psychiatry research

Turner, R.J.

Citation

Turner, R. J. (2023, November 14). *Safe anytime-valid inference: from theory to implementation in psychiatry research*. Retrieved from <https://hdl.handle.net/1887/3663083>

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/1887/3663083>

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

Nederlandse Samenvatting

Dit proefschrift, *veilige altijd valide inferentie: van theorie naar implementaties voor wetenschappelijk onderzoek in de psychiatrie*, gaat over het doorontwikkelen van een nieuw paradigma in de statistiek, veilige altijd valide inferentie (SAVI), en het toewerken naar een toepassing hiervan in wetenschappelijk onderzoek in de psychiatrie. De gangbare, meest gebruikte methoden voor statistische inferentie (het doen van uitspraken over de gehele populatie op basis van een kleine steekproef) zijn niet geschikt voor flexibele onderzoeksopzetten. Voorbeelden van dit soort opzetten zijn continue analyse van onderzoeksdata en projecten waarbij algoritmes gedeeld worden tussen verschillende instanties. In de eerste twee hoofdstukken van dit proefschrift wordt onderzocht hoe SAVI doorontwikkeld kan worden voor enkele veel voorkomende onderzoeksvragen in dit soort flexibele onderzoeksopzetten. In de drie daaropvolgende hoofdstukken wordt onderzocht hoe data verzameld tijdens psychiatrische zorg geschikt kan worden gemaakt voor continue analyse in meerdere ziekenhuizen, en wordt exploratief gezocht naar patronen in deze retrospectieve data met Bayesiaanse netwerkanalyse. Tot slot worden in het afsluitende hoofdstuk de nieuwe statistische methoden verder ontwikkeld voor het beantwoorden van onderzoeksvragen met een vorm zoals de vragen die volgden uit de Bayesiaanse netwerkanalyse, en wordt geïllustreerd hoe een flexibele onderzoeksopzet om een van deze vragen te beantwoorden er uit zou kunnen zien.

Leren in meerdere instellingen Data verzameld tijdens routinematige zorg en/ of medisch onderzoek blijft nu meestal exclusief binnen de zorginstellingen waar de zorg is verleend. Dit is goed voor de privacy van patiënten, maar maakt het lastig om te *leren* en om *patronen* te ontdekken in de data, om later de zorg te verbeteren. Vooral als het doel is te leren voor kleine groepen patiënten, om uiteindelijk *gepersonaliseerde aanbevelingen* te kunnen doen, is data van grote aantallen patiënten nodig voor algoritmes om patronen te ontdekken en bevestigen. Om leren van data bij verschillende zorginstellingen mogelijk te maken terwijl ook de privacy van patiënten gewaarborgd blijft is het *Enabling Personalized Interventions* consortium (Nederlandse vertaling van de naam: faciliteren van gepersonaliseerde behandelingen) opgericht, waar het werk verricht voor dit proefschrift onderdeel van uitmaakt. In dit consortium wordt gewerkt aan een kader en softwarepakket waarbinnen ten eerste patiëntdata bij zorginstellingen op een veilige manier bereikt kan worden, ten tweede regelgevende beperkingen opgelegd kunnen worden op basis van toestemming die zorginstellingen en indi-

viduele patiënten hebben gegeven en ten derde algoritmes worden ontwikkeld die gefedereerd kunnen leren. Dit houdt in dat als meerdere zorginstellingen samen een algoritme willen trainen de data niet verzameld hoeft te worden op een centraal punt. In plaats daarvan wordt het algoritme naar de veilige omgevingen in de aparte instellingen gestuurd, waarbij de (niet-herleidbare) getrainde algoritmes worden teruggestuurd en gecombineerd op een centraal punt. In dit proefschrift worden methoden ontwikkeld en onderzocht die bij uitstek geschikt zijn voor dit soort toepassingen, waarbij de data verdeeld zijn over meerdere instellingen.

Datastromen Het bovengenoemde SAVI, een centraal concept in dit proefschrift, is een van deze methoden. Resultaten van SAVI-analyse van verschillende zorginstellingen kunnen met elkaar gecombineerd worden door simpele verenigingsvuldiging. Daarnaast is SAVI-analyse ook geschikt voor *continue analyse van datastromen*; het biedt de mogelijkheid data opnieuw te analyseren na de inclusie van iedere nieuwe patiënt. Na iedere nieuwe analyse kan een *veilige* beslissing genomen worden over de uitkomst van het onderzoek. Dit houdt in dat de methode een zogenoemde type-I fout garantie geeft op ieder moment: de kans dat we onterecht de nulhypothese van het onderzoek verwerpen (bijvoorbeeld de kans dat we onterecht besluiten dat een nieuw middel superieur is vergeleken met een placebo) is begrensd door een door de onderzoeker ingestelde maximaal acceptabele kans op een fout, op welk moment we de beslissing ook nemen. Dit is iets dat met de traditioneel toetsen waarbij bijna altijd gebruik wordt gemaakt van p-waardes niet kan: hierbij moet van tevoren vastgesteld worden met hoeveel patiënten de analyse verricht zal worden. Traditionele p-waarde toetsen zijn dus niet geschikt voor continue analyse van onderzoeksdata. In hoofdstuk 2 van dit proefschrift worden algemene SAVI toetsen voor het vergelijken van twee of meer datastromen ontwikkeld, en wordt een specifieke implementatie met bijbehorende software voor binaire datastromen ontwikkeld. Deze wordt ook vergeleken met de klassieke p-waarde tegenhanger, Fishers exacte toets.

Effecten schatten in datastromen Naast het uitvoeren van hypothesetoetsen zoals “is het nieuwe middel beter dan een placebo” is het voor het nemen van beslissingen essentieel een goede effectschatting te hebben, om bijvoorbeeld te kunnen zeggen hoeveel winst een patiënt zou kunnen verwachten bij gebruik van het nieuwe middel. Traditionele manieren om effecten te schatten, zoals klassieke betrouwbaarheidsintervallen, hebben last van dezelfde beperkingen als p-waardes: deze zijn alleen valide als ze gebruikt worden met een van tevoren vastgestelde hoeveelheid data. In hoofdstuk 3 worden de in hoofdstuk 2 ontwikkelde SAVI toetsen uitgebreid en gebruikt als bouwsteen voor het maken van *altijd valide betrouwbaarheidsintervallen*. Deze altijd valide betrouwbaarheidsintervallen zouden bijvoorbeeld kunnen worden gebruikt in onderzoeksdashboards, waarbij datastromen in werkelijke tijd geanalyseerd worden en de effectschatting direct door onderzoekers en klinici gevolgd kan worden. Hierbij kunnen de onderzoekers op ieder moment beslissen dat de effectschatting precies genoeg is om de studie te kunnen stoppen en beleid aan te passen.

Leren van vrije tekst Voordat flexibele onderzoeken met als doel inferentie zoals hierboven beschreven kunnen worden opgezet, is het onmisbaar dat onderzoeksdata in het juiste format ontsloten wordt en geëxploreerd wordt voor hypothesevorming. Binnen de psychiatrie wordt veel informatie opgeslagen in de vorm van vrije tekst, wat zowel exploratief onderzoek als statistische inferentie bemoeilijkt. Tegelijkertijd is het onmogelijk van klinici te verwachten dat zij alle gegevens in tabulaire vorm in het elektronisch patiëntdossier registreren. Om deze redenen spelen *tekstmining* en informatie extractie een steeds belangrijkere rol in de exploratieve fase van medisch onderzoek. In hoofdstuk 4 wordt onderzocht of met deze technieken informatie over behandeluitkomsten uit routinematig geschreven klinische teksten geëxtraheerd kan worden, om later patronen te kunnen herkennen in de combinatie van deze uitkomsten, patiëntkarakteristieken en behandelkeuzes.

Netwerkanalyse Een methode voor het ontdekken van patronen bij uitstek passend bij de complexiteit van psychiatrische pathologie is Bayesiaanse netwerk-analyse. Met deze analysemethode kunnen causale of voorspellende verbanden tussen patiëntkarakteristieken, behandelkeuzes en een of meerdere uitkomsten ontdekt worden. De uitkomst van de analyse kan inzichtelijk gevisualiseerd worden in netwerkvorm. In hoofdstuk 5 van dit proefschrift wordt Bayesiaanse netwerk-analyse gecombineerd met de tekstmining methode ontwikkeld in hoofdstuk 4 om patronen te ontdekken en hypothesen te ontwikkelen over de uitkomsten na behandeling met verschillende typen antidepressiva op basis van routinematig verzamelde klinische data in twee instituten voor mentale gezondheidszorg. Bayesiaanse netwerken bieden ook de mogelijkheid voor het integreren van *voorkennis*: in hoofdstuk 6 wordt voorkennis uit medische literatuur gecombineerd met expertkennis van psychiaters om een Bayesiaans netwerk voor het modelleren van uitkomsten na behandeling met electroconvulsietherapie te verbeteren.

Leren voor specifieke groepen patiënten Het onderzoek in hoofdstuk 5 en 6 bracht meerdere kansrijke hypothesen over behandel-effecten bij specifieke, kleine groepen patiënten voort, die verdere validatie middels prospectief onderzoek behoeven om een robuuste effectschatting te kunnen maken. In hoofdstuk 7 worden de methodes ontwikkeld in hoofdstuk 2 en 3 uitgebreid voor onderzoeksvragen van deze vorm, waarbij behandel-effecten in verschillende datastromen prospectief waren geanalyseerd en ook nog rekening gehouden wordt met het type patiënt: de data wordt *gestratificeerd* naar het type patiënt. Om de mogelijkheid om behandel-effecten te detecteren met deze onderzoeksmethode te verbeteren, wordt het effect van informatieuitwisseling tussen de patiëntgroepen op een zelflerende manier onderzocht. Uiteindelijk wordt geïllustreerd hoe een onderzoeksopzet om een van de resulterende hypothesen uit hoofdstuk 5 er uit zou kunnen zien.

Makkelijker valide medisch onderzoek In dit proefschrift zijn nieuwe methoden ontwikkeld om medisch onderzoek op een meer flexibele en makkelijkere manier op te zetten: met deze nieuwe methoden is het mogelijk om onderzoeksresultaten in werkelijke tijd, al tijdens de studie te analyseren en eventueel eerder te beslissen of langer door te gaan als meer informatie verzameld moet worden. Ook zijn deze

methoden direct uit te breiden naar continue analyse over meerdere onderzoeksinstellingen. Continuering van het werk van het Enabling Personalized Interventions consortium en soortgelijke initiatieven in de komende jaren zal hopelijk de complexe data-infrastructuur die nodig is voor dit soort flexibele studies met meerdere deelnemende ziekenhuizen mogelijk maken, zodat daadwerkelijk toegewerkt kan gaan worden naar gepersonaliseerde behandelingen door valide effectschattingen in specifieke groepen patiënten.