Evidence synthesis methods for continuous outcomes

Papadimitropoulou, A.

Nederlandse Samenvatting

Meta-analyse is de methodologie die wordt gebruikt om bewijs uit verschillende bronnen te synthetiseren om: a) effecten te detecteren, b) hun grootte en variabiliteit te schatten, c) de factoren (covariaten of moderatoren) te analyseren die hierop van invloed zijn. Hoewel de term in de jaren zeventig door sociale wetenschappers werd bedacht, vindt de bredere term 'onderzoek synthese' zijn oorsprong in de in 1885 presidentiële toespraak van natuurkundige Lord Rayleigh, waarbij de nadruk werd gelegd op de noodzaak van "niet alleen het verkrijgen van nieuw materiaal", maar ook op "het integreren en het begrijpen van het oude" (Rayleigh, 1885). Meer dan een eeuw later levert een algemene Google-zoekopdracht naar "meta-analyse" meer dan 600 miljoen resultaten op, terwijl het zoeken naar 'meta-analyse' in PubMed meer dan 120000 artikelen oplevert. Men kan stellen dat de aanbevelingen van Rayleigh worden opgevolgd.

De introductie van formele bewijs-synthese methoden heeft geleid tot een verschuiving in het perspectief van wetenschappers op de resultaten van hun onderzoek. De output van een enkele primaire studies kan nu worden gezien als een onderdeel van het antwoord op een wetenschappelijke vraag in plaats van als de sluitende waarheid.

Bovendien heeft de introductie van deze methoden bijgedragen aan de verandering binnen de wetenschappelijke gemeenschap naar een meer open gemeenschap met hogere standaarden voor het rapporteren van resultaten. Het bekende probleem dat vooral significante resultaten worden gepubliceerd wordt ook aangepakt, nu onderzoekers zich realiseren dat niet-significante resultaten mogelijk net zo waardevol zijn (Button e.a., 2013).

Dit proefschrift

Dit proefschrift beschrijft nieuwe statistische methodologie en nieuwe toepassingen op het gebied van bewijs-synthese van vergelijkbare studies met continue uitkomstmaten. Speciale aandacht wordt gevestigd op de analyse van continue geaggregeerde data door het generen van pseudo individuele patiënten data (pseudo-IPD). Op deze manier kunnen voor de her-analyse zogenoemde lineaire mixed modellen (LMM) gebruikt worden. Deze nieuwe aanpak maakt gebruik van de hiërarchische aard van de meta-analytische gegevens en het grote aanbod aan statistische methoden die beschikbaar zijn voor het gebruik van mixed models. Daarnaast doen we op basis van empirische gegevens aanbevelingen voor het uitvoeren van meta-analyse op continue uitkomstmaten, die voor en na een behandeling zijn gemeten.

Dit proefschrift behandelt ook de kwestie van ontbrekende uitkomst data in (netwerk) meta-analyse, waarbij een nieuwe methodologie op basis van een zogenoemd pattern mixture model wordt gebruikt. Hieronder geven we een gedetailleerde beschrijving van elk hoofdstuk.

Hoofdstuk 1 geeft een algemene inleiding van bestaande bewijs-synthese methoden. We beginnen met het benadrukken van deze methoden en de brede acceptatie ervan in verschillende disciplines. De rest van de inleiding is opgesplitst in twee delen: de meta-analyse en netwerk-meta-analyse methoden.

We introduceren het concept van meta-analyse en de verschillende typen ervan op basis van de gestelde onderzoeksvraag. Daarnaast bespreken we kort de belangrijkste datastructuren voor meta-analyse: de geaggregeerde data, welke de openbaar beschikbare resultaten zijn van de uitkomsten op studieniveau, en de IPD, de onbewerkte data op deelnemersniveau. We introduceren de relevante terminologie en de twee fundamentele statistische modellen: de common-effect- en de random-effects-modellen. Het verschil tussen deze twee typen modellen ligt in de aanname dat de werkelijke onderliggende effecten van de verschillende studies identiek zijn (common) of afkomstig zijn uit een steekproef van mogelijke werkelijke effecten. Een belangrijk punt dat genoemd moet worden, is dat het testen op heterogeniteit niet de drijfveer mag zijn voor de beslissing tussen de twee type modellen. De beslissing moet uitsluitend gebaseerd zijn op de vraag welk model past bij de verdeling van de waargenomen effectgroottes. Meestal is het random-effects-model het meest geschikt. We bespreken kort de voordelen van het uitvoeren van een meta-analyse met IPD in vergelijking met geaggregeerde data. Hoewel meta-analyse op IPD, meer moeite kost, heeft dit om verschillende redenen de voorkeur boven het gebruik van de geaggregeerde data, mits dit mogelijk is.

In het tweede deel van de inleiding beschrijven we de overgang van meta-analyse naar technieken die de vergelijking mogelijk maken tussen meerdere behandelingen die niet direct zijn vergeleken in een studie, d.w.z. indirecte behandelingsvergelijkingen en netwerk-meta-analyse. Deze technieken voor het synthetiseren van bewijsmateriaal zijn de laatste 15 jaar aanzienlijk verbeterd. We leggen de onderliggende aannames van transitiviteit en consistentie uit en beschrijven een Bayesiaanse netwerk-meta-analyse aanpak.

Hoofdstuk 2 presenteert een nieuwe methode voor de meta-analyse van een enkele continue uitkomstmaat, gemeten in twee groepen. Onze voorgestelde methode is gebaseerd op de eigenschap van *voldoende steekproeffuncties*, dit zijn de statistieken die alle essentiële informatie voor de modelparameter die van belang is bevatten. In het geval van een meta-analyse van normaal verdeelde uitkomsten zijn de voldoende functies de gemiddelden, standaarddeviaties en de steekproefomvang en deze gegevens zijn vaak beschikbaar in gepubliceerde onderzoeksresultaten. We kunnen dan pseudo-IPD generen met dezelfde gemiddelden, standaarddeviaties en steekproefgroottes. Als we LMM gebruiken om deze pseudo-IPD te analyseren, zal de zogenoemde likelihood identiek zijn aan de likelihood van de oorspronkelijke echte IPD. Naast het algoritme om dergelijke pseudo-IPD te genereren, presenteren we de scripts voor verschillende statistische software programmas.

Zodra de pseudo-IPD zijn geconstrueerd, kunnen ze op dezelfde eentraps manier geanalyseerd worden als de echte IPD, namelijk met LMM. Een overvloed aan modelleeropties is beschikbaar en deze worden uitgebreid beschreven in **hoofdstuk 2**. Het gebruik van LMM biedt aanzienlijke voordelen ten opzichte van de traditioneel gebruikte methoden voor geaggregeerde data. Het staat bijvoorbeeld complexe/-meer realistische aannames voor de residuele varianties binnen de studie toe. We

illustreren onze voorgestelde methode aan de hand van twee echte data voorbeelden afkomstig uit een meta-analyse waarin gemiddelde plasmaspiegels tussen Alzheimer patiënten en gezonde controles worden vergeleken. In totaal worden 12 IPD modellen van toenemende complexiteit gebruikt en worden de resultaten vergeleken met traditionele methoden voor geaggregeerde data. Daarnaast presenteren we een simulatie studie waarin we echte en pseudo-IPD generen en analyseren. We bestuderen de effecten van het aantal onderzoeken en de steekproefomvang per arm op verschillende statistieken zoals de dekkingsgraad, bias en mean squared error (MSE) voor de eentraps pseudo-IPD met de random-effects REML aanpak. De resultaten van de echte en de pseudo-IPD zijn grotendeels identiek, met uitzondering van enkele gevallen waarin problemen met de convergentie werden aangetroffen. Over het algemeen geeft de pseudo-IPD methode goede resultaten.

In **Hoofdstuk 3** breiden we de pseudo-IPD aanpak uit naar de situatie waarin continue uitkomstmaten worden gerapporteerd op baseline en bij follow-up. Een standaard meta-analyse van deze gegevens omvat doorgaans het berekenen en samenvoegen van gemiddelde verschillen van follow-up scores of veranderingsscores tussen onderzoeken. We stellen dat een ANCOVA op gegenereerde pseudo-IPD, gebaseerd op de voldoende steekproeffuncties, hiervoor een geschiktere methode is, aangezien het rekening houdt met de onbalans in de baselinemetingen en de correlatie van baseline en follow-up metingen. De voldoende functies in dit geval zijn: de gemiddelden op groepsniveau, standaarddeviaties en steekproefomvang op baseline en bij follow-up samen met de correlatie binnen de groep tussen de baseline en follow-up metingen. De pseudo-IPD kan vervolgens met behulp van ofwel een eentraps gestratificeerde studie of met een random studie ANCOVA of met tweetrapsmodellen geanalyseerd worden.

De pseudo-IPD method maakt het bovendien mogelijk om op een eenvoudige manier een interactie-effect tussen de behandeling en baselinemetingen op te nemen, waardoor de een verschillende effect van behandeling dat afhangt van de beginwaarden kan worden onderzocht. De voordelen van onze voorgestelde pseudo-IPD aanpak worden geïllustreerd aan de hand van twee voorbeelden. In het eerste voorbeeld gebruiken we de gerapporteerde geaggregeerde data en laten we zien dat (een- en tweetraps) pseudo-IPD ANCOVA modellen identieke resultaten opleveren als de echte IPD ANCOVA, zoals uitgevoerd in Riley e.a., 2013. In het tweede voorbeeld zijn onbalans op baseline en effectmodificatie aanwezig. De pseudo-IPD ANCOVA methode kan rekening houden met onbalans op baseline en nauwkeurige schattingen van het behandelingseffect geven. Daarnaast wordt een statistisch significante interactie gevonden tussen de baselinemetingen en het behandeleffect, die in een meta-regressie analyse gemist kan worden.

Hoofdstuk 4 beschrijft geschikte en minder geschikte meta-analytische technieken voor geaggregeerde continue uitkomsten gemeten op baseline en bij follow-up. Hiervoor gebruiken we een voorbeeld van negen onderzoeken waarin gewichtsmetingen zijn uitgevoerd, voor en na calciumsuppletie, waarbij aanzienlijke onbalans op baseline werd aangetroffen. Deze dataset werd oorspronkelijk geanalyseerd met behulp van een geaggregeerde data ANCOVA methode, die gevoelig is voor aggregatie valkuil.

Er worden verschillende meta-analyse methoden voor geaggregeerde data geïntroduceerd; van speciale aandacht zijn: a) een bekende, maar in de praktijk weinig gebruikte methode voor het recontrueren van ANCOVA-schattingen uit gepubliceerde samenvattende statistieken, waarbij wordt uitgegaan van gelijke groepsvariantie voor en na de behandeling en b) de pseudo-IPD methode, die pseudo-IPD genereert, die vervolgens met behulp van ANCOVA geanalyseerd wordt in een of tweetrapsmodellen. Over het algemeen hebben methoden die ANCOVA gebruiken de voorkeur boven de routinematig gebruikte meta-analyse methoden die follow-up of veranderingsscores gebruiken. De eentraps pseudo-IPD-methode biedt extra voordelen ten opzichte van de ANCOVA methode, omdat deze in staat is om effecten binnen onderzoeken en tussen onderzoeken te scheiden en de interactie tussen behandeling en baselinemetingen kan modelleren. De tweetraps IPD-methode wordt ook aanbevolen, omdat deze vergelijkbare voordelen biedt als de eentraps aanpak en minder statistische expertise vereist. We geven een uitgewerkt voorbeeld van een meta-analyse waarbij een groot deel van de samenvattende statistieken ontbreekt en geven een stapsgewijze datavoorbewerking gevolgd door analyse.

In **Hoofdstuk 5** presenteren we een webapplicatie voor de meta-analyse van studies met continue uitkomstmaten (rekening houdend met hun waardes op baseline), waardoor het uitvoeren van deze analyses toegankelijk wordt voor een breder publiek. Deze webapplicatie is bedoeld om de flexibiliteit van het LMM raamwerk te combineren met traditionele methoden voor meta-analyse van continue gegevens, en biedt veel modelleringsopties. In eerdere hoofdstukken hebben we besproken dat het gebruik van eindscores en veranderingsscores in het algemeen minder nauwkeurige schattingen van het behandeleffect zullen opleveren en dat ANCOVA methoden de voorkeur hebben, zowel met en zonder onbalans in de situatie op baseline.

De webapplicatie stelt zowel technisch als niet-technisch publiek in staat om de analyse eenvoudig uit te voeren onder een gebruiksvriendelijke interface.

Hoofdstukken 6 and **7** gaan in op netwerk-meta-analyse methoden.

Hoofdstuk 6 behandelt een toepassing van bewijs-synthese, waarbij een systematisch literatuur onderzoek en Bayesiaanse netwerk-meta-analyse worden uitgevoerd om de relatieve werkzaamheid van farmacologische en niet-farmacologische interventies bij therapieresistente depressieve patiënten te beoordelen. Deze toepassing is de eerste synthese van verschillende soorten interventies (farmacologisch versus somatisch), onder de aanname van 'transitieve ' placebo/sham armen, die grondig worden onderzocht in Sectie 6.3.3.

De belangrijkste bijdrage van dit werk is dat het een behandelingshiërarchie biedt over een reeks behandelingen, die waarschijnlijk niet rechtstreeks met elkaar kunnen worden vergeleken. De therapieresistente depressieve patiënten zijn ernstige depressieve patiënten, voor wie het grootste aanbod aan beschikbare therapie opties mogelijk niet nuttig is en het is van groot belang dat gezondheidswerkers beschikken over informatie over de rangschikking van behandelingen op basis van hun relatieve werkzaamheid.

In **Hoodfdstuk 7** introduceren we een eentraps pattern mixture model methode om het probleem van ontbrekende continue uitkomstgegevens in netwerk-meta-analyse aan te pakken. Ontbrekende uitkomstgegevens zijn alomtegenwoordig op studieniveau en kan leiden tot vertekende conclusies wanneer onderzoekers er geen rekening mee houden. We introduceren een eentraps pattern mixture model, waarmee we afwijkingen van de aanname, dat de ontbrekende data willekeurig zijn, kunnen bestuderen. De afwijkingen worden gekwantificeerd door de twee parameters genaamd IMDoM en IMRoM, die respectievelijk het contrast en de verhouding zijn tussen de niet-waargenomen gemiddelde waarde in de ontbrekende gegevens en de gemiddelde waarde in de waargenomen gegevens.

In plaats van vaste waarden aan te nemen voor de informatieve ontbrekende parameters, kunnen we er zogenoemde prior verdelingen aan toekennen. Een van de voordelen van het eentraps pattern mixture model is de mogelijkheid om meer te weten te komen over de ontbrekende gegevens door gegevens uit onderzoeken met verschillende fracties aan ontbrekende gegevens in één stap te combineren. Onze voorgestelde methode wordt geïllustreerd in twee netwerken van studies die een aanzienlijke hoeveelheid ontbrekende uitkomstgegevens bevatten. In deze toepassingen leren we weinig over de informatieve ontbrekende parameters, behalve voor één behandelingscombinatie, waar we zouden kunnen stellen dat de ontbrekende deelnemers van deze arm de tijd totdat de Parkinson symptomen terugkeren korter is in vergelijking tot de patiënten zonder ontbrekende gegevens.

BELANGRIJKSTE CONCLUSIES EN ASPECTEN VOOR VERVOLGONDERZOEK

Omdat continue uitkomstgegevens minder vaak voorkomen in systematische reviews hebben bewijs-synthesemethoden voor continue uitkomstgegevens minder aandacht gekregen in vergelijking tot methoden voor binaire gegevens. In dit proefschrift introduceren we statistische methodologie, die het mogelijk maakt om continue meta-analytische gegevens op meer geavanceerde manieren te analyseren, met standaard statistische software.

We introduceren een pseudo-IPD-aanpak op continue gegevens voor, die één keer worden gemeten (na de behandeling), en de uitbreiding ervan om ook waarden voor de behandeling (baseline) mee te nemen. Dit stelt onderzoekers in staat om de geaggregeerde data te analyseren zonder de aanname te maken dat de residuele varianties vast en bekend zijn, een aanname die veel gemaakt wordt in standaard meta-analytische methoden voor geaggregeerde data. De pseudo-IPD methode biedt dus de mogelijkheid om de residuele varianties binnen de studie te schatten, uitgaande van realistische scenario's. Eenvoudigere gevallen, bijvoorbeeld de aanname dat de varianties binnen het onderzoek per groep verschillen, maar niet tussen onderzoeken, kunnen formeel worden getest met behulp van een likelihood ratiotest.

De pseudo-IPD-methode biedt ook een eenvoudige aanpassing voor de vrijheidsgraden in de noemer, waarbij de normale verdeling vervangen wordt door een t-verdeling met een bepaald aantal vrijheidsgraden middels het gebruik van LMM-software. De betrouwbaarheidsintervallen voor de parameters die van belang zijn, meestal de parameter voor het behandelingseffect, zullen daardoor dus breder en

conservatiever zijn.

Het uitbreiden van de pseudo-IPD-methode om baseline metingen op te nemen is met name interessant omdat de literatuur in de bewijs-synthese van continue uitkomstgegevens nog steeds verdeeld is overde beste aanpak om deze data te analyseren. In de afwezigheid van IPD is het combineren van ANCOVA schattingen over studies de meest geschikte methode, maar dergelijke schattingen worden in de praktijk zelden gerapporteerd. Onze voorgestelde pseudo-IPD methode en de gereconstrueerde ANCOVA aanpak bieden een onmiddellijke oplossing voor dit probleem en zullen hopelijk een paradigmaverschuiving initiëren in het uitvoeren van meer ANCOVA-meta-analyses in de toekomst, samen met het bevorderen van betere en meer gedetailleerde rapportage van de statistische kengetallen door de onderzoekers. De ANCOVA schatter is efficiënter dan methoden die follow-up scores en verschilcores gebruiken. We hopen dat met de introductie van de pseudo-IPD ANCOVA aanpak de discussie over de vraag of veranderingsscores of follow-up scores geanalyseerd moeten worden, overbodig wordt. De pseudo-IPD methode maakt het bovendien mogelijk om interactie effecten tussen baseline en het behandelingseffect te onderzoeken, wat statistisch gezien de juiste manier is om de impact van kenmerken op patiëntniveau op de behandeling te onderzoeken.

Een belangrijke bijdrage van de pseudo-IPD-aanpak is dat het de kwestie van toegang tot de oorspronkelijke IPD mogelijk overbodig maakt, iets wat nog lastiger is sinds de nieuwe Algemene Verordening Gegevensbescherming die in 2018 van kracht werd. Hoewel de verordening, terecht, als doel heeft de privacy van gegevens te beschermen, bemoeilijkt het de openbaarmaking van IPD. De pseudo-IPD-methode biedt een oplossing, omdat het alleen gebruik maakt van geaggregeerde informatie die zonder privacy problemen openbaar kan worden gemaakt. Bovendien zou onze voorgestelde methode kunnen werken in een kader waarin sommige onderzoeken geaggregeerde data en andere IPD rapporteren, door pseudo-IPD te genereren voor de studies met geaggregeerde data en deze te combineren met de echte IPD.

In het eenvoudigste geval, dat we in **Hoofdstuk 2** beschouwen, zijn er geen metingen op baseline beschikbaar (of wordt er geen rekening mee gehouden) en dus is de 'natuurlijke' meta-analytische schatting het verschil in de ruwe (eind)gemiddelden. Onze voorgestelde eentraps pseudo-IPD methode schat dit gemiddelde verschil als de coëfficiënt van de behandelgroepindicator in het LMM. Deze maat voor effectgrootte is nuttig wanneer de uitkomsten in alle onderzoeken op dezelfde schaal worden gemeten, maar er zijn gevallen waarin verschillende instrumenten/schalen worden gebruikt om dezelfde uitkomst te beoordelen, bijvoorbeeld voor de ernst van depressie met behulp van HAM_D- of MADRS-schalen. De juiste maat voor effectgrootte is dan het gemiddelde gestandaardiseerde verschil. Uitbreiding van de pseudo-IPD methode tot deze gemiddelde gestandaardiseerde verschillen is mogelijk door de nodige gegevenstransformaties toe te passen alvorens het modelleren. Als de pseudo-IPD set wordt gegeneerd, moet hiervoor per onderzoek het gemiddelde in de controlegroep worden afgetrokken van het gemiddelde van de behandelingsgroep. Dit verschil wordt vervolgens gedeeld door een gemeenschappelijke standaarddeviatie, meestal de gepoolde standaarddeviatie tussen de twee groepen. Het is echter niet eenvoudig omdat er verschillende schatters voor de gemiddelde gestandaardiseerde verschillen zijn, zoals Glass' Δ, Cohen's d, Hedges' g, en de literatuur is bovendien

inconsistent met betrekking tot de schatting van de variantie. Ervan uitgaande dat men de getransformeerde uitkomst heeft verkregen, door te kiezen tussen de schatters voor de gemiddelde gestandaardiseerde verschillen op deelnemersniveau, kan een mixed effects-model worden toegepast op dezelfde manier als de modellen die in **Hoofdstuk 2** gepresenteerd zijn, uitgaande van vaste en/of willekeurige effecten voor de gestandaardiseerde gemiddelde uitkomst in de controlearm en voor het behandeleffect. Meer details zijn te vinden in het baanbrekende onderzoek van Goldstein e.a., 2000 en in Walwyn en Roberts, 2017, waarin clustereffecten zijn opgenomen.

De pseudo-aanpak van **Hoofdstuk 2** kan op een vrij eenvoudige manier worden uitgebreid tot netwerk-meta-analyse door voor elke aanvullende behandeling vaste en willekeurige termen aan het LMM toe te voegen. Bovendien, hoewel de methode wordt voorgesteld voor parallelle gerandomiseerde gecontroleerde onderzoeken met twee groepen, kan deze worden aangepast aan andere designs — twee perioden cross-over-onderzoeken met twee behandelingen, mits informatie over de correlatie binnen de deelnemers beschikbaar is.

Uitbreidingen van de pseudo-IPD ANCOVA-aanpak beschreven in **Hoofdstuk 3** meta-analyse van het netwerk en tot twee periodes, twee-behandelingen cross-over studies zijn ook eenvoudig. Een toekomstig onderzoeksonderwerp dat moet worden overwogen, is het opnemen van andere covariaten dan de baseline, ervan uitgaande dat de samenvattende statistieken beschikbaar zijn. In dit geval zouden de voldoende functies de gemiddelden en de variantie-covariantiematrix per behandelingsgroep zijn.

Ten slotte heeft ons werk dezelfde beperkingen die inherent zijn aan het bewijssynthese raamwerk, zoals publicatiebias. Deze vertekening kan ontstaan wanneer de (netwerk) meta-analyse niet al het bewijs met betrekking tot de onderzoeksvraag bevat, maar slecht een deelverzameling, vaak bestaande uit onderzoeken met statistisch significante effecten. In dit geval kan een meta-analyse een vertekend beeld opleveren, dat niet kan worden verholpen door onze voorgestelde methoden. Bovendien kunnen de gepoolde effecten die in (netwerk)meta-analyses worden geschat nauwkeuriger zijn dan de effecten die in enkelvoudige studies worden verkregen, op voorwaarde dat de studies van hoge kwaliteit zijn.

Deze dissertatie biedt vooruitgang op het gebied van bewijssynthese van continue resultaatgegevens en pakt belangrijke problemen aan, zoals het combineren van resultaten wanneer er baselineverschillen bestaan tussen de onderzoeksarmen en het ontbreken gegevens als gevolg van het verlies van patienten gedurende het onderzoek. We hopen dat dat de pseudo IPD ANCOVA-aanpak de standaardmethode wordt voor het meta-analyseren van dergelijke resultaten in de toekomst en dat onze nieuw ontwikkelde tool de gebruikers in staat stelt deze methode toe te passen.