



Zorginstituut Nederland

Kader voor analysemethoden (D3)

Datum	24 juni 2020
Status	Concept

Colofon

Projectnaam	Regie op Registers Dure Geneesmiddelen
Versienummer	1.0
Projectleider	Anke ter Horst
Volgnummer	2020015900
Opdrachtgever	Ministerie van VWS
Contactpersoon	mw. M.J.H.J. Dekker +31 (0)6 30917071
Afdeling	Zorg I
Uitgebracht aan	Ministerie van VWS

Inhoud

Colofon—1

Samenvatting—5

Inleiding—7

1 Evalueren Kwaliteit Patiëntenregistratie—11

2 Studieopzet voor analyse (studiedesign)—13

- 2.1 Basisdesigns & Analyse mogelijkheden—13
 - 2.1.1 Basisdesigns—13
 - 2.1.1.1 Cohortstudie & Case-cohort studie—13
 - 2.1.1.2 Cross-sectionele studie—13
 - 2.1.1.3 Case-control studie & nested case-control studie—13
 - 2.1.1.4 Cross-sectioneel cohort design—13
 - 2.1.2 Analyse mogelijkheden—13
 - 2.1.2.1 Case-only designs—13
 - 2.1.2.2 Incident user design—13
 - 2.1.2.3 Regression discontinuity design—13
 - 2.1.2.4 Difference-in-difference estimation—13
 - 2.1.2.5 Interrupted time series—13
- 2.2 Overall design & tijdsankers—14
 - 2.2.1 Onderzoeksvraag en studiepopulatie—14
 - 2.2.2 Selectie comparator—14
 - 2.2.3 In-en exclusiecriteria—14
 - 2.2.3.1 Cohort/study entry datum—14
 - 2.2.3.2 Enrollment window—14
 - 2.2.3.3 Exclusion definition window—14
 - 2.2.3.4 Interval voor bepalen van 'incident exposure'—14
 - 2.2.4 Exposure definition—14
 - 2.2.4.1 Exposure assessment window—14
 - 2.2.4.2 Switching/add-on—14
 - 2.2.5 Outcome definition—14
 - 2.2.5.1 Interval voor bepalen 'incident outcomes'—14
 - 2.2.5.2 Follow-up window—14
 - 2.2.5.3 Censoring criteria—14
 - 2.2.5.4 Outcome event datum—14
 - 2.2.6 Covariate selection—14
 - 2.2.6.1 Covariate assessment window—14
 - 2.2.7 Treatment heterogeneity—14
 - 2.2.8 Effectieve sample size—14

3 Uitdagingen bij de analyse van observationele data & overzicht van methoden voor analyse—15

- 3.1 Selection bias—15
- 3.2 Information bias—15
- 3.3 Tijdsgelateerde bias—15
- 3.4 Kleine sample size—15
- 3.5 Interactie—15
- 3.6 Time-varying exposure—15
- 3.7 Competing risk—15

3.8	Natural history data—15
3.9	Confounding—15
3.9.1	Definitie—16
3.9.2	Confounding by indication—16
3.9.3	Measured vs. unmeasured confounding—16
3.9.4	Tijdsafhankelijke confounding—16
3.9.5	Methoden voor omgang met confounding—17
3.9.5.1	Design—17
3.9.5.2	Stratificatie—18
3.9.5.3	Regressie—18
3.9.5.4	Disease risk score—18
3.9.5.5	Propensity score—19
3.9.5.6	Marginal structural models—19
3.9.6	Methoden voor omgang met unmeasured confounding—19
3.9.6.1	Instrumental variable analyse—19
3.9.7	Methoden voor omgang met tijdsafhankelijke confounding—20
3.9.8	Sensitiviteitsanalyse—20
3.10	Missing data—21
3.10.1	Patronen in missing data—21
3.10.2	Mechanismen voor missing data—22
3.10.2.1	MCAR – Missing completely at random—22
3.10.2.2	MAR – Missing at random—22
3.10.2.3	MNAR – Missing not at random—22
3.10.3	Methoden voor omgang met missing data—22
3.10.3.1	Complete case analysis & available case analysis—22
3.10.3.2	'Simpele' methoden—23
3.10.3.3	Single en multiple imputation—23
3.10.3.4	Maximum – likelihood methods—24
3.10.3.5	Hot and cold deck imputatie—24
3.10.3.6	Predictive mean matching—25
3.10.3.7	Verdere methoden.—25
3.10.4	Sensitiviteitsanalyse—25
3.10.5	Algemene aanbevelingen—25
4	Richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming—26
4.1	Richtlijnen voor rapportage—26
4.2	Evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming—26
5	Kennislacunes—28
5.1	Kennislacunes—28
	Bijlage I EunetHTA REQUEST tool—29
	Bijlage II: RECORD-PE statement—37
	Bijlage III: Grace Checklist—45
	Referenties—48

Samenvatting

Een belangrijk onderdeel van het project Regie op Registers voor Dure Geneesmiddelen (RORDGM) is het vaststellen van een richtlijn voor de analyse van observationele gegevens uit patiëntenregistraties ten behoeve van pakketbeheer. De doelstellingen zoals die geformuleerd zijn in de uitvoeringstoets voor het eerste jaar van het project, de definitiefase, zijn:

- Het definiëren van een conceptkader voor deze richtlijnen gebaseerd op reeds bestaande relevante literatuur en rapporten;
- Het inventariseren van mogelijke methodologische lacunes die moeten worden geadresseerd, bijvoorbeeld door extra onderzoek.

Dit document is het conceptkader voor de richtlijn voor de analyse van gegevens uit patiëntenregistraties. Het uiteindelijke doel van deze richtlijn is om handvaten te bieden voor het uitvoeren van accurate en betrouwbare analyses op basis van deze observationele gegevens. Het document bevat, naast een beschrijving van de relevante methodologische overwegingen, verwijzingen naar bijbehorende literatuur. Een inventarisatie van methodologische lacunes maakt ook deel uit van het kader. Conform de uitvoeringstoets zal het huidige conceptkader tijdens de case studiefase van RORDGM verder uitgewerkt worden tot een definitieve richtlijn voor de analyse van (observationele) registratiedata. Bij deze verdere uitwerking zal nadrukkelijk aangesloten worden bij de reeds bestaande "Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg" van het Zorginstituut Nederland.[1]

De inhoud van het conceptkader voor de richtlijn is tot stand gekomen in afstemming met de Expertisegroep. In het kort worden de volgende onderwerpen behandeld:

- "Evalueren kwaliteit patiëntenregistratie" (Hoofdstuk 1);
- "Studieopzet voor analyse (studiedesign)" (Hoofdstuk 2);
- "Uitdagingen bij de analyse van observationele data en een overzicht van methoden voor analyse" (Hoofdstuk 3);
- "Richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming" (Hoofdstuk 4);
- "Kennislacunes" (Hoofdstuk 5).

Het conceptkader bevat per hoofdstuk een introductie op het thema, verwijzingen naar de belangrijkste literatuur en een overzicht van de onderwerpen die we in de definitieve richtlijn verder willen uitwerken. Een aantal onderwerpen zijn al in detail uitgewerkt: 'missing data' en 'confounding' (Hoofdstuk 3). Deze onderwerpen hebben bij de uitwerking prioriteit gekregen, omdat dit uitdagingen zijn die zeer vaak optreden bij de analyse van observationele data. Daarnaast werd door de Expertisegroep aangegeven dat er specifiek voor het onderwerp 'missing data' behoefte was aan een richtlijn.

Over 'studiedesign' (Hoofdstuk 2) zijn reeds uitgebreide handboeken en richtlijnen geschreven. In het huidige conceptkader is volstaan met een inhoudsopgave van de meest relevante studiedesigns en keuzes binnen deze designs (b.v. exposure en outcome definition). Tijdens de case studiefase van RORDGM zal dit hoofdstuk verder uitgewerkt worden, waarbij ook nadrukkelijk de aansluiting wordt gezocht bij reeds bestaande handboeken en richtlijnen om een herhaling van reeds bestaand werk te voorkomen.

De meeste onderwerpen die binnen het huidige conceptkader behandeld worden zijn

technisch en methodologisch van aard. Er zijn echter ook enkele onderwerpen die raken aan vraagstukken op het gebied van governance en de plaatsbepaling van studies op basis van registratiedata binnen pakketbeheer. Voorbeelden hiervan zijn de evaluatie van de kwaliteit van patiëntenregistraties (Hoofdstuk 1) waarbij de "Registry Evaluation and Quality Standards (REQueST) Tool" als uitgangspunt wordt genomen. Om deze tool goed te kunnen gebruiken is gedetailleerde informatie over de patiëntenregistratie nodig, die niet altijd publiekelijk beschikbaar is. Binnen het huidige conceptkader is nog niet uitgewerkt welke stakeholders deze evaluatie uitvoeren en op welke wijze deze informatie wordt aangeleverd.

Een ander voorbeeld is Hoofdstuk 4, waarin richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies worden behandeld. In het huidige kader is nog niet uitgewerkt wat de exacte plaatsbepaling is van deze evaluatie voor het gebruik van registratiedata binnen pakketbeheer. Deze beide vraagstukken zullen tijdens de case studiefase van ROR DGM in afstemming met de Expertisegroep, veldpartijen en secretarissen van de Wetenschappelijke Adviesraad (WAR) en beoordelaars van het Zorginstituut verder worden uitgewerkt.

Ten slotte zal in samenwerking met bovengenoemde stakeholders een visie worden ontwikkeld op de governance van studiedesign en uitvoering van dataanalyse. Vraagstukken hierbij zijn 'welke stakeholder(s) voeren de dataanalyse uit?', 'op welke wijze worden stakeholders betrokken bij keuzes op het gebied vraagstelling, studiedesign en dataanalyse?', en 'welke aspecten van de dataanalyse moeten ten minste beschreven worden in de verslaglegging?'.

Inleiding

Het gebruik van data uit de dagelijkse praktijk om tot bewijs te komen dat bij kan dragen aan de beoordeling van effectiviteit, kosteneffectiviteit en gepast gebruik van (dure) geneesmiddelen leidt al jaren tot verhitte discussies. In vele beoordelingskaders zoals Grading of Recommendations Assessment, Development and Evaluation (GRADE) en in de hiërarchie van bewijs is de randomized clinical trial (RCT) aangemerkt als de gouden standaard om verschillen tussen behandelingen zo goed als mogelijk vast te stellen en is men terughoudend met betrekking tot gebruik van bewijs dat gebaseerd is op data uit de praktijk (ook wel observationele data of real-world data (RWD)). Deze terughoudendheid vindt zijn oorsprong in allerlei methodologische vragen rondom betrouwbaarheid van dergelijk bewijs. Het bekendste voorbeeld van een methodologisch probleem dat voorkomt bij dit type bewijs is confounding by indication, waarbij de indicatie voor behandeling afhankelijk is van patiëntenkarakteristieken die sterk gebonden zijn aan de prognose van de patiënt. Dit heeft als gevolg dat patiënten met en zonder behandeling niet zomaar kunnen worden vergeleken.

Er is echter een groeiend besef dat, gezien de veranderingen in de zorg en de ontwikkeling van behandelingen die steeds meer op individueel niveau worden afgestemd, observationele gegevens van essentieel belang zijn om de kwaliteit van zorg beter vast te kunnen stellen. In het specifieke geval van (dure) geneesmiddelen komen er bijvoorbeeld steeds vaker geneesmiddelen op de markt die voor hele kleine patiëntenpopulaties zijn ontwikkeld en waarbij RCTs niet beschikbaar zijn of moeilijk kunnen worden uitgevoerd. Daarnaast worden in de praktijk geneesmiddelen steeds vaker in combinatie gebruikt, terwijl (meestal) alleen toepassing van individuele geneesmiddelen in RCTs is onderzocht. Ook zijn er recente ontwikkelingen, bijvoorbeeld in de oncologie, waarbij tumoragnostische geneesmiddelen op de markt komen. Deze geneesmiddelen worden goedgekeurd voor gebruik in alle tumoren met bepaalde markers. Daarmee kan het zo zijn dat er niet voor alle typen kanker individuele RCT data beschikbaar is over de effectiviteit van deze middelen. Tenslotte blijkt ook steeds vaker dat de externe validiteit van de RCTs waarin dure geneesmiddelen worden onderzocht beperkt is. Dat wil zeggen dat een groot deel van patiënten die met deze geneesmiddelen behandeld wordt, wezenlijk verschilt van de patiënten die zijn opgenomen in de RCT.

Om deze redenen wordt steeds vaker onderkend dat bewijs gebaseerd op observationele data een (aanvullende) rol kan spelen naast bewijs dat is gebaseerd op RCTs. Er bestaan echter nog steeds veel zorgen over de kwaliteit van bewijs dat is gebaseerd op observationele data. In de afgelopen jaren zijn er daarom vele methoden ontwikkeld om op meer verantwoorde manier met observationele data om te gaan. Daarbij gaat het bijvoorbeeld om nieuwe statistische en epidemiologische methoden om voor allerlei vormen van vertekening (bias) te corrigeren. Ook worden nieuwe studiedesigns ontworpen en oude studiedesigns geoptimaliseerd. Daarnaast worden er ook methoden ontwikkeld om observationele data te linken aan data uit RCTs om zo specifiek (op het niveau van de subpopulatie of zelfs het individu) en zo betrouwbaar mogelijk de effectiviteit van een behandeling te schatten. Een voorbeeld hiervan is het gebruik van historische controles (natural history data).

Deze ontwikkelingen passen direct binnen het kader van het project op Regie op Registers voor Dure Geneesmiddelen. Om in de toekomst het gebruik van data uit patiëntenregistraties voor pakketbeheer en monitoring van gepast gebruik van dure geneesmiddelen mogelijk te maken, zal duidelijk moeten worden aan welke eisen die data, en het bewijs dat gebaseerd is op deze data, moet voldoen om bruikbaar zijn te voor deze doelen. Daarbij gaat het niet alleen om de kwaliteit van de data in de patiëntenregistraties zelf, maar ook om de opzet van de studies die gebruikt worden om met deze data tot bewijs te komen en de methoden die gebruikt worden om dit bewijs op waarde te kunnen schatten.

Deze richtlijn is bedoeld om handvaten te bieden aan indieners en beoordelaars van analyses die gebaseerd zijn op data uit patiëntenregistraties en betrekking hebben op vragen in het kader van pakketbeheer (breed genomen als effectiviteit, kosteneffectiviteit en gepast gebruik) van (dure) geneesmiddelen

Deze richtlijn is niet bedoeld als handboek voor het opzetten van een registratie. Ook is deze richtlijn is niet bedoeld als leerboek voor het inrichten van studies en het uitvoeren van statistische analyses. Wij nemen aan dat gebruikers van de richtlijn voldoende epidemiologische basiskennis hebben om zelfstandig studies uit te voeren en te analyseren.

Een belangrijk uitgangspunt bij deze richtlijn is dat er een verschil is tussen de kwaliteit van (de gegevens in) een patiëntenregistratie en de kwaliteit van de studie die op basis van deze gegevens wordt uitgevoerd om bewijs te genereren. Om die reden hebben we in deze richtlijn een duidelijk onderscheid gemaakt tussen een patiëntenregistratie an sich en een studie die op basis van de data in een registratie wordt uitgevoerd (hier zal in dit document naar verwezen worden als 'registerstudie' (zie ook [2]).

Dit is het eerste concept van de richtlijn. De richtlijn is nog volop in ontwikkeling en niet alle onderdelen zijn volledig uitgewerkt in het huidige document. De secties die nog niet zijn geadresseerd zijn, zullen in de case studiefase van het project Regie op Registers voor Dure Geneesmiddelen verdere invulling krijgen. Per onderwerp wordt een beschrijving van de relevante methodologische overwegingen met verwijzingen naar bijbehorende literatuur opgenomen. Voor de onderwerpen 'selection bias' en 'natural history data' zal een verdiepingmodule geschreven worden zodat deze onderwerpen in meer detail behandeld kunnen worden. Dit omdat dit relevante onderwerpen zijn waarbij de behoefte aan een richtlijn geconstateerd is door de Expertisegroep en andere initiatieven zoals Innovative Medicines Initiative (IMI) GetReal.[3]

Er is binnen deze richtlijn specifiek aangegeven welke secties zijn uitgewerkt en welke secties verdere aandacht behoeven. Hierbij wordt de keuze voor prioritering van onderwerpen gemotiveerd.

- In hoofdstuk 1 is een eerste aanzet gedaan om te beschrijven hoe we de kwaliteit van een patiëntenregistratie willen evalueren. Daarbij is van een bestaand Europees concept uitgegaan. Wel moeten in een volgende fase afspraken aangaande praktische invulling gemaakt worden.
- In hoofdstuk 2 is het thema studieopzet (studiedesign) uitgewerkt. Omdat hoofdstuk 2 basisonderwerpen aandoet, waarover reeds veel literatuur beschikbaar is, is er voor gekozen een inhoudsopgave te maken van de onderwerpen die binnen deze sectie geadresseerd moeten worden. De invulling zal in een latere fase vorm krijgen.

- Hoofdstuk 3 beschrijft de uitdagingen bij de analyse van observationele data en geeft een overzicht van methoden voor analyse. Hierbij is gekozen voor een primaire uitwerking van de onderwerpen 'missing data' en 'confounding'. Voor het onderwerp missing data is de behoefte aan een richtlijn geconstateerd; het is in de praktijk lastig hier goed mee om te gaan. 'Confounding' wordt gezien als de grootste uitdaging bij het gebruik van observationele data voor het beantwoorden van (kosten)effectiviteitsvragen.
- In hoofdstuk 4 worden richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming besproken. Ook bij dit hoofdstuk zullen in een volgende fase afspraken aangaande praktische invulling gemaakt moeten worden.
- Hoofdstuk 5 beschrijft kennislacunes. Dit hoofdstuk is nu deels uitgewerkt en zal op basis van de uitwerking van de andere hoofdstukken en de aanvulling van experts worden gefinaliseerd.

1 Evalueren Kwaliteit Patiëntenregistratie

Voor beoordeling van de kwaliteit van een patiëntenregistratie bevelen wij de EunetHTA REQueST tool als uitgangspunt aan, zoals hieronder wordt beschreven. Dit hoofdstuk heeft nog nadere uitwerking op praktisch vlak. De gedachte is om het gebruik van de REQueST tool te testen binnen de case studies die tijdens de case studiefase van het project Regie op Registers voor Dure Geneesmiddelen worden uitgevoerd. Op basis van deze ervaringen kan worden vastgesteld of (1) deze tool standaard voor alle registraties gebruikt moet worden en (2) of er eventueel aanvullende instrumenten moeten worden gebruikt. Omdat niet alle informatie over registraties zoals gevraagd in het instrument vrij beschikbaar is, zullen er mogelijk ook afspraken moeten worden gemaakt over het aanleveren van informatie over de registratie.

Zoals ook uit de REQueST tool blijkt, is datakwaliteit is een belangrijk aspect van de kwaliteit van een registratie. Datakwaliteit heeft betrekking op de consistentie, compleetheid, precisie en tijdigheid van de data als vastgelegd in de registratie. In de case studiefase van het project Regie op Registers voor Dure Geneesmiddelen zal hiervoor een aparte richtlijn worden opgesteld.

Als registerstudies worden gebruikt ten behoeve van besluitvorming, moet er, naast vertrouwen in de kwaliteit van de studie, vertrouwen zijn in de kwaliteit van de bron: de registratie. Dit hoofdstuk biedt handvaten voor het evalueren van de kwaliteit van de registratie. Het evalueren van de kwaliteit van de registerstudie zelf wordt beschreven in hoofdstuk 4.

Het uitgangspunt bij dit hoofdstuk is om zover als mogelijk gebruik te maken van instrumenten die (1) aansluiten bij het doel om (data uit) patiëntenregistraties een rol te geven binnen pakketbeheer en (2) relevant zijn voor de Nederlandse setting. Daarom ligt het voor de hand de REQueST tool voor het evalueren van de kwaliteit van registraties als startpunt te gebruiken.[4] De Registry Evaluation and Quality Standards Tool, (REQueST) is recent ontwikkeld door The European Network for Health Technology Assessment (EUnetHTA), waar Zorginstituut Nederland ook deel van uitmaakt. REQueST is mede gebaseerd op eerdere richtlijnen voor registraties en het gebruik van data uit patiëntenregistraties ten behoeve van besluitvorming (zie referenties [2, 5-14] voor deze en aanvullende achtergrondliteratuur), en is getest in enkele Health Technology Assessment (HTA) omgevingen.

REQueST bestaat uit drie onderdelen. Het eerste onderdeel raakt aan hoe de registratie is opgezet en welke data de registratie bevat. Hier wordt bijvoorbeeld gevraagd om informatie over de grootte van de registratie en de lengte van follow-up. In het tweede deel wordt gekeken naar essentiële standaarden: minimale eisen die gesteld worden aan een registratie om datakwaliteit en databescherming te kunnen garanderen, zoals informed consent. In de derde stap kan er gekeken worden naar 'elementen' die relevant kunnen zijn met betrekking tot het gebruik van (data uit) de registratie voor specifieke doeleinden, zoals de mogelijkheid om de data uit de registratie met data uit een andere informatiebron te combineren. De velden uit de REQueST tool zijn opgenomen in bijlage I.

Het is hierbij belangrijk om te realiseren dat het niet mogelijk is de kwaliteit van een registratie te beoordelen zonder context, ofwel zonder te weten voor welk doel de (data uit) de registratie gebruikt wordt. Daarom zal voor elke registerstudie die binnen een specifieke registratie wordt opgezet aan de hand van de velden in het instrument gekeken moeten worden waar mogelijke onzekerheden liggen binnen de registratie die relevant zijn met betrekking tot de onderzoeksvraag.

2 Studieopzet voor analyse (studiedesign)

Binnen hoofdstuk 2 wordt het thema 'studieopzet voor analyse (studydesign)' uitgewerkt. Er is een inhoudsopgave gemaakt van onderwerpen die binnen dit onderdeel geadresseerd gaan worden (op basis van referenties [7, 15-17] en input van de Expertisegroep). Deze onderwerpen zijn verdeeld over twee secties. De eerste sectie zal informatie over basale studiedesigns en analysemethoden bevatten. De tweede sectie is opgebouwd uit onderwerpen die relateren aan de opbouw van het design van een individuele studie inclusief beschrijving van de chronologie van de studie. Deze onderwerpen zijn, ongeacht het gekozen basisdesign, relevant. Er zal binnen dit hoofdstuk ook aandacht worden besteed aan de grafische weergave van studiedesigns [16]. Deze ondersteunen rapportage over studies zoals besproken in hoofdstuk 4 'richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming'. Daarnaast zullen de connecties met hoofdstuk 3 'Uitdagingen bij de analyse van observationele data en overzicht van methoden voor analyse' duidelijk geïllustreerd worden.

Omdat dit hoofdstuk basisonderwerpen aandoet, waarover reeds veel literatuur beschikbaar is, is er voor gekozen om voor nu te volstaan met een inhoudsgave. De onderwerpen zullen in de case studiefase van het project Regie op Registers voor Dure Geneesmiddelen verder uitgewerkt worden.

2.1 Basisdesigns & Analyse mogelijkheden

2.1.1 *Basisdesigns*

- 2.1.1.1 Cohortstudie & Case-cohort studie
- 2.1.1.2 Cross-sectionele studie
- 2.1.1.3 Case-control studie & nested case-control studie
- 2.1.1.4 Cross-sectioneel cohort design

2.1.2 *Analyse mogelijkheden*

- 2.1.2.1 Case-only designs
 - 1 Case-crossover design
 - 2 Case-time-controlled design
 - 3 Self-controlled case-series design
- 2.1.2.2 Incident user design
- 2.1.2.3 Regression discontinuity design
- 2.1.2.4 Difference-in-difference estimation
- 2.1.2.5 Interrupted time series

2.2 Overall design & tijdsankers

2.2.1 Onderzoeksvraag en studiepoulatie

2.2.2 Selectie comparator

Keuze van groepen gebruikt voor vergelijking. In relatie tot onderzoeksvraag, potentie voor bias en confounding, belang van vaststellen initiatieperiode, exposurewindow en dergelijke.

2.2.3 In-en exclusiecriteria

2.2.3.1 Cohort/study entry datum

Datum waarop participant deel uit gaat maken van de studiepoulatie.

2.2.3.2 Enrollment window

Eventueel, als eisen worden gesteld aan de tijd dat een individu gevolgd is in de registratie alvorens deze wordt opgenomen in de studiepoulatie.

2.2.3.3 Exclusion definition window

Periode waarin exclusiecriteria worden toegepast (voor cohort entry datum).

2.2.3.4 Interval voor bepalen van 'incident exposure'

Als er binnen deze periode geen blootstelling is, wordt nieuwe blootstelling geclassificeerd als 'incident exposure'. Als er wel blootstelling is wordt dit gezien als 'prevalent exposure' (voor cohort entry datum).

2.2.4 Exposure definition

2.2.4.1 Exposure assessment window

Periode waarin 'exposure status' wordt bepaald, indien blootstelling niet bepalend is voor cohort entry (voor follow-up window).

2.2.4.2 Switching/add-on

2.2.5 Outcome definition

2.2.5.1 Interval voor bepalen 'incident outcomes'

Als er geen uitkomst is geobserveerd binnen dit interval, wordt de volgende uitkomst als 'incident outcome' beschouwd.

2.2.5.2 Follow-up window

Periode waarin de studiepoulatie 'at risk' is voor het ontwikkelen van de uitkomst.

2.2.5.3 Censoring criteria

Eventuele maxima die gesteld worden aan de analytische tijdsperiode voor follow-up.

2.2.5.4 Outcome event datum

Datum dat uitkomst optreedt.

2.2.6 Covariate selection

Hier zal de link tussen confounding en het selecteren van covariaten behandeld worden (inclusief Directed Acyclic Graphs).

2.2.6.1 Covariate assessment window

Ook wel baseline; periode waarin covariaten worden gemeten (voor exposure assessment window).

2.2.7 Treatment heterogeneity

Variatie in behandelingen en behandeleneffect.

2.2.8 Effectieve sample size

3 Uitdagingen bij de analyse van observationele data & overzicht van methoden voor analyse

Binnen hoofdstuk 3 zijn de thema's 'uitdagingen bij analyse van observationele data' en 'methoden voor analyse' uitgewerkt. Er is een inhoudsopgave gemaakt van onderwerpen die binnen dit hoofdstuk geadresseerd gaan worden (op basis van referenties [7, 13, 15, 18, 19-23] en input van de Expertisegroep). De onderwerpen 'confounding' en 'missing data' zijn al uitgewerkt. Deze onderwerpen hebben bij de uitwerking prioriteit gekregen, omdat er voor missing data nog geen richtlijn bestaat en confounding wordt gezien als een van de grootste uitdagingen bij het gebruik van observationele data voor het beantwoorden van (kosten)effectiviteitsvragen.

De onderwerpen die nog niet geadresseerd zijn, zullen in de case studiefase van het project Regie op Registers voor Dure Geneesmiddelen verdere invulling krijgen. Voor 'selection bias' en 'natural history data' zal een verdiepingmodule opgesteld worden zodat deze onderwerpen in meer detail behandeld kunnen worden. Selection bias is een belangrijke bron van vertekening in observationele studies die vaak wordt onderschat. Natural history data is een zeer actueel onderwerp omdat er steeds vaker geneesmiddelen op de markt komen op basis van single-arm trials en een vergelijking met onbehandelde patiënten alleen kan worden gemaakt op basis van historische controles.

- 3.1 Selection bias**
- 3.2 Information bias**
- 3.3 Tijdsgerelateerde bias**
- 3.4 Kleine sample size**
- 3.5 Interactie**
- 3.6 Time-varying exposure**
- 3.7 Competing risk**
- 3.8 Natural history data**

- 3.9 Confounding**

Referenties [7, 15-28] hebben bij uitwerking van dit stuk als uitgangspunt gediend. Aanvullende bronnen voor de lezer worden gegeven onder het kopje "Relevante bronnen".

Confounding is een van de grootste uitdagingen bij het werken met observationele data, en speelt een nog significantere rol als er op basis van observationele data uitspraken worden gedaan over (kosten)effectiviteit. Deze sectie gaat in op de typen confounding en adresseert de meest gangbare methoden om met confounding om te gaan.

3.9.1 *Definitie*

Er is sprake van confounding als de associatie tussen de blootstelling ('exposure') en de uitkomst verstoord wordt door aanwezigheid van een derde variabele (de confounder). Een variabele is een confounder als deze geassocieerd is met (1) de blootstelling, (2) de uitkomst, en (3) niet in het causale pad tussen de blootstelling en uitkomst ligt, ofwel geen intermediaire variabele is. Als dit laatste wel het geval is, kan er sprake zijn van tijdsafhankelijke confounding (3.9.4).

3.9.2 *Confounding by indication*

Confounding by indication, ook wel 'channeling' of confounding by severity genoemd, is een type confounding dat vaak voorkomt bij farmacoepidemiologisch onderzoek. Confounding by indication treedt op als de keuze voor behandeling afhankelijk is van (bekende of onbelende) patiëntkenmerken die een associatie hebben met de uitkomst onder studie, zoals ernst van ziekte. Over het algemeen kunnen de methoden zoals beschreven in deze sectie toegepast worden bij confounding by indication. Echter is het binnen effectiviteitsstudies meer uitdagend om op de juiste manier met confounding by indication om te gaan aangezien de associatie tussen de behandeling en uitkomst daar de primaire uitkomst is.

Relevante bronnen

Het artikel van McMahon (2003) introduceert methoden voor omgang met confounding by indication binnen effectiviteitsstudies op basis van observationele data.[29]

3.9.3 *Measured vs. unmeasured confounding*

Er wordt onderscheid gemaakt tussen 'gemeten' en 'ongemeten' confounding. Omdat geen enkele database informatie over alle mogelijke confounders bevat, zal er altijd sprake zijn van ongemeten confounders waarvoor niet gecorrigeerd kan worden. Dit resulteert in residuele confounding. Randomisatie is de enige manier om compleet voor ongemeten confounding te corrigeren. Daarom is binnen observationele studies van belang om de mogelijke impact van residuele confounding in te schatten en ter discussie te stellen. Omdat de impact op uiteindelijke resultaten groot kan zijn, is het sterk aan te raden sensitiviteitsanalyses uit te voeren. Dit wordt verder behandeld in de corresponderende sectie (3.9.8).

Relevante bronnen

In Fewell et al. (2007) wordt de impact van ongemeten en residuele confounding bediscussieerd en geïllustreerd aan de hand van simulatiestudies.[30]

3.9.4 *Tijdsafhankelijke confounding*

Bij tijdsafhankelijke confounding beïnvloeden confounders en risicofactoren elkaar. De variabele is zowel een confounder als een intermediaire uitkomst. Een voorbeeld is het effect van een HbA-verlagende medicatie op een cardiovasculaire uitkomst (HbA: maat voor gemiddelde bloedglucosewaarden over de afgelopen maanden). HbA is een confounder van dit effect omdat een hoog HbA een reden is voor gebruik van de medicatie. Tegelijk is HbA ook een risicofactor voor de cardiovasculaire

uitkomst.

3.9.5 *Methoden voor omgang met confounding*

Met betrekking tot omgang met confounding zijn er twee veelgebruikte benaderingen te onderscheiden: conditioneel en marginaal. De associatie tussen de blootstelling (in het kader van farmacoepidemiologische studies vaak de behandeling) en de uitkomst wordt gecorrigeerd voor andere variabelen (conditioneel) of er wordt gezorgd voor een evenredige verdeling van (mogelijke) confounders over de te vergelijken groepen (marginaal). Regressie (3.9.5.3), propensity score (3.9.5.5) en instrumental variable analyse (3.9.6.1) kennen een conditionele aanpak, matching (3.9.5.1), restrictie (3.9.5.1), stratificatie (3.9.5.2.), disease risk score (3.9.5.4) en marginal structural models (3.9.5.6) kennen een marginale aanpak.

3.9.5.1 Design

Er zijn enkele algemene designkeuzes die kunnen helpen om confounding en andere vormen van bias te beperken. De keuze van de vergelijkende groep (comparator; zie sectie 2.2.1), de keuze voor covariaten (zie sectie 2.2.6) en het includeren van 'incident vs. prevalent' gebruikers kunnen ook invloed hebben op de omvang van bias en confounding in een studie. Berger et al. (2012) en The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP; 2010) gaan dieper in op algemene 'good research practices' die het optreden van bias en confounding tegen kunnen gaan.[7, 19] Enkele van deze keuzes kunnen gezien worden als vorm van restrictie.

Restrictie

Restrictie, waarbij er eisen worden gesteld aan wie wel en wie niet tot de studiepopulatie toegelaten wordt, kan toegepast worden om de populatie onder studie meer homogeen te maken met betrekking tot potentiële confounders. Restrictie kan de externe validiteit van een studie beperken, maar bij farmacoepidemiologische studies hoeft dit niet altijd het geval te zijn: uitsluiting van een bepaalde groep patiënten bij toediening van een behandeling kan de situatie in de praktijk reflecteren.

Relevante bronnen

Schneeweiss et al. (2007) demonstreren het effect van toenemende restrictie op puntschattingen en bias.[31]

Matching

Bij matching wordt elke participant met de blootstelling onder studie gekoppeld aan een participant in de vergelijkende groep die vergelijkbaar is met betrekking tot één of meerdere karakteristieken (de mogelijke confounder(s)). Er kan exact of inexact gematched worden. Bij exacte matching wordt gematched op het niveau van de individuele variabele(n). Bij inexacte matching wordt gematched op basis van een functie van variabelen. Vaak wordt hierbij een 'nearest neighbour' of 'propensity score' (3.9.5.5) methode toegepast.

Relevante bronnen

Het stuk van Faria et al. (2015) behandelt matching in detail.[18]

3.9.5.2 Stratificatie

Bij stratificatie wordt de data in subcategorieën, ook wel strata, verdeeld op basis van de waarde van één of meerdere variabelen. Vervolgens wordt binnen deze strata gekeken naar associaties, puntschattingen en dergelijke. Stratificatie wordt vaak alleen toegepast om effectmodificatie te verkennen, echter is het aan te raden altijd een gestratificeerde analyse te doen. Hoewel het gepaard gaat met verlies aan precisie en soms ook verlies aan informatie (dichetomisatie), geeft het een goed beeld van de data dat verdere analyse kan informeren. In het kader van confounding kan er gestratificeerd worden op de levels van de (mogelijke) confounder.

Relevante bronnen

Tripepi et al. (2010) illustreren het gebruik van stratificatie om voor confounding te corrigeren.[32]

3.9.5.3 Regressie

Een zeer bekende techniek voor het corrigeren voor mogelijke confounders is multivariabele regressie. Hierbij wordt een schatting gemaakt van de associatie tussen de blootstelling en uitkomst onder studie, die gecorrigeerd is voor alle variabelen die in het regressiemodel zijn opgenomen. Het is van belang om de keuze voor variabelen die in het model worden opgenomen goed te onderbouwen (zie ook sectie 2.2.6).

Relevante bronnen

Alexopoulos (2010) geeft een toegankelijke introductie tot multivariabele regressie.[33]

3.9.5.4 Disease risk score

Sommige methoden voor omgang met confounding, waaronder de disease risk and propensity score (sectie 3.9.5.5), maken gebruik van een multivariabele confounder score, waarin informatie over een groot aantal (mogelijke) confounders is samengevat (ook wel confounder summary method genoemd).

De disease risk score kan worden gezien als samenvatting van de relatie tussen deze variabelen en de uitkomst onder studie. Anders gezegd: de disease risk score is de kans op de uitkomst, gegeven deze variabelen. Deze disease risk score wordt vervolgens gebruikt in de analyse (in plaats van de individuele variabelen) of wordt gebruikt om de studiegroepen vergelijkbaar te maken ten aanzien van de kans op het ontwikkelen van de uitkomst (matching).

De disease risk score is gevoelig voor misspecificatie, en er bestaat grote variatie in de manier waarop deze techniek wordt toegepast. Toch lijkt de disease risk score goede resultaten te geven bij zeldzame binaire uitkomsten (weinig 'events' per variabele). Ook is het wellicht gepast deze methode toe te passen binnen studies naar middelen die net op de markt zijn gekomen, omdat er dan vaak sprake is van een snel veranderende gebruikersgroep.

Relevante bronnen

Arbogast en Ray (2009) beschrijven en illustreren het gebruik van disease risk scores en detailleren de constructie ervan.[34] Het artikel van Glynn et al. (2012)

gaat dieper in op de mogelijke rol van disease risk scores binnen comparative effectiveness research.[25]

3.9.5.5 Propensity score

Net als de disease risk score is de propensity score een multivariabele confounder score waarin informatie over een groot aantal (mogelijke) confounders is samengevat. De propensity score is de kans op blootstelling (vaak een behandeling), gegeven deze confounders. Tijdens analyse wordt de propensity score gebruikt in plaats van de individuele variabelen. De propensity score kan ook gebruikt worden om ervoor te zorgen dat confounders evenredig verdeeld zijn over de te vergelijken groepen (te vergelijken met het concept van randomisatie; matching op basis van de score). Echter hoeft dit niet te gelden voor de ongemeten confounders. Een voordeel van het gebruik van propensity scores is dat het mogelijk is om voor een groot aantal covariaten te corrigeren, ook als sample size beperkt is.

Relevante bronnen

Faria et al. (2015) beschrijft de verschillende manieren waarop de propensity score toegepast kan worden.[18] Arbogast en Ray (2011) vergelijken de performance van de disease risk en propensity score met de performance van multivariabele regressie.[35]

3.9.5.6 Marginal structural models

Bij een marginale benadering tot confounding wordt de confounder evenredig verdeeld over alle niveaus van de blootstelling alvorens de relatie tussen de blootstelling en de uitkomst wordt geschat. Bij marginale structurele modellen wordt deze balans verkregen door een gewicht toe te wijzen aan elke observatie. Vervolgens worden de te vergelijken groepen opnieuw gewogen. Alle observaties die ondergerepresenteerd zijn, krijgen een groter gewicht en andersom. Herweging zorgt daarmee voor een gelijke distributie van (mogelijke) confounders over de te vergelijken groepen.

Relevante bronnen

Williamson en Ravani (2017) geven een uitgebreide introductie op marginal structural models.[27]

3.9.6 *Methoden voor omgang met unmeasured confounding*

3.9.6.1 Instrumental variable analyse

Het uitgangspunt van instrumental variable methods is een variabele (instrument genoemd), die voldoet aan de volgende aannames: (1) gecorreleerd met de blootstelling, (2) niet gecorreleerd met andere confounders van de relatie tussen de blootstelling en uitkomst en (3) alleen geassocieerd met de uitkomst via de blootstelling. Vervolgens wordt gekeken naar de associatie tussen de uitkomst en het instrument en de associatie tussen de blootstelling en het instrument. Met behulp van de Wald ratio (een statistische test) wordt vervolgens het zogenaamde "instrumental variable estimate" berekend. De instrumental variable estimate kan gezien worden als schatting van de werkelijke associatie tussen de blootstelling en uitkomst (de associatie gecorrigeerd voor confounding, ook wel gezien als indicatie

voor het wel of niet bestaan van een causale relatie tussen de bestudeerde variabelen).

Het is over het algemeen uitdagend om een instrument te vinden dat aan alle aannames voldoet. Daarbij zijn niet alle aannames volledig te testen. Een veel voorkomend probleem bij instrumental variable analysis is dat het instrument zwak is: er is dan een zwakke associatie tussen het instrument en de blootstelling, ofwel het instrument verklaart maar een klein deel van de variatie in de blootstelling. Als dit het geval is, zal de IV-estimate 'biased' zijn in de richting van de door confounding vertekende associatie tussen de blootstelling en uitkomst.

Relevante bronnen

Brookhart et al. (2010) geven een toegankelijke en praktische introductie tot het gebruik van instrumental variable methoden in het kader van comparative effectiveness research.[36]

3.9.7 *Methoden voor omgang met tijdsafhankelijke confounding*

Bij tijdsafhankelijk confounding adresseren methoden zoals stratificatie, matching en regressie het probleem niet voldoende. Vaak wordt bij tijdsafhankelijke regressie g-estimation, structural nested models of marginal structural models toegepast (zie 3.9.5.6). Marginal structural models heeft hierbij het voordeel over g-estimation in dat (1) het gemakkelijker toe te passen is bij dichotome uitkomsten en (2) meer vergelijkbaar is met standaard statistische modellen.

Relevante bronnen

In Daniel et al. (2013) worden g-computation, marginal structural models en structural nested models in detail behandeld.[37]

3.9.8 *Sensitiviteitsanalyse*

Het doel van de sensitiviteitsanalyse is om inzicht te krijgen in residuele confounding en de mogelijke impact op de studieresultaten te kwantificeren. Dit kan op verschillende manieren benaderd worden. Men kan bijvoorbeeld een simulatie doen waarin de prevalentie van de confounder in de blootgestelde en niet blootgestelde groep en de sterkte van de associatie tussen de confounder en uitkomst worden gevarieerd om een wijde range aan risicoschattingen te verzamelen. Een ander manier is om te kijken hoe sterk de residuele confounding zou moeten zijn om de geobserveerde associatie tussen de blootstelling en uitkomst te verklaren.

Er zijn ook meer algemene manieren om de impact van bias en confounding in een studie te verkennen. Eén daarvan is het verzamelen van additionele uitkomsten die verwacht worden niet geassocieerd te zijn met de behandeling. Als vervolgens op deze uitkomsten een verschil wordt geobserveerd tussen de behandelde en niet behandelde groep, is dit een teken dat er ongemeten confounding is. Het gebruik van meerdere vergelijkende groepen kan ook helpen om vertrouwen in de resultaten te vergroten als à priori is vastgesteld dat de te vergelijken groepen verschillen met betrekking tot een confounder die niet gemeten kan worden. Een voorbeeld is het betrekken van een groep patiënten die niet in een registratie is opgenomen als er reden is om te verwachten dat deze groep van de registerpopulatie verschilt met betrekking tot een belangrijke karakteristiek. Als in beide populaties soortgelijke resultaten worden gevonden, kan er meer vertrouwen

worden gesteld in het resultaat.

Verdere bronnen

In Schneeweiss et al. (2006) worden verschillende manieren beschreven om een sensitiviteitsanalyse uit te voeren om de mogelijke impact van residuele confounding te onderzoeken.[38]

3.10 Missing data

Referenties [13, 23, 39-56] hebben bij uitwerking van dit stuk als uitgangspunt gediend. Aanvullende bronnen voor de lezer worden gegeven onder het kopje "Relevante bronnen".

"Missing data", ofwel het ontbreken van gegevens/waarden in een dataset, is een bekend probleem dat bij allerlei typen onderzoek een rol speelt. Echter is de omvang van het probleem vaak groter binnen registraties. Dit heeft verschillende redenen. Het is binnen registraties bijvoorbeeld ongebruikelijk dat er enige vorm van mandaat is om gegevens vast te leggen. Ook zijn er over het algemeen geen 'periodieke' meetmomenten. Daarnaast kan het combineren van gegevens uit verschillende databronnen de omvang van het missing data probleem binnen een registratie vergroten (bijvoorbeeld als er sprake is van ongelijke registratie).

Aangezien missing data tot vertekening van resultaten kan leiden, is het essentieel dat er goed mee om wordt gegaan. Dit hoofdstuk geeft een inleiding op missing data en introduceert methoden en technieken voor omgang met missing data. Een ander hoofdstuk binnen deze richtlijn gaat in op het rapporteren over missing data (hoofdstuk 4 'Richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming').

3.10.1 Patronen in missing data

Bij missing data is het belangrijk om een beeld te vormen van de omstandigheden die (mogelijk) geleid hebben tot het ontbreken van gegevens in de dataset. Het zogenaamde mechanisme dat missings veroorzaakt, bepaalt namelijk grotendeels welke methoden geschikt zijn om met de missing data om te gaan.

Een eerste stap hierin is om op conceptueel niveau na te denken over mogelijke oorzaken van missing data en een 'hypothese' op te stellen. Een tweede stap is om de data te verkennen en te zien in hoeverre de bevindingen in overeenkomst zijn met deze hypothese. Bij het verkennen van de data kan onder andere gekeken worden naar het patroon van missing data in de database. Veel statische software biedt visualisatieopties voor missing data.

Er zijn drie patronen van missing data:

- (1) univariaat / multivariaat: missing data heeft een univariaat patroon als één variabele missende waarden heeft.
- (2) monotoon: missing data heeft een monotoon patroon als de variabelen zo gerangschikt kunnen worden dat, als er binnen een variabele een missende waarde optreedt, de daaropvolgende waarden binnen die rij ook missing zijn. Dit gebeurt vaak bij longitudinale studies.
- (3) niet-monotoon / arbitrair: als het patroon niet monotoon is, is het niet-monotoon ofwel arbitrair.

3.10.2 *Mechanismen voor missing data*

Missing data wordt geclassificeerd als missing completely at random (MCAR), missing not at random (MNAR) of missing at random (MAR).

3.10.2.1 MCAR – Missing completely at random

Bij MCAR data is de reden dat een waarde mist, 'toeval'. De kans dat een waarde mist, is niet gerelateerd aan andere karakteristieken van de participant. Een typisch voorbeeld van MCAR data is als een bloedmonster in het lab valt zodat er geen bepalingen op gedaan kunnen worden. Omdat de subset van participanten zonder missing data nog steeds een willekeurige subset is van de bronpopulatie, geeft complete case analyse of available case analyse (zie 3.10.3.1) valide resultaten bij MCAR data. Omdat een deel van de data niet wordt gebruikt, verliest de analyse hiermee wel aan precisie.

3.10.2.2 MAR – Missing at random

Bij MAR data is de kans dat een waarde mist, gerelateerd aan andere karakteristieken van de participant die geobserveerd zijn. Een voorbeeld van MAR data is als een longfunctietest alleen bij een subgroep van participanten is gedaan. Omdat de subset van participanten zonder missende waarden geen willekeurige subset is van de bronpopulatie, introduceert het toepassen van complete case analysis (zie 3.9.1.2) hoogstwaarschijnlijk selectiebias. Over het algemeen leidt toepassing van 'simpele' methoden (zie 3.9.3.2) voor omgang met missing data bij MAR tot vertekening van resultaten.

3.10.2.3 MNAR – Missing not at random

Bij MNAR data is de kans dat een waarde mist, gerelateerd aan informatie die niet geobserveerd is, bijvoorbeeld de waarde van de observatie zelf. Een voorbeeld van MNAR data is als participanten gevraagd worden hun gewicht in te vullen en mensen met een hoger gewicht dit veld vaker leeg laten. Er is geen standaard aanpak voor omgang met MNAR data.

3.10.3 *Methoden voor omgang met missing data*

Veel van de hieronder besproken technieken zijn imputatietechnieken, waarbij de missende waarde wordt 'vervangen'. Deze vervangende waarde kan een gemiddelde waarde of mediaan zijn, maar ook een geschatte waarde. Dat het principe van "vervanging" ten grondslag aan imputatietechnieken ligt is logisch als men zich bedenkt dat een uitgangspunt van de klassieke statistiek is dat herhaling van een studie soortgelijke resultaten op zou moeten leveren (en dus onafhankelijk is van de subset van de bronpopulatie die heeft deelgenomen).

3.10.3.1 Complete case analysis & available case analysis

Bij complete case analyse, ook wel listwise deletion genoemd, worden alle participanten met één of meer missende waarden op de reeks van (gebruikte) variabelen in de dataset, verwijderd uit de analyse. Bij available case analyse, ook

wel pairwise deletion genoemd, wordt binnen de analyse gebruik gemaakt van alle beschikbare waarden voor een variabele. Elke puntschatting kan daarmee gebaseerd zijn op een andere subset van participanten, wat interpretatie van resultaten kan compliceren.

Beide technieken kunnen toegepast worden bij MCAR data. De analyse verliest hiermee wel aan precisie omdat een deel van de data niet wordt gebruikt. Bij voorkeur wordt complete case analysis gebruikt omdat goede prestatie van available case analyse nooit helemaal zeker is. De prestatie van available case analyse (ofwel de kans dat toepassing juiste resultaten oplevert) is namelijk afhankelijk van de sterkte van de correlatie tussen de variabelen in de dataset.

Bij MNAR en MAR data kan complete case analyse toegepast worden in de vorm van een sensitiviteitsanalyse, waarbij de conclusies als getrokken op basis van de complete case analyse en de conclusies als getrokken op basis van de analyse waarbij een techniek voor missing data is toegepast, kunnen worden vergeleken.

3.10.3.2 'Simpel' methoden

Verschillende methoden ofwel technieken die vaak worden toegepast bij missing data kunnen geschaard worden onder de noemer 'simpele methoden'. Hieronder valt bijvoorbeeld:

- Substitutie van het gemiddelde of de mediaan, waarbij de missende waarde wordt vervangen door het gemiddelde of de mediaan van alle waarden voor de betreffende variabele (bijvoorbeeld bij cross-sectioneel onderzoek);
- Last observation carried forward/backward, waarbij de missende waarde wordt vervangen door de voorgaande / daar op volgende waarde (herhaalde meting);
- Last & Next, waarbij het gemiddelde van de voorgaande en volgende waarde wordt genomen (herhaalde meting);
- Substitutie op basis van (multivariabele) regressie, waarbij de missende waarde wordt vervangen door de geschatte waarde.
- Deterministische hot deck methoden (zie 3.10.3.5).

Het is af te raden deze methoden te gebruiken omdat toepassing van deze methoden vrijwel altijd tot vertekening van resultaten leidt.

Relevante bronnen

Een toegankelijke introductie tot het de wijze waarop de meer simpele methoden voor omgang met missing data tot vertekende resultaten leiden, is te vinden in Donders et al. (2006).[40]

3.10.3.3 Single en multiple imputation

Vaak wordt met single imputation gerefereerd naar imputatie op basis van stochastische regressie. Bij imputatie op basis van stochastische regressie worden de waarden van de missende observaties geschat aan de hand van (multivariabele) regressie, maar wordt een willekeurige error (een residu) toegevoegd. Daarmee blijft de spreiding in de data gehandhaafd. De standaardfout is wel te klein omdat de onzekerheid over de "echte" waarde van de geïmputeerde observatie niet is meegenomen (er wordt 'gedaan' alsof de observatie niet missing was).

Bij multiple imputation worden meerdere datasets gecreëerd, waarin elke keer

andere waarden zijn geïmputeerd (stochastische regressie) voor de missende observaties. Alle resultaten worden vervolgens gepoold tot één resultaat (puntschatting en standaardfout). Daarmee wordt ook de onzekerheid omtrent de "echte" waarden van de geïmputeerde observaties geborgd. Aannames bij multiple imputation zijn dat de data MAR of MCAR is en dat continue variabelen normaal verdeeld zijn. Als dit niet het geval is, kan de data getransformeerd worden voor imputatie, en na imputatie terug getransformeerd worden. Ook kan predictive mean matching toegepast worden (zie sectie 3.10.3.6). Theoretisch gezien is multiple imputation de beste manier om met missing data om te gaan: correcte toepassing van multiple imputation bij MAR of MCAR data zal over het algemeen tot onvertekende resultaten met juiste standaardfouten leiden. Praktisch gezien kan het echter redelijk ingewikkeld zijn om multiple imputation toe te passen. Er zijn voor non-parametrische testen bijvoorbeeld geen standaard methoden voor de pooling tot één resultaat. (Anders gezegd, er is bijvoorbeeld geen standaard antwoord op de vraag hoe men meerdere medianen poolt).

Relevante bronnen

Het boek flexible imputation of missing data van Van Buuren (2008) biedt een uitgebreide introductie en beschrijving van multiple imputation, geïllustreerd met praktijkvoorbeelden.[41]

3.10.3.4 Maximum – likelihood methods

Maximum likelihood (ML) is een generieke benadering tot schatting van de meest plausibele waarde(s), die binnen de statistiek breed toegepast wordt. Het toepassen van ML methoden bij MAR data is efficiënt en levert onvertekende resultaten op.

Relevante bronnen

In het boek van Allison (2001) worden maximum likelihood methoden beschreven en geïllustreerd met praktijkvoorbeelden.[42]

3.10.3.5 Hot and cold deck imputatie

Bij hot deck imputatie worden missende waarden van de ene participant vervangen door de geobserveerde waarden voor een andere participant in de dataset die vergelijkbaar is met betrekking tot de geobserveerde karakteristieken (ook wel donor genoemd). Er wordt een onderscheid gemaakt tussen random en deterministische hot deck methoden. Bij random hot deck methoden wordt de donor random getrokken uit een set van mogelijke donoren. Bij deterministische hot deck methoden wordt één donor geïdentificeerd. Dit wordt vaak gedaan op basis van het principe van "nearest neighbour". Een voordeel van hot deck methoden is dat alleen plausibele waarden geïmputeerd worden. Het toepassen van hot deck methoden lijkt een goede oplossing te zijn bij het imputeren van missing data in surveys, omdat daar ook rekening gehouden moet worden met eventuele relaties tussen antwoorden. Cold deck methoden zijn vergelijkbaar met hot deck methoden, echter worden de donoren dan uit een andere dataset verkregen.

Relevante bronnen

Een uitgebreide introductie met voorbeelden is te vinden in Andridge & Little (2010).[44]

3.10.3.6 Predictive mean matching

Bij predictive mean matching wordt de geschatte waarde voor een missende observatie berekend volgens een imputatiemodel (zoals hierboven beschreven). Vervolgens wordt er binnen de complete cases een set van mogelijke donoren gevormd. Deze donoren hebben de geschatte waarden die het dichtst liggen bij de geschatte waarde voor de missende observatie. Er wordt een willekeurige donor gekozen. Net als bij hot deck methoden is een voordeel van het toepassen van predictive mean matching dat uiteindelijk alleen plausibele waarden geïmputeerd worden.

3.10.3.7 Verdere methoden.

De hoofdstuk introduceert de meest gangbare methoden voor omgang met MCAR en MAR missing data. Voor verdere methoden, zoals weighting methods en Bayesian methods verwijzen wij naar andere literatuur. Het boek van Allison (2001) bevat een introductie op omgang met MNAR data en kan hiervoor als uitgangspunt worden genomen.

3.10.4 *Sensitiviteitsanalyse*

Naast de voor de hand liggende vergelijking met een analyse gemaakt op complete case basis, kunnen onderzoekers overwegen om een scenario-gebaseerde sensitiviteitsanalyse te doen. Dit kan op basis van de vragen die het meest relevant zijn met betrekking tot de onderzoeksvraag. Een voorbeeld van een worst-case sensitiviteitsanalyse: de aanname dat alle uitkomsten binnen de blootgestelde groep 'events' zijn, en alle uitkomsten van de niet blootgestelde groep 'non-events' of andersom.

3.10.5 *Algemene aanbevelingen*

Relevante uitgangspunten met betrekking tot het werken met missing data zijn:

- Start met het opstellen van een 'hypothese' voor de reden van het optreden van de missing data en het bekijken van patronen in de missing data.
- Pas de 'simpele' technieken niet toe.
- Voer een sensitiviteitsanalyse uit door de uiteindelijke resultaten te vergelijken met de resultaten op basis van een complete case analyse.
- Vertrouw niet blind op de technieken en kijk kritisch naar geïmputeerde waarden;
- Imputeer met een stochastische component (error).

Relevante bronnen

Goede naslagwerken op dit gebied zijn de boeken van Van Buuren (2012) en Allison (2001).[41, 42] Voor het werken met missing data binnen de software R is het artikel over het MICE package van Van Buuren & Groothuis-Oudshoorn (2011) een goed startpunt. [57]

4 Richtlijnen voor rapportage en evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming

Hoofdstuk 4 gaat in op richtlijnen van rapportage over observationele studies en evaluatie van de kwaliteit van observationele studies. We raden aan om te rapporteren volgens het RECORD-PE statement. Daarnaast bevelen wij de GRACE checklist aan voor het evalueren van die observationele studies die gebruikt zijn ter ondersteuning van besluitvorming. Er moet nog nader worden uitgewerkt hoe deze checklist in de praktijk toegepast zullen worden.

Met het toenemende gebruik van observationele data in het kader van besluitvorming, wordt het belang van juiste en complete rapportage over onderliggende studies onderstreept. Ook zijn verschillende checklist, vragenlijsten en tools ontwikkeld om de kwaliteit van observationele studies, die gebruikt worden ter ondersteuning van besluitvorming, te beoordelen.

4.1 Richtlijnen voor rapportage

Het is aan te raden om de STROBE, RECORD en RECORD-PE statements te gebruiken bij rapportage van farmacoepidemiologisch werk binnen een patiëntenregistratie (zie bijlage II).

De strengthening the reporting of observational studies in epidemiology (STROBE) statement is een bekende standaard voor het rapporteren over observationele studies.[58] Omdat specifieke elementen van database studies (waaronder registerstudies), zoals de selectie van de studiepopulatie, binnen STROBE suboptimaal geadresseerd worden, is de reporting of studies conducted using observational routinely-collected health data (RECORD) statement ontwikkeld in verlenging van het STROBE statement.[59] Idealiter worden deze statements samen toegepast bij het rapporteren over een registerstudie. Voor rapportage over farmacoepidemiologisch werk op basis van database studies is in verlenging van STROBE én RECORD the reporting of studies conducted using observational-routinely collected health data statement for pharmacoepidemiology (RECORD-PE) statement ontwikkeld. [60] Hierin wordt bijvoorbeeld ook gevraagd naar rapportage over de definitie van blootstelling aan de behandeling onder studie.

Om vertrouwen in de resultaten van een observationele studies te vergroten, zou consistente registratie van studieprotocollen, waarin de belangrijkste designkeuzes en geplande analyses worden beschreven, een goede ontwikkeling zijn. Er is hier momenteel geen gericht platform voor, echter wordt er sterk aangeraden ontwikkelingen op dit gebied te volgen (bijvoorbeeld van het Real-World Evidence Transparency Initiative [61]).

4.2 Evaluatie van de kwaliteit van observationele studies in het kader van besluitvorming

Het is aan te bevelen om de GRACE checklist te gebruiken voor het beoordelen van de kwaliteit van observationele studies in het kader van besluitvorming (zie bijlage III).

Er kan onderscheid gemaakt worden tussen checklists die specifiek ontwikkeld zijn voor het evalueren van comparative effectiveness studies op basis van observationele data, zoals het werk van Kreif et al. (2013), en meer algemene checklist voor het beoordelen van de kwaliteit van observationele studies in het kader van besluitvorming zoals die ontwikkeld zijn door Berger et al. (2014) en het Good Research practices voor comparative effectiveness (GRACE) initiatief (2014).[62-64]

Faria et al. hebben ook de quality of effectiveness estimates from non-randomised studies (QUEENS) checklist ontwikkeld (2015).[18] Deze checklist is gebaseerd op een set van eerdere, soortgelijke documenten, waaronder de hierboven genoemde. (Hierbij is op basis van thema's gewerkt, items zijn niet één op één overgenomen). Desondanks is de GRACE checklist is aan te bevelen, omdat deze uitgebreid gevalideerd is.

5 Kennislacunes

Dit is een initiële inventarisatie gevormd op basis van referenties [7, 19, 24] een de input van de Expertisegroep. De lijst zal op basis van de uitwerking van de andere hoofdstukken en met behulp van experts op dit gebied worden aangevuld tijdens de case studiefase van het project Regie op Registers voor Dure Geneesmiddelen.

In dit hoofdstuk worden kennishiaten op het gebied van de analyse van gegevens uit patiëntenregistraties, en breder, observationele gegevens ofwel gegevens uit de praktijk, uitgezet. Enerzijds ter bewustwording van de ontwikkelende status van dit veld, anderzijds om richting te geven aan verdere methodiek- en richtlijnontwikkeling.

5.1 Kennislacunes

- Methoden voor non-comparative individual patient data (individual patient data van één enkele groep, bijvoorbeeld alleen een groep blootgesteld aan het geneesmiddel onder studie), onder andere met betrekking tot het gebruik van historische controles bij de beoordeling van een nieuw geneesmiddel;
- Methoden voor geaggregeerde data (wanneer de individual patient data niet beschikbaar is);
- Methoden om de individual patient data van gerandomiseerde en niet-gerandomiseerde studies te combineren;
- Verwerken van de structurele onzekerheid die geassocieerd is met niet-gerandomiseerde studies in decision modelling en het value of information framework;
- Omgang met kleine patiëntenaantallen;
- Minimalisatie van selectiebias;
- Gebruik van data uit de praktijk in netwerk meta analyse;
- Het selecteren van de juiste variabelen voor propensity models.

Bijlage I EunetHTA REQUEST tool

Bron: National Institute for Health and Care Excellence, Croatian Institute of Public Health, Agency for Health Quality and Assessment of Catalonia, French National Authority for Health, 2019. REQUEST Tool. Via <https://eunetha.eu/request-tool-and-its-vision-paper/>

Methodological information

Item number	Area	Item and format required
1	Type of registry	Specify the type of registry which defines the patient population, all the health interventions included in the registry and the registry objectives (primary and secondary).
2	Use for registry-based studies and previous publications	Can the registry be used as a platform for prospective registry-based studies? Provide weblink to publications.
3	Geographical and organisational setting	Specify the geographical area of the registry and organisational setting. List the data providers (type of providers and the number of sites) participating in the registry.
4	Duration	Specify the start and, if relevant, final date of data collection (duration). Indicate if the content (e.g. variables or coding) of the registry has changed in any significant way over time)

5	Size	<p>Provide the total number of patients included in the registry. When was this number calculated?</p> <p>Provide the percentage of the patient population who meet selection criteria and who have participated in the registry.</p>
6	Inclusion and exclusion criteria	<p>List the inclusion and exclusion criteria.</p>
7	Follow-up	<p>Describe the methodology for the follow-up. What is the average follow-up period per patient in months? How do you predict and prevent loss to follow-up?</p>
	Confounders	<p>Are data relating to potential confounders collected and identified for a specific registry use as appropriate. Specify techniques to prevent or control the potential confounders.</p>

Essential Standards

Item number	Area	Minimum standard	Assessment criteria
9	Registry aims and methodology	Registry has stated aims, objectives and methodology.	Registry has specified objectives, target population, exposures of interest, primary and secondary outcomes, data sources, linkage (and analysis plans if any). If the documentation is more than 5 years old, the current status should be checked with the registry coordinator or participant.
10	Governance	Registry governance is in place.	An independent steering committee or a governing body and a data quality team with specified responsibilities are in place. These should include patient representation. Registry governance should have an audited process for declarations of interest covering all financial contributions to the work. Employees of the relevant manufacturers, close relatives who have a position of responsibility within these manufacturing companies or close relatives with financial interests in the capital of these manufacturers could have a declared role in data analysis for the specified HTA project as long as the declared interests are considered not to affect the validity of the data.

<p>11</p>	<p>Informed consent</p>	<p>Protection of privacy rights is assured for the persons whose health-related data is recorded.</p>	<p>The informed consent document should explain to potential participants:</p> <ul style="list-style-type: none"> • the nature, purpose of the registry and whether secondary analyses may be undertaken, • why they are candidates for participating in the registry, • what risks, benefits, and alternatives are associated with the participation • what rights they have as research subjects. <p>If the documentation is more than 5 years old, the current status should be checked with the registry holder.</p>
<p>12</p>	<p>Data dictionary</p>	<p>The data set has a data dictionary or similar.</p>	<p>The data dictionary should contain identifying attributes (name, ID), definitional attribute (definition of data element, where also the purpose of the data element is described), and representational attributes (permissible values, representation class, data type, format).</p> <p>The data dictionary defines terms needed to answer the registry's research questions and objectives.</p> <p>If the documentation is more than 5 years old, the current status should be checked with the registry coordinator or participant.</p>
<p>13</p>	<p>Minimum data set</p>	<p>The registry has a defined minimum data set.</p>	<p>The registry has a defined minimum data set that is able to answer the registry's research questions and objectives. If new fields are required for a specific purpose, the registry is able and willing to make the necessary changes.</p> <p>If the documentation is more</p>

			than 5 years old, the current status should be checked with the registry coordinator or participant.
14	Standard definitions, terminology and specifications	Standard definitions, terminology and specifications are being used.	Name of the standard, category of data (diagnosis, procedure, medication) and usage of the standard (organising, storing, managing or protecting the data sets) should be provided.
15	Data collection	Data collection is described.	Data collection methods are realistic (e.g. software requirements acceptable to submitters) for the proposed population and treating centres with clear access rights.
16	Quality assurance	The registry has a quality assurance plan including assured delivery of continuous and comprehensive data submission.	Quality assurance activities relevant for the registry need to be described.
17	Data cleaning	A plan for cleaning the data is described.	There is a plan for cleaning the data that includes the time required for cleaning after closure to data submission.
18	Missing data	A plan to manage missing data is described.	The percentage of missing data for the core outcomes has been provided. An explanation is given for whether missing data may potentially bias results.

<p>19</p>	<p>Financing</p>	<p>The financing of the registry is transparently presented.</p>	<p>Financial security to the end of the evidence development period should be demonstrated in the financial plan, solvency with a summary of income and expenditure for the previous 2 years is recommended. Also, funding sources are identified and the approx. proportions (%) of total sum from each funding source is indicated. If the documentation is more than 5 years old, the current status should be checked with registry coordinator or participant.</p>
<p>20</p>	<p>Protection, security and safeguards</p>	<p>Data security risks, policies and procedures are described. The registry has a policy for data sharing.</p>	<p>The security controls specific for the registry should be specified. Risks should be identified and appropriate mitigation described.</p>

Additional Requirements

Item number	Area	Standard/requirement and format of the requested data	Assessment criteria
21	Interoperability and readiness for data linkage	Registry shares data with interested parties. 1. Document file format 2. Free text 3. Free text 4. Free text 5. Free text	1. Data access and sharing procedures documentation are uploaded. 2. Average time to answer an information query and to undertake data linkage is specified. 3. Statements regarding usage of data and consent for data sharing (and with whom) are provided. These cover sharing of registry data with interested parties from other countries and/or international organisations. 4. Technical standards, data structure, and standard sets for measuring health outcomes and Internationally agreed minimum data set is specified.

			5. Existence of specific fees in providing access to data and data linkage are clarified.
22	Data sources	All data sources are identified. Predefined, multiple choices. If 'Other' - Free text	Sources of data are identified. If the sources of data for the registry are not listed, please select 'Other' and describe the data sources in text.
23	Ethics	An ethical committee is involved in governing the usage of data. Free text	Consideration of research ethics requirements has been reported. If a research ethics committee approved the working procedures/ methodology of the registry, the process of obtaining approval is described.

Bijlage II: RECORD-PE statement

Bron: Langan et al. 2018. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*; 363: k3532

Item No	STROBE items	RECORD items	RECORD-PE items
1	<p>(a) Indicate the study's design with a commonly used term in the title or the abstract.</p> <p>(b) Provide in the abstract an informative and balanced summary of what was done and what was found.</p>	<p>1.1: The type of data used should be specified in the title or abstract. When possible, the name of the databases used should be included.</p> <p>1.2: If applicable, the geographical region and timeframe within which the study took place should be reported in the title or abstract.</p> <p>1.3: If linkage between databases was conducted for the study, this should be clearly stated in the title or abstract.</p>	—
2	Explain the scientific background and rationale for the investigation being reported.	—	—
3	State specific objectives, including any prespecified hypotheses.	—	—
4	Present key elements of study design early in the paper.	—	<p>4.a: Include details of the specific study design (and its features) and report the use of multiple designs if used.</p> <p>4.b: The use of a</p>

Item No	STROBE items	RECORD items	RECORD-PE items
			<p>diagram(s) is recommended to illustrate key aspects of the study design(s), including exposure, washout, lag and observation periods, and covariate definitions as relevant.</p>
5	<p>Describe the setting, locations, and relevant dates, including periods of recruitment, exposure, follow-up, and data collection.</p>	—	—
6	<p>(a) Cohort study—give the eligibility criteria, and the sources and methods of selection of participants. Describe methods of follow-up. Case-control study—give the eligibility criteria, and the sources and methods of case ascertainment and control selection. Give the rationale for the choice of cases and controls. Cross sectional study—give the eligibility criteria, and the sources and methods of selection of participants. (b) Cohort study—for matched studies, give matching criteria and number of exposed and unexposed. Case-control study—for</p>	<p>6.1: The methods of study population selection (such as codes or algorithms used to identify participants) should be listed in detail. If this is not possible, an explanation should be provided. 6.2: Any validation studies of the codes or algorithms used to select the population should be referenced. If validation was conducted for this study and not published elsewhere, detailed methods and results should be provided. 6.3: If the study involved linkage of databases, consider use of a flow diagram or other graphical display to demonstrate the data</p>	<p>6.1.a: Describe the study entry criteria and the order in which these criteria were applied to identify the study population. Specify whether only users with a specific indication were included and whether patients were allowed to enter the study population once or if multiple entries were permitted. See explanatory document for guidance related to matched designs.</p>

Item No	STROBE items	RECORD items	RECORD-PE items
	<p>matched studies, give matching criteria and the number of controls per case.</p>	<p>linkage process, including the number of individuals with linked data at each stage.</p>	
7	<p>Clearly define all outcomes, exposures, predictors, potential confounders, and effect modifiers. Give diagnostic criteria, if applicable.</p>	<p>7.1: A complete list of codes and algorithms used to classify exposures, outcomes, confounders, and effect modifiers should be provided. If these cannot be reported, an explanation should be provided.</p>	<p>7.1.a: Describe how the drug exposure definition was developed. 7.1.b: Specify the data sources from which drug exposure information for individuals was obtained. 7.1.c: Describe the time window(s) during which an individual is considered exposed to the drug(s). The rationale for selecting a particular time window should be provided. The extent of potential left truncation or left censoring should be specified. 7.1.d: Justify how events are attributed to current, prior, ever, or cumulative drug exposure. 7.1.e: When examining drug dose and risk attribution, describe how current, historical or time on therapy are considered. 7.1.f: Use of any comparator groups should be outlined</p>

Item No	STROBE items	RECORD items	RECORD-PE items
			and justified. 7.1.g: Outline the approach used to handle individuals with more than one relevant drug exposure during the study period.
8	For each variable of interest, give sources of data and details of methods of assessment (measurement). Describe comparability of assessment methods if there is more than one group.	—	8.a: Describe the healthcare system and mechanisms for generating the drug exposure records. Specify the care setting in which the drug(s) of interest was prescribed.
9	Describe any efforts to address potential sources of bias.	—	—
10	Explain how the study size was arrived at.	—	—
11	Explain how quantitative variables were handled in the analyses. If applicable, describe which groupings were chosen, and why.	—	—
12	(a) Describe all statistical methods, including those used to control for confounding. (b) Describe any methods used to examine subgroups and interactions. (c) Explain how missing data were addressed.	—	12.1.a: Describe the methods used to evaluate whether the assumptions have been met. 12.1.b: Describe and justify the use of multiple designs, design features, or analytical approaches.

Item No	STROBE items	RECORD items	RECORD-PE items
	<p>(d) Cohort study—if applicable, explain how loss to follow-up was addressed. Case-control study—if applicable, explain how matching of cases and controls was addressed. Cross sectional study—if applicable, describe analytical methods taking account of sampling strategy.</p> <p>(e) Describe any sensitivity analyses.</p>		
12	—	<p>12.1: Authors should describe the extent to which the investigators had access to the database population used to create the study population.</p> <p>12.2: Authors should provide information on the data cleaning methods used in the study.</p>	—
12	—	<p>12.3: State whether the study included person level, institutional level, or other data linkage across two or more databases. The methods of linkage and methods of linkage quality evaluation should be provided.</p>	—
13	<p>(a) Report the numbers of individuals at each stage of the study (eg, numbers potentially</p>	<p>13.1: Describe in detail the selection of the individuals included in the study (that is, study population selection)</p>	—

Item No	STROBE items	RECORD items	RECORD-PE items
	<p>eligible, examined for eligibility, confirmed eligible, included in the study, completing follow-up, and analysed).</p> <p>(b) Give reasons for non-participation at each stage.</p> <p>(c) Consider use of a flow diagram.</p>	<p>including filtering based on data quality, data availability, and linkage. The selection of included individuals can be described in the text or by means of the study flow diagram.</p>	
14	<p>(a) Give characteristics of study participants (eg, demographic, clinical, social) and information on exposures and potential confounders.</p> <p>(b) Indicate the number of participants with missing data for each variable of interest.</p> <p>(c) Cohort study—summarise follow-up time (eg, average and total amount).</p>	—	—
15	<p>Cohort study—report numbers of outcome events or summary measures over time.</p> <p>Case-control study—report numbers in each exposure category, or summary measures of exposure.</p> <p>Cross sectional study—report numbers of outcome events or summary measures.</p>	—	—

Item No	STROBE items	RECORD items	RECORD-PE items
16	<p>(a) Give unadjusted estimates and, if applicable, confounder adjusted estimates and their precision (eg, 95% confidence intervals). Make clear which confounders were adjusted for and why they were included.</p> <p>(b) Report category boundaries when continuous variables are categorised.</p> <p>(c) If relevant, consider translating estimates of relative risk into absolute risk for a meaningful time period.</p>	—	—
17	Report other analyses done—eg, analyses of subgroups and interactions, and sensitivity analyses.	—	—
18	Summarise key results with reference to study objectives.	—	—
19	Discuss limitations of the study, taking into account sources of potential bias or imprecision. Discuss both direction and magnitude of any potential bias.	19.1: Discuss the implications of using data that were not created or collected to answer the specific research question(s). Include discussion of misclassification bias, unmeasured confounding, missing data, and changing eligibility over time, as they pertain to the study being reported.	19.1.a: Describe the degree to which the chosen database(s) adequately captures the drug exposure(s) of interest.

Item No	STROBE items	RECORD items	RECORD-PE items
20	Give a cautious overall interpretation of results considering objectives, limitations, multiplicity of analyses, results from similar studies, and other relevant evidence.	—	20.a: Discuss the potential for confounding by indication, contraindication or disease severity or selection bias (healthy adherer/sick stopper) as alternative explanations for the study findings when relevant.
21	Discuss the generalisability (external validity) of the study results.	—	—
22	Give the source of funding and the role of the funders for the present study and, if applicable, for the original study on which the present article is based.	—	—
22	—	22.1: Authors should provide information on how to access any supplemental information such as the study protocol, raw data, or programming code.	—

Bijlage III: Grace Checklist

Bron: Dreyer et al. 2016. The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness. *Journal of Managed Care and Specialty Pharmacy*; 22(10):1107-13. Via <https://www.graceprinciples.org/>

Component Item	Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)
Data	
D1. Were treatment and/or important details of treatment exposure adequately recorded for the study purpose in the data source(s)? Note: not all details of treatment are required for all research questions.	(+) Yes, reasonably necessary information to determine treatment or intervention was adequately recorded for study purposes (e.g., for drugs, sufficient detail on dose, days supplied, route, or other important data. For vaccines, consider the importance of batch, dose, route, and site of administration, etc. For devices, consider type of device, placement, surgical procedure used, serial number, etc.). (-) No, data source clearly deficient, <i>or</i> not enough information in article.
D2. Were the primary outcomes adequately recorded for the study purpose (e.g., available in sufficient detail through data sources)?	(+) Yes, information to ascertain outcomes were adequately recorded in the data source (e.g., if clinical outcomes were ascertained using ICD-9-CM diagnosis codes in an administrative database, the level of sensitivity and specificity captured by the codes were sufficient for assessing the outcome of interest). (-) No, data source clearly deficient (e.g., the codes captured a range of conditions that was too broad or narrow, and supplementary information such as that from medical charts was not available), <i>or</i> not enough information in article.
D3. Was the primary clinical outcome(s) measured objectively rather than subject to clinical judgment (e.g., opinion about whether the patient's condition has improved)?	(+) Yes, clinical outcomes were measured objectively (e.g., hospitalization, mortality). (+) Not applicable; primary outcome not clinical (e.g., PROs). (-) No (e.g., clinical opinion about whether patient's condition improved) <i>or</i> not enough information in article.

Component Item	Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)
D4. Were primary outcomes validated, adjudicated, or otherwise known to be valid in a similar population?	(+) Yes, outcomes were validated, adjudicated, or based on medical chart abstractions with clear definitions (e.g., a validated instrument was used to assess patient-reported outcomes [e.g., SF-12 Health Survey]; a clinical diagnosis via ICD-9-CM code was used, with formal medical record adjudication by committee to confirm diagnosis or other procedures to achieve reasonable sensitivity and specificity; and billing data were used to assess health resource utilization). (-) No, <i>or</i> not enough information in article.
D5. Was the primary outcome(s) measured or identified in an equivalent manner between the treatment/intervention group and the comparison group?	(+) Yes. (-) No, <i>or</i> not enough information in article.
D6. Were important covariates that may be known confounders or effect modifiers available and recorded? Important covariates depend on the treatment and/or outcome of interest (e.g., body mass index should be available and recorded for studies of diabetes; race should be available and recorded for studies of hypertension and glaucoma).	(+) Yes, most if not all important known confounders and effect modifiers available and recorded (e.g., measures of medication dose and duration). (-) No, at least 1 probable known confounder or effect modifier not available and recorded (as noted by authors or as determined by user's clinical knowledge), <i>or</i> not enough information in article.
Methods	
M1. Was the study (or analysis) population restricted to new initiators of treatment or those starting a new course of treatment? Efforts to include only new initiators may include restricting the cohort to those who had a washout period (specified period of medication nonuse) before the beginning of study follow-up.	(+) Yes, only new initiators of the treatment of interest were included in the cohort, or for surgical procedures and devices, including only patients who never had the treatment before the start of study follow-up. (-) No, <i>or</i> not enough information in article.
M2. If 1 or more comparison groups were used, were they concurrent comparators? If not, did the authors justify the use of historical comparison groups?	(+) Yes, data were collected during the same time period as the treatment group (concurrent), or historical comparators were used with reasonable justification (e.g., when it is impossible for researchers to identify current users of older treatments or when a concurrent comparison group is not valid, as when uptake of new product is so rapid that concurrent comparators differ greatly on factors related to the outcome). (-) No, historical comparators used

Component Item	Scoring as Fit for Purpose: Sufficient (+), Insufficient (-)
<p>M3. Were important confounding and effect-modifying variables taken into account in the design and/or analysis? Appropriate methods to take these variables into account may include restriction, stratification, interaction terms, multivariate analysis, propensity score matching, instrumental variables, or other approaches.</p>	<p>without being scientifically justifiable, <i>or</i> not enough information in article.</p> <p>(+) Yes, most if not all important covariates that would be likely to change the effect estimate substantially were accounted for (e.g., measures of medication dose and duration).</p> <p>(-) No, some important covariates were available for analysis but not analyzed appropriately, <i>or</i> at least 1 important covariate was not measured, <i>or</i> not enough information in article.</p>
<p>M4. Is the classification of exposed and unexposed person-time free of "immortal time bias," i.e., "immortal time" in epidemiology refers to a period of cohort follow-up time during which death (or an outcome that determines end of follow-up) cannot occur.</p>	<p>(+) Yes.</p> <p>(-) No, <i>or</i> not enough information in the article.</p>
<p>M5. Were any meaningful analyses conducted to test key assumptions on which primary results are based (e.g., were some analyses reported to evaluate the potential for a biased assessment of exposure or outcome, such as analyses where the impact of varying exposure and/or outcome definitions was tested to examine the impact on results)?</p>	<p>(+) Yes, and primary results did not substantially change.</p> <p>(-) Yes, and primary results changed substantially.</p> <p>(-) None reported, <i>or</i> not enough information in article.</p>

Referenties

- [1] Zorginstituut Nederland. (2016) Richtlijn voor het uitvoeren van economische evaluaties in de gezondheidszorg. Via <https://www.zorginstituutnederland.nl/over-ons/publicaties/publicatie/2016/02/29/richtlijn-voor-het-uitvoeren-van-economische-evaluaties-in-de-gezondheidszorg>
- [2] The cross-committee task force on patient registries. (2018). Discussion paper: Use of patient disease registries for regulatory purposes – methodological and operational considerations.
- [3] Innovative Medicines Initiative (IMI) GetReal. Advancing Evidence Generation for New Drugs: IMI GetReal's Recommendations on Real-World Evidence. Via <https://www.imi-getreal.eu/>
- [4] National Institute for Health and Care Excellence, Croatian Institute of Public Health, Agency for Health Quality and Assessment of Catalonia, French National Authority for Health, 2019. REQUEST Tool. Via <https://eunetha.eu/request-tool-and-its-vision-paper/>
- [5] Kodra Y et al. (2018). Recommendations for Improving the Quality of Rare Disease Registries. *Int J Environ Res Public Health*; 15: 1644.
- [6] Glicklich et al. (2012). Standards in the Conduct of Registry Studies for Patient-Centered Outcomes Research - A Guidance Document for the Patient-Centered Outcomes Research Institute. Via <https://www.pcori.org/assets/Standards-in-the-Conduct-of-Registry-Studies-for-Patient-Centered-Outcomes-Research.pdf>.
- [7] The European Network of Centres for Pharmacoepidemiology and Pharmacovigilance (ENCePP). Guide on Methodological Standards in Pharmacoepidemiology (Revision 7). (2018). EMA/95098/2010. Via http://www.encepp.eu/standards_and_guidances
- [8] U.S. Department of Health and Human Services Food and Drug Administration. (2013). Best Practices for Conducting and Reporting Pharmacoepidemiologic Safety Studies Using Electronic Healthcare Data. Via <https://www.fda.gov/media/79922/download>.
- [9] Zalatel & Kralj (Eds). (2015) Methodological guidelines and recommendations for efficient and rational governance of patient registries. National Institute of Public Health, Slovenia. Via https://ec.europa.eu/health/sites/health/files/ehealth/docs/patient_registries_guidelines_en.pdf.
- [10] ANQ, FMS, H+, SAMS & University Medicine Switzerland. (2016) Recommendations for the development and operation of health-related registries. Via https://www.anq.ch/fileadmin/redaktion/deutsch/20160926_Empfehlungen_Register_final_en.pdf.
- [11] MedTech Europe. (2017). Medical Technology Registries Six Key Principles. Via <https://www.medtecheurope.org/wp-content/uploads/2017/07/Medical-Technology-Registries-Six-Key-Principles-170509.pdf>
- [12] Wilson et al. (2015). Review Body for Interventional Procedures Programme (ReBIP) - Evaluating Databases. Via https://www.sheffield.ac.uk/polopoly_fs/1.43797!/file/Evaluating-Databases-15-March-2006.pdf.
- [13] Agency for Healthcare Research and Quality (AHRQ). (2014). Registries for evaluating patient outcomes: A user's guide. Via <https://effectivehealthcare.ahrq.gov/products/registries-guide-4th-edition/prior-editions>
- [14] Australian Commission on Safety and Quality in Health Care. (2014) Framework for Australian clinical quality registries. Via

<https://www.safetyandquality.gov.au/our-work/national-arrangements-clinical-quality-registries>.

- [15] Agency for Healthcare Research and Quality. (2013). Developing a Protocol for Observational Comparative Effectiveness Research: A User's Guide. Via <https://effectivehealthcare.ahrq.gov/products/observational-cer-protocol/research>
- [16] Schneeweiss et al. 2019 Graphical depiction of longitudinal study designs in health care databases. *Annals of internal medicine*; 170: 398 – 406.
- [17] Wang et al. (2017). Reporting to Improve Reproducibility and Facilitate Validity Assessment for Healthcare Database Studies V1.0. *Pharmacoepidemiological Drug Safety*; 26(9): 1018 - 1032.
- [18] Faria et al. (2015) NICE DSU Technical Support Document 17: The use of observational data to inform estimates of treatment effectiveness for Technology Appraisal: Methods for comparative individual patient data. Via <http://www.nicedsu.org.uk>
- [19] Berger ML et al. (2012). Prospective observational studies to assess comparative effectiveness: the ISPOR Good Research Practices Task Force Report. *Value in Health*; 15(2): 217 - 230.
- [20] Berger ML et al. (2017). Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. *Value in Health*; 26: 1033 – 1039.
- [21] Berger ML et al. (2009). Good research practices for comparative effectiveness research: defining, reporting and interpreting nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part I. *Value in Health*; 12(8): 1044 - 1052.
- [22] Cox E et al. (2009). Good research practices for comparative effectiveness research: approaches to mitigate bias and confounding in the design of nonrandomized studies of treatment effects using secondary data sources: the International Society for Pharmacoeconomics and Outcomes Research Good Research Practices for Retrospective Database Analysis Task Force Report--Part II. *Value in Health*; 12(8): 1053 - 1061.
- [23] Johnson ML et al. (2009). Good research practices for comparative effectiveness research: analytic methods to improve causal inference from nonrandomized studies of treatment effects using secondary data sources: the ISPOR Good Research Practices for Retrospective Database Analysis Task Force Report--Part III. *Value in Health*; 12(8): 1063-1073.
- [24] De Graaf et al. (2011). Matching, an appealing method to avoid confounding? *Kidney disease and population health*; 118: c315 – c318.
- [25] Glynn et al. (2012). Role of disease risk scores in comparative effectiveness research with emerging therapies. *Pharmacoepidemiology and drug safety*; 21(2): 138 – 147.
- [26] Tadrous et al. (2013). Disease Risk Score (DRS) as a Confounder Summary Method: Systematic Review and Recommendations. *Pharmacoepidemiology and drug safety*; 22(2): 122 - 129.
- [27] Williamson & Ravani. (2017). Marginal structural models in clinical research: when and how to use them? *Nephrology Dialysis Transplantation*; 32(2): ii84 – ii90.
- [28] Pazzagli et al. (2017). Methods for time-varying exposure related problems in pharmacoepidemiology: An overview. *Pharmacoepidemiology and drug safety*; 27(2).
- [29] McMahon. (2003). Approaches to combat with confounding by indication in observational studies of intended drug effects. *Pharmacoepidemiology and drug safety*; 12(7): 551 – 558.
- [30] Fewell et al. 2007. The impact of residual and unmeasured confounding in epidemiologic studies: a simulation study. *American journal of epidemiology*;

166(6): 646 – 655.

[31] Schneeweiss et al. (2010). Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results. *Medical care*; 45(10): s131 – s142.

[32] Tripepi et al. (2010). Stratification for confounding – part 1: the Mantel-Haenzel formula. *Nephron clinical practice*; 116(4): c317 – c321.

[33] Alexopoulos. (2010). Introduction to multivariate regression analysis. *Hippokratia*; 14(1): 23 – 28.

[34] Arbogast & Ray. (2009). Use of disease risk scores in pharmacoepidemiological studies. *Statistical methods in medical research*; 18(1), 67 – 80.

[35] Arbogast & Ray. (2011). Performance of disease risk scores, propensity scores, and traditional multivariable outcome regression in the presence of multiple confounders. *American journal of epidemiology*; 174(5): 613 – 620.

[36] Brookhart et al. (2010) Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiology and drug safety*; 19(6): 537 – 554.

[37] Daniel et al. (2013) Methods for dealing with time-dependent confounding. *Statistics in medicine*; 32(9): 1584 – 1618.

[38] Schneeweiss. (2006). Sensitivity analysis and external adjustment for unmeasured confounders in epidemiologic database studies of therapeutics. *Pharmacoepidemiology and drug safety*; 15(5).

[39] Mack et al. (2018) Managing Missing Patient Data in Patient Registries. White Paper, addendum to Registries for Evaluating Patient Outcomes: A User's Guide, Third Edition. AHRQ Publication No. 17(18)-EHC015-EF. Rockville, MD: Agency for Healthcare Research and Quality. Via www.effectivehealthcare.ahrq.gov.

[40] Donders et al. (2006). Review: a gentle introduction to imputation of missing values. *Journal of clinical epidemiology*; 59: 1087-1091

[41] Van Buuren. (2018). Flexible imputation of missing data. Second edition. Chapman and Hall / CRC.

[42] Allison. (2001). Missing data. Sage Publications Inc.

[43] Van Kuijk. (2017). Missing data: van gatenkaas naar valide uitkomsten. Maastricht UMC+.

[44] Andridge & Little. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review*; 78(1): 40 – 64.

[45] OECD. (2013). Glossary of statistical terms: deck imputation. Via <https://stats.oecd.org/glossary/detail.asp?ID=3379>

[46] Montaquila & Ponikowski. (1995). An evaluation of alternative imputation methods. Proceedings of the Section on Survey Research Methods, American Statistical Association.

[47] Rubin DB. (1987). Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons.

[48] Greenland et al. (1995). A critical look at methods for handling missing covariates in epidemiologic regression analyses. *American journal of epidemiology*; 142(12): 1255 – 1264.

[49] Hernan et al. (2002). Causal knowledge as a prerequisite for confounding evaluation: an application to birth defects epidemiology. *American journal of epidemiology*; 155(2): 176 - 184.

[50] Lash et al. (2009). Applying quantitative bias analysis to epidemiologic data. Springer.

[51] Pigott et al. (2001). A review of methods for missing data. *Educational research and evaluation*; 7(4): 353 – 383.

[52] Batista & Monard. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied artificial intelligence*; 17(5-6): 519 – 533.

[53] Dong & Peng. (2013). Principled missing data methods for researchers. *Springerplus*; 2(222).

- [54] Horton & Kleinman. (2007). Much ado about nothing: a comparison of missing data methods and software to fit incomplete data regression models. *The American statistician*; 61(1): 79 – 90.
- [54] Schafer & Graham. (2002). Missing data: our view of the state of the art. *Psychological methods*; 7(2): 147-77.
- [56] Engels & Diehr. 2003. Imputation of missing longitudinal data: a comparison of methods. *Journal of clinical epidemiology*; 56(10): 968 – 976.
- [57] Van Buuren & Groothuis-Oudshoorn. (2011). MICE: multivariate imputation by chained equations in R. *Journal of statistical software*; 45(3).
- [58] Von Elm E et al. (2007). The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) Statement: guidelines for reporting observational studies. *PloS medicine*; 4(10): e296.
- [59] Benchimol et al. (2015). The REporting of studies Conducted using Observational Routinely-collected health Data (RECORD) Statement. *PLoS Medicine*; 12(10): e1001885.
- [60] Langan et al. 2018. The reporting of studies conducted using observational routinely collected health data statement for pharmacoepidemiology (RECORD-PE). *BMJ*; 363: k3532.
- [61] International society for pharmacoconomics and outcomes research. (2020). Real-world evidence transparency initiative. Via <https://www.ispor.org/strategic-initiatives/real-world-evidence/real-world-evidence-transparency-initiative>
- [62] Kreif et al. (2013) Statistical methods for cost-effectiveness analyses that use observational data: a critical appraisal tool and review of current practice. *Health economics*; 22(4): 486 – 500.
- [63] Berger ML et al. (2014). A Questionnaire to Assess the Relevance and Credibility of Observational Studies to Inform Health Care Decision Making: An ISPOR-AMCP-NPC Good Practice Task Force Report. *Value Health*; 17: 143 - 156.
- [64] Dreyer et al. 2016. The GRACE Checklist: A Validated Assessment Tool for High Quality Observational Studies of Comparative Effectiveness. *Journal of Managed Care and Specialty Pharmacy*; 22(10): 1107 - 1113. Via <https://www.graceprinciples.org/>