

# Onderzoeksmethodologie 1.

## EPIDEMIOLOGIE.

Onder een *epidemie* verstaat men een opmerkelijke toename in de mate van vóórkomen van bepaalde ziekten of ziekteverschijnselen in en bepaald tijdvak (weken, maanden, jaren, decennia).

De nieuwe frequentie van de ziekteverschijnselen wordt dus impliciet vergeleken met het tot dan toe gebruikelijke achtergrondniveau.

Wanneer ziekten of ziekteverschijnselen gedurende langere tijd op een constant hoog frequentieniveau in een populatie aanwezig zijn, spreekt men van een *endemie*.

Kenmerken epidemiologie:

- In de epidemiologie vormt ziekte de centrale variabele Het object van onderzoek in de epidemiologie is ziekte bij mensen.
- In de epidemiologie gaat het om het vóórkomen van ziekte in menselijke populaties.
- De epidemiologie beschouwt het voorkomen van ziekte veelal in relatie tot het voorkomen van andere verschijnselen. Het gaat daarbij om factoren waarvan men vermoedt dat ze hetzij van invloed zijn op het ontstaan van de desbetreffende ziekte (*etiologische factoren*), hetzij op een andere wijze een indicatie geven van de aanwezigheid van de ziekte in kwestie (*diagnostische factoren*), hetzij samenhangen met het verloop van de ziekte (*prognostische factoren*).

Ieder individu uit de onderzoekspopulatie wordt in epidemiologisch onderzoek gezien als een vertegenwoordiger van een groep individuen met (gedeeltelijk) dezelfde kenmerken. Het gaat om het vaststellen van de ziektefrequentie per groep en van verschillen in ziektefrequentie tussen groepen met verschillende kenmerken. Per individu wordt vastgesteld of de ziekte al dan niet aanwezig is en vervolgens wordt geteld hoeveel zieke individuen er in de totale groep zijn. Aldus ontstaat de epidemiologische breuk of epidemiologische fractie;

Aantal zieke individuen/totaal aantal personen in de groep waaruit deze zieke individuen afkomstig zijn.

De breuk krijgt het karakter van een *incidentiemaat* indien in een groep gedurende een bepaalde periode het aantal nieuwe gevallen van of nieuwe personen met de desbetreffende ziekte wordt geteld.

De breuk krijgt het karakter van een *prevalentiemaat* indien in een groep op een bepaald moment of gedurende een bepaalde periode het aantal bestaande gevallen van of aanwezige personen met een ziekte wordt geteld.

Frequentiematen, in het bijzonder incidentiematen, zijn tevens *risicomaten*: ze geven voor de mensen die tot de desbetreffende groep behoren aan wat de kans is op het krijgen van een ziekte.

Personen die daadwerkelijk kans hebben om de ziekte te krijgen worden de *population at risk* genoemd.

Het in kaart brengen van het distributiepatroon van een ziekte behoort tot het domein van de beschrijvende epidemiologie. De tak van de epidemiologie die de oorzakelijke factoren die aan deze verspreiding van ziekte ten grondslag liggen tracht te identificeren, wordt de verklarende of analytische epidemiologie genoemd.

Er zijn 3 verschillende kenmerken die gerelateerd zijn aan het voorkomen van een ziekte: etiologische, diagnostische en prognostische.

*Etiologische factoren:* determinanten die in oorzakelijke zin (mede) verantwoordelijk zijn voor het ontstaan van een ziekte.

*Prognostische factoren:* determinanten die, zodra de ziekte eenmaal in gang is gezet, van invloed zijn op het verloop van het ziekteproces.

*Diagnostische factoren:* factoren die onderscheiden welke personen wel en welke niet een bepaalde ziekte hebben.

De factoren die het ontstaan of het beloop van een ziekte beïnvloeden zijn: genen, gedrag en omgeving.

$$Z=f(D_i)$$

Deze epidemiologische functie – in statistische termen een regressievergelijking– stelt dat het voorkomen van de ziekte (Z) een functie is van een serie van k determinanten ( $D_i$ , met  $i=1, \dots, k$ ). In deze vergelijking is de ziekte de afhankelijke variabele of uitkomstvariabele. De determinanten fungeren in de formule als onafhankelijke variabelen.

Sommige associatie maten zijn gebaseerd op het berekenen van het verschil in ziektefrequentie tussen personen die wel en personen die niet aan de determinant in kwestie zijn blootgesteld (of tussen personen die zijn blootgesteld aan verschillende waarden of categorieën van de determinant). Dit zijn de *verschilmaten*. Andere associatiematen zijn gebaseerd op het berekenen van de ratio van de ziektefrequenties van de personen die wel en de personen die aan de determinant zijn blootgesteld (of van personen die aan verschillende niveaus van de determinant zijn blootgesteld). Dit zijn de *verhoudingsmaten*.

*Effectmodificatoren:* een effectmodifier is een determinant die de relatie tussen de centrale determinant en de ziekte kan modificeren (wijzigen). Dit betekent dat de associatiemaat die gebruikt wordt om de relatie tussen de centrale determinant en de bestudeerde uitkomst te kwantificeren (bijvoorbeeld het relatief risico), andere uitkomsten te zien geeft voor verschillende waarden van de effectmodifier.

Een *confounder* is een *verstorende variabele*, die verantwoordelijk is voor een vertekende weergave van de relatie tussen de centrale determinant en de ziekte.

Confounding is te herleiden tot het gegeven dat in een populatie verscheidende determinanten naast elkaar opereren.

$$Z=f_{mi}(D|C_i)$$

Deze vergelijking zegt het volgende. Het voorkomen van ziekte (Z) is een functie (f) van de centrale determinant (D). Het gaat erom de sterkte van de relatie tussen Z en D te meten, rekening houdend met (na correctie voor) eventuele confounders ( $|C_i$ ). De sterkte van de relatie kan verschillen, afhankelijk van de waarden die weer andere variabelen, modificatoren ( $M_i$ ) van het effect van de centrale determinant op de ziekte, aannemen.

Een onderzoek noemt men geldig (*valide*) als de schatting van de associatie tussen de centrale determinant en de ziekte uit het onderzoek overeenkomt met de werkelijke associatie. Fouten in het onderzoek resulteren in *bias*: een vertekende weergave van de werkelijke associatie (gebrek aan validiteit).

Verschillende manieren waarop *bias* in een epidemiologisch onderzoek ken worden geïntroduceerd:

- *Selectiebias*: vertekening als gevolg van fouten bij de samenstelling van de groepen die men wil vergelijken om inzicht te krijgen in de relatie tussen centrale determinant en ziekte.
- *Confounding*: vertekening als gevolg van het feit dat onvoldoende wordt afgerekend met het versturende effect van andere determinanten (confounders) op de relatie tussen centrale determinanten en ziekte.
- *Informatiebias*: vertekening als gevolg van fouten bij de meting van de onderzoeksvariabelen (determinantstatus, ziektestatus) bij de leden van de onderzoekspopulatie.

Gaat het bij validiteit om een zo goed mogelijke overeenkomst tussen de gevonden associatie en de werkelijke associatie tussen determinanten en ziekte, *precisie* houdt in dat het aantal waarnemingen voldoende groot is om een toevalsbevinding uit te sluiten.

De validiteit wordt gewaarborgd door een goede operationalisatie van het te meten concept tot een concrete meetprocedure en door een correcte uitvoering hiervan. De reproduceerbaarheid (precisie) hangt af van de mate waarin herhaalde metingen van hetzelfde concept bij dezelfde individuen dezelfde meetuitkomsten te zien geven.

Onderzoekers kunnen de resultaten van hun onderzoek:

- Bevestigen (*confirmatie*)
- Tegenspreken (*contradictie*)
- Weerleggen (*falsificatie*)
- Specificeren (*elaboratie*)

De *beschrijvende epidemiologie* houdt zich bezig met de bestudering van het vóórkomen van ziekten, al of niet in relatie tot bepaalde kenmerken.

De *verklarende epidemiologie* zoekt naar factoren die het voorkomen van ziekten kunnen verklaren.

*Diagnostische factoren* zeggen iets over de aanwezigheid van de ziekte, maar behoeven geenszins causaal met het ontstaan of de prognose van de ziekte te zijn verbonden.

Bij diagnostisch onderzoek zijn geen confounders of effectmodificatoren.

De kans dat een ziekte daadwerkelijk zal ontstaan is de resultante van de aanwezige etiologische determinanten (risicoprofiel).

Door de processen nauwkeurig te beschrijven in goed gedefinieerde populaties, geeft epidemiologisch onderzoek inzicht in het klinisch verloop van de ziekte.

Determinanten (risicofactoren) die een rol spelen op een moment dat de ziekte nog niet aanwezig is, of op een moment dat de ziekte al wel aanwezig maar nog niet gedetecteerd is, worden *etiologische factoren* genoemd.

Tot de etiologische factoren behoren tevens interventies die zijn gericht op de (ogenshijnlijk) gezonde bevolking.

Etiologische factoren kunnen aan de basis staan van het ontstaan van een ziekte (*initiatoren*) of een reeds in gang gezet ziekteproces bevorderen (*promotoren*).

Factoren die van invloed zijn op het verdere verloop van een ziekte, nadat deze aan het licht is gekomen is door middel van diagnostiek of screening noemt men *prognostische factoren*. Deze factoren zeggen iets over de vermoedelijke afloop van de ziekte: ze beïnvloeden de prognose. De prognostische factoren kunnen voor een deel dezelfde zijn als de etiologische factoren.

Variabelen die louter iets zeggen van de aanwezigheid van de verschillende stadia in het ziekteverloop: de *diagnostische factoren*.

Veel (beschrijvend) epidemiologisch onderzoek is echter gericht op het verkrijgen van kennis die heel specifiek gebonden is aan tijd en plaats. Men spreekt dan van *particularistische kennis*. Deze kennis is particularistisch, dat wil zeggen geldig voor dat moment en die locatie.

Waar het in de epidemiologie allemaal om draait is het voorkomen van ziekte, uitgedrukt in een epidemiologische breuk:

Aantal zieke individuen/ totaal aantal personen in de groep waaruit deze zieke individuen afkomstig zijn.

Bij een objectieve benadering van ziekte is er sprake van ongezondheid op het organisch niveau (disease), voor zover dat van buitenaf kan worden vastgesteld.

Bij een subjectieve benadering van ziekte is er sprake van ongezondheid op het niveau van het individu (illnes).

Bij sociale dimensie benadering is er sprake van ongezondheid op het sociale niveau (sickness).

*Prevalantie* heeft betrekking op het aantal personen dat op een bepaald moment ziek is. Dit aantal bestaande ziektegevallen wordt dan gedeeld door het aantal personen in de populatie waartoe de ziektegevallen behoorden.

Het deel (proportie of percentage) van de populatie waarbij een bepaalde gezondheidstoestand op een bepaald tijdstip aanwezig is wordt de prevalentie van die toestand genoemd.

Daarbij doet het er niet toe of er sprake is van een cohort of van een dynamische populatie: prevalentie impliceert immers een dwarsdoorsnede van de populatie, een momentopname.

Wanneer een prevalentie laag is (minder dan 10%) en er is sprake van een steady state (een constante prevalentie doordat incidentie enerzijds en genezing en sterfte anderzijds met elkaar in evenwicht zijn), kan deze prevalentie worden geschat door het product van de gemiddelde ziekteduur (T) en de incidentiedichtheid (ID).

$$P = ID \times T$$

Voor hogere prevalenties geldt:  $P = \frac{ID \times T}{1 + (ID \times T)}$

De noemer van deze formule zal voor lage prevalenties ongeveer 1 zijn.

Bij hogere ID en/of T zal de corresponderende stijging van de prevalentie steeds minder worden.

Wanneer bij een dwarsdoorsnede van een populatie wordt nagegaan of een bepaalde toestand bij een individu tot op dat moment ooit aanwezig was, spreken we van *lifetime prevalentie*.

*Periode prevalentie* is het deel van de populatie dat de desbetreffende ziekte heeft gehad in een bepaalde periode.

*Incidentie* daarentegen heeft betrekking op het aantal personen dat ziek wordt in een bepaalde periode. Men deelt dan deze nieuwe ziektegevallen door het aantal personen in de populatie aan het begin van (of tijdens) die periode.

Frequentiematen, in het bijzonder incidentiematen, zijn tevens risicomaten: ze geven voor de mensen die tot de desbetreffende groep aan wat de kans is op het krijgen van een ziekte.

De *Cumulatieve incidentie* bestaat uit het deel van de leden van het cohort (op  $t_0$ ) dat gedurende de follow-up deze aandoening krijgt.

De cumulatieve incidentie ligt tussen de 1 en de 0, en kan ook opgevat worden als een individueel risico.

De incidentie in een dynamische populatie wordt niet uitgedrukt ten opzichte van het aantal leden op een bepaald tijdstip (zoals bij een cohort), maar ten opzichte van de totale persoonstijd die is geobserveerd (meestal uitgedrukt in persoonsjaren). De desbetreffende maat wordt *incidentiedichtheid* genoemd (in het Engels: *incidence density* of *hazard rate*). De dimensie is daarmee  $1/\text{jaar}$  ( $\text{jaar}^{-1}$ ) en de waarden zullen liggen tussen 0 en oneindig.

In principe wordt de *incidentiedichtheid* (ID) berekend door het aantal nieuwe gevallen te delen door het aantal geobserveerde persoonsjaren.

Sterfte (mortaliteit) is in feite de incidentie van overlijden.

Een bruto-sterftecijfer beschrijft het aantal overledenen gedurende een bepaalde periode (meestal 1 jaar).

Het bruto-sterftecijfer in de indexpopulatie gedeeld door het indirect gestandaardiseerde sterftecijfer vormt de *standardized mortality ratio* (SMR). Een SMR van 1,25 duidt bijv. op 25% oversterfte in de indexpopulatie ten opzichte van de standaard.

Het *letaliteitspercentage* (*case-fatality rate*) is de proportie van de patiënten met een bepaalde aandoening die ten gevolge van die aandoening overlijdt.

In feite is het letaliteitspercentage de oorzaak specifieke sterfte onder de incidentie gevallen met de desbetreffende aandoening.

Het proportionele sterftecijfer is het deel van alle sterfgevallen in een populatie dat het gevolg is van een bepaalde aandoening.

*Cohort en dynamische populatie.*

Een cohort is een gesloten populatie.

In tegenstelling tot het cohort heeft de dynamische populatie een open karakter. Het lidmaatschap hangt samen met een bepaalde toestand en eindigt zodra een individu niet meer in deze toestand verkeert. Het lidmaatschap van een dynamische populatie is derhalve van variabele duur.

*Centrale tendentiematen en spreidingsmaten.*

Voorbeelden van centrale tendentiematen zijn:

- Het *gemiddelde*, de som van alle meetuitkomsten, gedeeld door het aantal metingen.
- De *mediaan*, de waarde waar 50% van de meetuitkomsten boven en eveneens 50% van de meetuitkomsten onder ligt.
- De *modus*, de meetwaarde die het vaakst voorkomt.

Voorbeelden van spreidingsmaten zijn:

- De *standaardafwijking* of *standaarddeviatie* (SD) van de gemiddelde meetwaarde, dit is bij benadering de absolute waarde van het gemiddelde verschil tussen de individuele meetuitkomsten en de gemiddelde waarde van alle meetuitkomsten.
- De *interpercentielspreiding*, een zone, gemarkeerd door twee meetuitkomsten, waarbinnen een bepaald percentage van de waarnemingen valt
- De *range*, de afstand tussen de hoogste en de laagste meetwaarde in de rangschikking van alle meetwaarden.

## ASSOCIATIES

Determinanten in etiologisch onderzoek zijn grofweg in drie categorieën in te delen: gedrag, constitutie en omgeving.

Constitutionele determinanten zijn soms te beïnvloeden door medische interventies. Factoren als leeftijd en geslacht zijn voorbeelden van niet-beïnvloedbare determinanten.

Een verschil in ziektefrequentie tussen beide groepen duidt op een relatie tussen de determinant en de ziekte, maar kan alleen oorzakelijk geïnterpreteerd worden als beide groepen voor alle andere relevante determinanten gelijk zijn.

*Confounder*: een andere determinant van de ziekte die men bestudeert, die verschillend is verdeeld over de categorieën van de determinant waarin men primair geïnteresseerd is.

*Effectmodifier*: effectmodificatoren zijn overige determinanten die de sterkte van de relatie die men bestudeert, doen toe- of afnemen.

*Regressievergelijking*: In zijn meest simpele vorm beschrijft een regressievergelijking een lineaire relatie van een ziekteparameter met één determinant:  $Z = b_0 + b_1 D_1$ .  
(blz 49tm 52)

Attributief risico.

Om het effect van de determinant op de ziekte uitkomst te kwantificeren ligt het voor de hand beide incidenties van elkaar af te trekken. Dit levert het attributief risico (AR), ofwel het risico verschil. Indien de geexponeerden (D1) en de niet-geexponeerden (D0) verder in alle opzichten vergelijkbaar zijn, is dit het extra risico (cumulatieve incidentie) dat voor rekening komt van de expositie:  $AR = I_1 - I_0$

In deze formulering veronderstelt men dat de associatie oorzakelijk (causaal) van aard is.

Het attributieve risico wordt ook wel risicoverschil of risk difference (RD) genoemd. Sepcifiekere maten zijn het cumulatieve incidence difference (CID) en het incidence density difference (IDD).

Indien de determinant een bepaalde interventie is, gebruikt men als ageleide associatiemaat ook wel de reciproke ( $1/AR$ ). Deze drukt dan het aantal personen uit dat behandeld moet worden om bij één persoon de beoogde uitkomst (bijv. genezing) te realiseren. Deze maat staat bekend als number needed to treat ( $NNT = 1/AR$ ).

Als de determinant niet met de ziekte is geassocieerd dan  $AR=0$ .

Relatief risico.

Het relatief risico geeft de verhouding tussen incidenties (CIR en IDR).

Men berekent dan het relatief risico (RR):  $RR = I1/I0$ .

Zowel de CIR als de IDR zijn dementieeloos. Voor de twee vormen van het relatief risico geldt dan ook dat ze elkaar getalsmatig zullen nadere naarmate:

- De follow-upperiode korter is
- De bestudeerde aandoening zeldzamer is

Dit gaat ook op voor de CID en de IDD.

Het relatief risico kan waarden aannemen tussen de 0 en oneindig, waarbij  $0 < RR < 1$  duidt op een beschermend effect,  $RR < 1$  op een risicoverhogend effect en  $RR = 1$  betekent dat er geen associatie is tussen de determinant en het optreden van de ziekte.

Odds ratio.

In een patiëntcontroleonderzoek, waarbij de patienten en de controles in de regel worden verzameld binnen een dynamische populatie, kan de ziekte-incidentie niet worden vastgesteld.

In een patiëntcontroleonderzoek gaat het om het vergelijken van de prevalentie van de expositie in de patientengroep met die van de controlegroep. Een gangbare maat hiervoor is de odds ratio (OR).

Een odds ratio is het quotiënt van twee odds.

Odds is de kans op een bepaalde gebeurtenis of toestand gedeeld door het complement van die kans.

Berekening OR zie blz 57.

De OR kan doorgaans worden opgevat als een RR.

## ONDERZOEKSOPZET.

De vraagstelling voor een epidemiologisch oorzaak-gevolgonderzoek moet in ieder geval de volgende elementen bevatten:

- De variabelen in het onderzoek, met name de afhankelijke en de onafhankelijke variabelen (wat);
- De onderzoekspopulatie (bij wie);
- De aard en de richting van het veronderstelde verband tussen de onderzoeksvariabelen (hoe);
- De termijn waarop een eventueel effect gerealiseerd dient te zijn (wanneer).

*Interne validiteit:* Een onderzoek heeft een hoge interne validiteit wanneer de resulterende conclusies geldig zijn voor de onderzochte populatie.

*Externe validiteit:* Een onderzoek heeft een hoge externe validiteit wanneer de conclusies ook van toepassing zijn op andere populaties dan de onderzochte.

Externe validiteit verwijst dus naar de generaliseerbaarheid van de onderzoeksresultaten en heeft te maken met de representativiteit van de populatie en de waarnemingen in een bepaald onderzoek.

Een geschikte onderzoekspopulatie in het kader van oorzaak-gevolgonderzoek wordt gekenmerkt door:

- Een duidelijk contrast in de blootstelling aan de determinant die centraal staat in het onderzoek (de vermoedelijke oorzaak);
- Een zo groot mogelijke vergelijkbaarheid ten aanzien van andere determinanten van de ziekte (potentiele confounders);
- Een zo groot mogelijke vergelijkbaarheid ten aanzien van de wijze waarop de verschillende variabelen bij de onderzoekspersonen gemeten worden.

Efficiëntie van het onderzoek.

Behalve om confounding buiten de deur te houden kan *restrictie* van de onderzoekspopulatie ook gebruikt worden als hulpmiddel om de efficiëntie van het onderzoek te verhogen.

Andere manieren om binnen een onderzoekspopulatie te zorgen voor o gering mogelijke verschillen in andere risicofactoren dan de bestudeerde expositiefactor zijn: *randomiseren*, *(pre)stratificeren en matchen*. Zie paragraaf 5.3.2.

Prestratificatie houdt in dat men de leden van de onderzoekspopulatie voorafgaand aan de randomisatieprocedure in lagen (strata) indeelt op grond van hun waarden voor bepaalde kenmerken. Op deze manier neemt de kans aanzienlijk toe dat iedere behandelingsgroep uiteindelijk bijvoorbeeld evenveel vrouwen bevat.

Bij de subpopulaties moet er sprake zijn van vergelijkbaarheid van:

- Populatiekenmerken;
- Expositiecondities;
- Meetprocedures.

Een placebo-interventie vormt tot op zekere hoogte een remedie tegen non-compliance.

De opzet is prospecties: vanaf het moment dat het startsein voor het onderzoek wordt gegeven, is de blik voorwaarts gericht.

## ONDERZOEKSDESIGNS.

In een *ecologisch onderzoek*, ook wel aangeduid met de termen *populatie- of correlatieonderzoek*, vindt de bestudering van de relatie tussen ziekte en andere verschijnselen plaats aan de hand van geaggregeerde gegevens over de relevante variabelen.

Twee subtypen van ecologisch onderzoek zijn *Tijdstrendonderzoek* (analyse van tijdseries) en *geografisch-correlatieonderzoek*.

*Individuele onderzoeken* kunnen op hun beurt worden onderscheiden in *transversale* en *longitudinale onderzoeken*.



Longitudinaal onderzoek kan opgesplitst worden in *experimenteel onderzoek* en *niet-experimenteel onderzoek*.

Een experiment waarbij voorts sprake is van random toewijzing van de interventie maatregel aan de onderzoekspersonen noemt men een *randomized controlled trial (RCT)*. Bij niet-experimenteel onderzoek (*observationeel onderzoek*) is er geen sprake van toewijzing aan de verschillende niveaus van de determinant door de onderzoeker.

Longitudinaal observationeel onderzoek valt weer uiteen in *cohortonderzoek* en *patiëntcontroleonderzoek*.

Cohortonderzoek.

Bij cohortonderzoek vormt de determinant het uitgangspunt. Bijvoorbeeld niemand ziek (*etiologisch onderzoek*), of allemaal hetzelfde klinische ziektebeeld (*prognostisch onderzoek*). Soms hebben alle relevante gebeurtenissen al plaatsgevonden op het moment dat het onderzoek van start gaat. Men spreekt dan van een *historisch cohortonderzoek*.

Een gerandomiseerd experiment is een bijzondere vorm van cohortonderzoek, met de random toewijzing als belangrijkste extra element.

Cohortonderzoek is in de regel niet exploratief.

(zie verder paragraaf 4.3.1)

Patiëntcontroleonderzoek.

Bij patiëntcontroleonderzoek vormt de ziekte het uitgangspunt.

Patiëntcontroleonderzoek volgt de relevante gebeurtenissen dus niet op de voet, maar kijkt juist terug in de tijd (*retrospectief*)

(zie verder paragraaf 4.3.2)

Verschillen tussen cohortonderzoek en patiëntcontroleonderzoek zie tabel 4.2 blz. 103.

VALIDITEIT EN PRECISIE.

Nogmaals epidemiologische functie:  $P(Z) = f(b_0 + b_1D_1 + b_2D_2 + \dots + b_kD_k)$

In woorden: de kans op ziekte ( $p(Z)$ ) wordt beschreven als een functie van een serie determinanten  $D_i$ . De regressiecoëfficiënten  $b_i$  geven hierbij een kwantitatieve schatting van de (onafhankelijke) bijdrage die elk van de determinanten aan de ziektekans levert.

De regressiecoëfficiënten uit de epidemiologische functie zijn de parameters die de sterkte van het verband tussen de betreffende determinant en de kans op de ziekte weergeven.

De bruikbaarheid van een meetinstrument wordt bepaald door:

- De Validiteit
- De reproduceerbaarheid
- De hanteerbaarheid

Zie blz. 125.

*Operationalisatie*: het meetbaar maken van kenmerken.

Voor een goede operationalisatie van een in een epidemiologische functie opgenomen kenmerk zal men de volgende stappen moeten ondernemen:

- 1) Vaststellen wat men wil weten: een *conceptuele definitie* van de variabele.
- 2) Vaststellen van de mogelijke waarden die men bij dit concept wil onderscheiden. Men noemt dit ook wel de *conceptuele schaal*.
- 3) Omzetten van de conceptuele definitie in een empirische (operationele) definitie van de variabele.
- 4) Vaststellen van de empirische schaal waarop men de onderzoekspersonen feitelijk gaat meten, ofwel de mogelijke waarden die men bij het meten in de praktijk wil onderscheiden.

Zie blz. 125-126.

Onderzoek is valide als systematische fouten ontbreken, onderzoek is precies als er weinig toevallige fouten zijn.

Systematische fouten vinden hun oorsprong in verkeerde beslissingen tijdens de opzet, de uitvoering of de analyse van een onderzoek. Toevallige fouten leiden tot niet-precieze effectschattingen, systematische fouten leiden even wel tot onjuiste effectschattingen. Zij tasten de validiteit van de onderzoeksuitkomsten aan.

De eis van precisie impliceert dat een herhaling van het onderzoek (met dezelfde opzet en methode) steeds (nagenoeg) hetzelfde resultaat levert.

Steekproefomvang, spreiding en meetfout bepalen de mate van precisie.

BIAS.

Validiteit is het ontbreken van bias.

Er zijn drie bedreigingen van de validiteit:

- 1) Selectiebias
- 2) Informatiebias
- 3) Confounding

Bias kan ervoor zorgen dat het effect wordt overschat, onderschat of zelfs een andere richting krijgt.

Indien in een patientcontroleonderzoek bijvoorbeeld een vrij duidelijk verhoogd risico wordt gevonden (OR= 3,0) en beredeneerd kan worden dat, zo er al bias in het spel is, deze in de richting van de nul moet gaan, dan zal de werkelijke OR groter zijn dan 3,0 en is er bijna zeker sprake van een verband. Indien een nauwelijks verhoogd risico wordt gevonden (OR=1,3) en beredeneerd kan worden dat eventuele bias van de nul weg gaat, dan is de werkelijke OR kleiner dan 1,3 en stelt dat wat overblijft weinig of niks meer voor.

Selectiebias.

Selectiebias is vertekening van de effectschatting als gevolg van fouten bij de selectie of de follow-up van de onderzoekspersonen.

Informatiebias.

Informatiebias is vertekening van de effectschatting als gevolg van fouten bij de meting van de variabelen in het onderzoek (determinant, ziekte-uitkomst, confounders, effectmodificatoren). En dus bij de classificatie van de onderzoekspopulatie in subcategorieën op grond van de gemeten variabelen.

Slechts differentiële selectie mondt uit in selectiebias.

Confounding.

Confounding resulteert in een onjuiste effectschatting, doordat de groepen die in het onderzoek worden vergeleken op verschillende wijze zijn blootgesteld aan andere risicofactoren voor de ziekte.

Doorgaans reserveert men het begrip confounding voor bekende en meetbare risicofactoren van de betreffende ziekte die ongelijk verdeeld zijn over de te vergelijken groepen.

Confounding berust op de verstremgeling van het effect van de te bestuderen ziektedeterminant met het effect van een of meer andere, externe determinanten.

Om als confounder van de relatie tussen een determinant (etiologische factor, prognostische factor) en een ziektevariabele te kunnen optreden, dient een factor te voldoen aan de volgende voorwaarden:

- 1) De desbetreffende factor is een onafhankelijke determinant van de bestudeerde ziekte;
- 2) De desbetreffende factor is geassocieerd met de determinant waarvan primair het effect wordt bestudeerd;
- 3) De desbetreffende factor vormt geen verbindingsschakel in de causale keten tussen determinant en ziekte.

Confounding kan resulteren in positieve bias (vanaf de nul) en negatieve bias (naar de nul), of zelfs in een omkering van de richting van het effect.

Differentiële en non-differentiële misclassificatie.

Differentiële misclassificatie van de primaire determinant en/of de ziekte-uitkomst leidt soms tot een onderschatting, soms tot een overschatting van het werkelijke effect van de determinant op de ziekte-uitkomst. In het meest extreme geval kan de richting van het effect zelfs omdraaien.

Non-differentiële misclassificatie zal niet altijd tot vertekening leiden, maar indien dat het geval is, dan altijd naar de nul toe. Een in werkelijkheid aanwezig effect wordt dus onderschat.

*Sampling error* is een belangrijk aspect van random error (toevallige fouten).

Sampling error drukt men in de statistiek uit met behulp van de standaardfout (standard error SE)

De standaardfout wordt bepaald door drie elementen:

- De meetfout van de afzonderlijke waarnemingen bij de individuen in de steekproef;
- De spreiding (heterogeniteit, variatie) van het kenmerk in de populatie waaruit de steekproef getrokken is;
- De omvang van de steekproef.

Betrouwbaarheidsinterval.

Een betrouwbaarheidsinterval geeft alleen informatie over de precisie (toevalsfout) en niet over de validiteit (systematische fout) van een epidemiologische parameter.

Het 95%-betrouwbaarheidsinterval van een epidemiologische parameter P waarvoor (ongeveer) een standaard-normale verdeling geldt, is dan:

$$95\text{-BI} = P \pm 1.96 SE_p$$

Voor parameters zoals RR en OR geldt geen standaard-normale verdeling, maar een ln-normale verdeling. Het 95%-betrouwbaarheidsinterval voor  $\ln(P)$  is dan:

$$95\text{-BI} = P \pm 1.96 SE_{\ln(P)}$$

Een betrouwbaarheidsinterval omvat de situatie van geen verhoogd risico (RR=1), van een zwak verhoogd risico (RR tussen 1 en 2) en van een matig verhoogd risico (RR tussen 2 en 3).

#### P-WAARDE.

Een p-waarde van bijvoorbeeld 0,03 drukt uit dat, indien de nulhypothese die stelt dat er geen effect is van een determinant op een ziekte juist is, er slechts 3% kans is ( $p$ = probability) om bij toeval toch de gevonden waarde of een nog extremere waarde van de effectmaat te treffen.

Hoe kleiner de p-waarde, of overschrijdingskans, hoe sterker de significantie.

Een p-waarde van 0,03 zegt dan dat de kans dat dit resultaat louter op grond van het toeval verkregen is (sampling error) zo klein is, dat dat idee (hypothese) verworpen moet worden. Evenzo is het gebruikelijk om een p-waarde groter dan 0,05 te interpreteren als dat het verkregen onderzoeksresultaat wel door het toeval veroorzaakt zou kunnen zijn.

*Statistische toetsen* beantwoordt de vraag of er een effect is, niet hoe groot die is of zou kunnen zijn.

De precisie wordt beter door een grotere steekproef , een kleinere meetfout en efficiëntere designs.



Afkortingen.

AR= attributief risico

CI= cumulatieve incidentie

CIR= cumulatieve incidentie ratio

ID= incidentiedichtheid

IDR= incidentie dichtheidsratio

NNT= number needed to treat

OR= odds ratio

RR= relatief risico

SD= standaarddeviatie

SE= standard error

SMR= standardized mortality ratio

