

Cochrane-speak: een introductie

J. VAN OS

SAMENVATTING Met behulp van een aantal praktische voorbeelden wordt een toelichting gegeven op de terminologie die wordt gebezigd in de systematische reviews van de Cochrane Library. De termen odds ratio, Mantel-Haenszel odds ratio, confidence interval, funnel plot, numbers needed to treat, numbers needed to harm, likelihood ratio for a positive test result, pre-test probability en post-test probability passeren de revue.

[TIJDSCHRIFT VOOR PSYCHIATRIE 42 (2000) 6, 423-428]

TREFWOORDEN Cochrane Library, epidemiologie, evidence-based psychiatrie

Een belangrijk deel van de informatie die wordt gebruikt bij het maken van klinische beslissingen, is gebaseerd op wetenschappelijk onderzoek, met name op gerandomiseerde gecontroleerde trials (GGT's). GGT's zijn veelal gericht op onderzoek naar het klinisch effect en de bijwerkingen van een therapeutische interventie. De klinische effecten van zowel psychologische als farmacologische therapieën in de psychiatrie zijn, in tegenstelling tot wat soms wordt verondersteld, relatief klein. Dit is belangrijk voor de interpretatie van GGT's, omdat in het geval van kleine effecten de factor kans een grotere rol gaat spelen. Stel dat middel x een echt, doch relatief klein therapeutisch effect heeft op een belangrijke psychiatrische uitkomst in vergelijking met placebo. Het is in een dergelijk geval niet ongebruikelijk dat zelfs een goed opgezette en uitgevoerde GGT een kans heeft van 1 op 5 à 10 op een resultaat waarbij de superioriteit van middel x niet kan worden aangetoond. Het is dus zaak zich niet te baseren op de resultaten van een enkel onderzoek, maar op meerdere vergelijkbare GGT's. Als van de tien GGT's die middel x vergeleken met placebo, er zeven lieten zien dat middel x superieur was aan placebo, twee geen verschil konden aantonen, en één placebo effectiever

vond dan middel x, is dat een aanwijzing voor een echt effect van middel x in vergelijking met placebo. Men kan de resultaten van de tien verschillende onderzoeken dan trachten te combineren in een enkel 'meta'resultaat. De resultaten van een dergelijke meta-analyse of systematic review vormen, mits goed uitgevoerd, vaak het best beschikbare bewijs voor de effectiviteit en veiligheid van een bepaalde therapie.

Het combineren van GGT's in één enkel resultaat is bijzonder tijdrovend en vereist speciale deskundigheid. Een ondeskundig uitgevoerde meta-analyse kan veel schade aanrichten. De Cochrane Collaboration (zie <http://hiru.mcmaster.ca/cochrane/default.htm>) bestaat uit een netwerk van onpartijdige deskundigen, die volgens zeer stringente criteria individuele onderzoeken zoals GGT's (1) opsporen in de medische wereldliteratuur en (2) voorzover mogelijk combineren in een enkel globaal resultaat, en die (3) hun werk laten beoordelen door andere deskundigen alvorens tot publicatie over te gaan. Het groeiende bestand van zogenaamde systematic reviews in de Cochrane Library (zie <http://hiru.mcmaster.ca/cochrane/cochrane/cdsr.htm>) is van onschatbare waarde voor de klinisch practicus, die niet de tijd heeft zelf de individuele GGT's op te zoeken en te

combineren, en de deskundigheid mist om ingewikkelde statistische manipulaties van moderne meta-analysen op hun merites te beoordelen. De *Cochrane Library* kan inmiddels in de meeste medische bibliotheken in Nederland worden geraadpleegd, en heeft onmiskenbaar een groeiende invloed. Het enige doel van de *Cochrane Collaboration* is om de onderzoeksresultaten die op onbereikbare plaatsen zijn begraven in de wereldliteratuur toegankelijk te maken voor klinici, zodat ze het wetenschappelijk bewijs gemakkelijk kunnen raadplegen indien zij dat wensen.

Anekdotisch bewijs suggereert dat de meeste Nederlandse psychiaters de *Cochrane Library* nog niet regelmatig raadplegen. Een reden die vaak naar voren wordt gebracht is dat *systematic reviews* hun cruciale bevindingen weergeven in een nogal specifieke en in eerste instantie weinig toegankelijke terminologie, die is terug te vinden in de literatuur van de evidence-based geneeskunde. Het doel van dit artikel is een toelichting te geven op de terminologie zoals die wordt gebezigd in de *systematic reviews* van de *Cochrane Library*. Omdat de *Cochrane Library* gepubliceerd wordt in de Engelse taal, wordt hier veelal gebruikgemaakt van de originele Engelse termen. Als referentiewerk kan worden gebruikt *Evidence-based Medicine* van Sackett e.a. (Churchill Livingstone 1997).

ODDS RATIO (OR)

In een (fictief) *systematic review* wordt de werking van het atypisch antipsychoticum ATYP vergeleken met dat van het typisch antipsychoticum TYP. Alle relevante GGT's die aan de inclusiecriteria voldoen, worden gecombineerd. In totaal zijn in 15 trials 700 psychotische patiënten met ATYP behandeld en 700 met TYP. Een van de uitkomsten die worden geanalyseerd, is: *geen globale verbetering na zes weken behandeling*. De odds ratio in de vergelijking ATYP versus TYP is 0.49. In de ATYP-groep is bij 250 patiënten *geen globale verbetering* opgetreden na zes weken, bij 450 wel.

In de TYP-groep is bij 375 *geen globale verbetering* opgetreden na zes weken, bij 325 wel. De odds op *geen globale verbetering* in de ATYP-groep is $250/450 = 0.56$. De odds op *geen globale verbetering* in de TYP-groep is $375/325 = 1.15$. De odds ratio is de ratio van de odds in de ATYP- en de TYP-groepen, en is een maat van de effectgrootte van ATYP. In dit geval is de odds ratio dus $0.56/1.15 = 0.49$. Een odds ratio van 1 geeft aan dat er geen verschil is tussen de experimentele en de controlegroep. Een odds ratio kleiner dan 1 duidt op een beschermend effect (in dit geval beschermt ATYP tegen het uitblijven van globale verbetering), en een odds ratio groter dan 1 duidt op een inducerend effect.

CONFIDENCE INTERVAL

Bij bovenstaande odds ratio hoort een betrouwbaarheidsinterval (*confidence interval*). De OR was 0.49, het betrouwbaarheidsinterval 0.2-0.8. Een dergelijk betrouwbaarheidsinterval is zinvol, omdat het aangeeft hoe zeker we van onze zaak zijn als we zeggen dat de effectgrootte 0.49 bedraagt. Volgens het betrouwbaarheidsinterval zou, als we het experiment 100 keer zouden herhalen met 100 verschillende steekproeven uit de totale populatie van patiënten met psychose, 95 van de 100 keer de OR tussen de 0.2 en de 0.8 liggen. Dus zelfs als we het experiment 100 keer zouden herhalen, zou het effect in 95% van de gevallen kleiner zijn dan 1, dat wil zeggen in het voordeel uitvallen van ATYP, hoewel de grootte van dat voordeel een factor 4 (0.8/0.2) kan verschillen. In het algemeen geldt dat als het getal 1.0 niet in het betrouwbaarheidsinterval zit, het resultaat statistisch significant is. Stel dat het betrouwbaarheidsinterval niet 0.2-0.8 was geweest, maar 0.4-1.1; een OR groter dan 1 geeft in dit geval aan dat patiënten met ATYP een grotere kans hebben op *geen globale verbetering* na zes weken. In het geval van een betrouwbaarheidsinterval van 0.4-1.1 kan dus niet met 95% zekerheid uitgesloten worden dat ATYP slechter werkt dan TYP.

NUMBERS NEEDED TO TREAT (NNT)

In de TYP-groep was bij 375 van de 700 patiënten, ofwel 54%, geen globale verbetering opgetreden na zes weken. In de ATYP-groep was dit 36% (250/700). Het verschil bedraagt dus 18%. Met andere woorden: als er in beide groepen 100 mensen behandeld waren, zou bij 18 mensen het uitblijven van globale verbetering zijn voorkomen dankzij het voordeel van ATYP boven TYP, en als er $100/18 = 5,6$ mensen waren behandeld met ATYP, zou bij één het uitblijven van een globale verbetering zijn voorkomen dankzij het voordeel van ATYP boven TYP. Het getal van 5,6 is de *number needed to treat*, en wordt altijd afgerond naar boven, in dit geval dus naar 6. Zes mensen moeten met ATYP worden behandeld opdat er bij één het uitblijven van globale verbetering wordt voorkomen. Anders gesteld: van de zes mensen die werden behandeld met ATYP, hadden er vijf net zo goed met TYP behandeld kunnen worden, voorzover het althans de uitkomst geen globale verbetering betreft. Het is duidelijk dat de klinische relevantie van een behandeling beter tot zijn recht komt in *numbers needed to treat* dan in een maat van relatief risico, zoals de odds ratio.

NUMBERS NEEDED TO HARM (NNH)

In de *systematic review* van ATYP versus TYP kwam het optreden van EPS ook aan de orde. EPS trad op over een periode van zes weken bij 25% van de TYP-groep, en bij 5% van de ATYP-groep. Het verschil was dus 20%, en de *number needed to harm* was $100/20 = 5$. Met andere woorden, bij elke vijf mensen die worden behandeld met TYP, is er één die een episode van EPS zal ervaren door het ongunstigere EPS-bijwerkingsprofiel van TYP ten opzichte van ATYP.

MANTEL-HAENSZEL ODDS RATIO

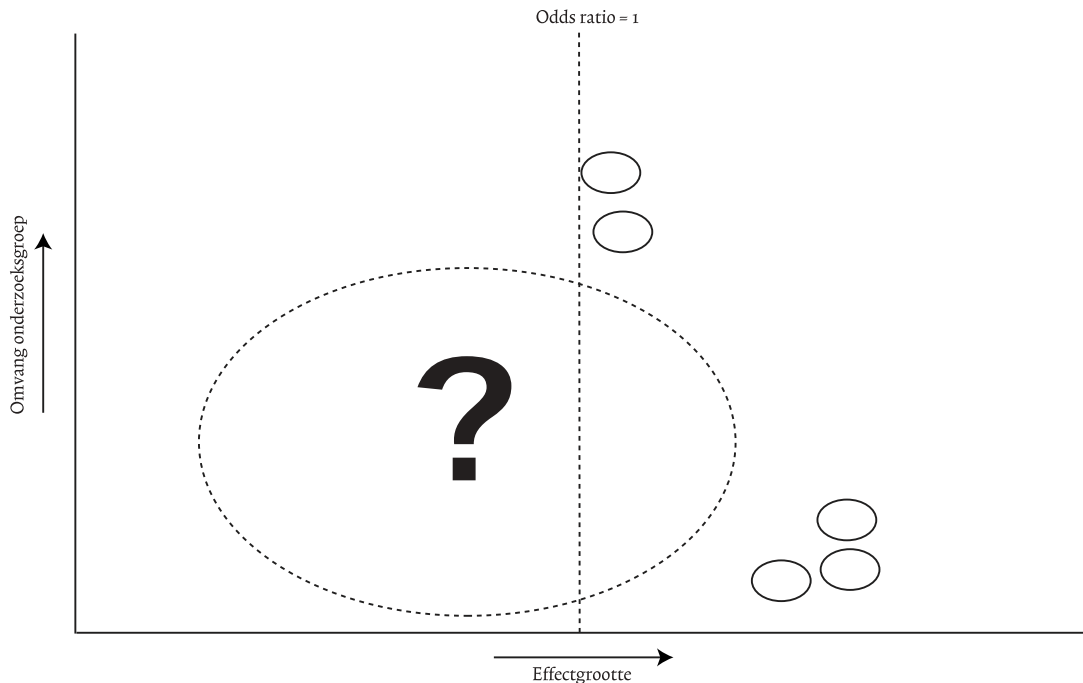
In tabel 1 zijn de resultaten samengevat van vijf (fictieve) onderzoeken die zijn opgenomen in een *systematic review* over het verband tussen een prenatale risicofactor PRENAT en psychose (PSYCH) later in het leven. In onderzoek 1 ontwikkelden 25 mensen die waren blootgesteld aan PRENAT psychose, en 11 niet (odds: $25/11 = 2.3$). In de groep van niet-blootgestelden waren dit respectievelijk 19 en 22 mensen (odds: $19/22 = 0.9$). De odds ratio in onderzoek 1 was dus $2.3/0.9 = 2.6$. De odds ratio's in de vijf onderzoeken waren: 2.6,

TABEL 1 Combinatie van vijf verschillende onderzoeken naar het verband tussen blootstelling aan de prenatale risicofactor PRENAT en psychose (PSY) op latere leeftijd

Onderzoek	PRENAT+	PRENAT-	OR (95% BI)	MH-wegingsfactor
1	PSY+: 25 PSY-: 11	PSY+: 19 PSY-: 22	2.6 (1.04 - 6.7)	2.7
2	PSY+: 26 PSY-: 12	PSY+: 25 PSY-: 11	2.4 (0.97 - 5.7)	3.1
3	PSY+: 350 PSY-: 324	PSY+: 381 PSY-: 393	1.1 (0.9 - 1.4)	85.3
4	PSY+: 23 PSY-: 10	PSY+: 22 PSY-: 25	2.6 (1.03 - 6.6)	2.8
5	PSY+: 315 PSY-: 293	PSY+: 201 PSY-: 217	1.2 (0.9 - 1.5)	57.4
Mantel-Haenszel OR onderzoek 1-5 gecombineerd (95% BI)			1.2 (1.04 - 1.4)	

FIGUUR 1 Funnel plot

Twee onderzoeken met een grote onderzoeksgroep hebben een kleine effectgrootte (OR in de buurt van 1); drie kleinere onderzoeken hebben een grote effectgrootte. De grote onderzoeken suggereren dat het ware effect in de buurt van de OR = 1 ligt, maar door het ontbreken van kleine onderzoeken met een geringe of tegenovergestelde effectgrootte wordt het totale resultaat mogelijk vertekend.



2.4, 1.1, 2.6 en 1.2. De gemiddelde odds ratio is dus de som van deze odds ratio's, gedeeld door vijf, ofwel 2.0. In de tabel echter is te zien dat de gecombineerde odds ratio volgens de methode van Mantel-Haenszel slechts 1.2 is. Dit komt doordat bij de berekening van de gecombineerde odds ratio veel meer gewicht wordt gegeven (zie kolom 'wegingsfactor' in tabel 1) aan de 'grote' onderzoeken 3 en 5 dan aan de kleine onderzoeken 1, 2 en 4. Omdat de grote onderzoeken in vergelijking met de kleinere onderzoeken slechts een klein effect vonden (OR's 1.1 en 1.2), is het resultaat van de gecombineerde analyse veel dicht bij de effectgrootte van de grote onderzoeken. Dat de grote onderzoeken zwaarder wegen bij de berekening van het gecombineerde resultaat is noodzakelijk, omdat het effect van de

grote onderzoeken veel 'preciezer' is. Volgens het betrouwbaarheidsinterval van bijvoorbeeld onderzoek 1 ligt het ware effect van de risicofactor PRENAT ergens tussen de 1.04 en de 6.7, terwijl onderzoek 3 een veel preciezere schatting geeft, namelijk tussen de 0.9 en de 1.4.

FUNNEL PLOT

In figuur 1 is de grootte van de onderzoeksgroep van de vijf onderzoeken naar het effect van risicofactor PRENAT op psychose uitgezet tegen de grootte van de odds ratio (de effectgrootte). Een dergelijke grafiek wordt een *funnel plot* genoemd. In het *funnel plot* is duidelijk te zien dat de grote onderzoeken zich dicht bij de lijn van $OR = 1$ bevinden, terwijl de kleine onderzoeken

relatief grote effectgroottes hebben. Het plot laat verder zien dat er een 'gat' is waar zich kleine onderzoeken met een odds ratio in de buurt van de 1 hadden moeten bevinden als (1) de risicofactor PRENAT het risico op psychose niet of nauwelijks verhoogt, en (2) een 'negatief' onderzoek evenveel kans had om gepubliceerd te worden als een 'positief' onderzoek. Aangezien de grote onderzoeken suggereren dat de ware effectgrootte marginaal groter is dan 1, lijkt het erop dat er sprake is van *publication bias*: kleine onderzoeken met een interessant effect werden geaccepteerd door de redacteurs van wetenschappelijke tijdschriften, maar kleine onderzoeken zonder een duidelijk effect werden door hen niet geaccepteerd, of door de auteurs niet voor publicatie aangeboden.

LIKELIHOOD RATIO FOR A POSITIVE TEST RESULT

Ook op het gebied van de diagnostiek wordt in de *Cochrane Library* en verwante publicaties gebruikgemaakt van speciale terminologie. In een (fictief) onderzoek wordt aan alle 500 22-jarige vrouwelijke studentes van de lokale universiteit door middel van een *E-mail questionnaire* gevraagd of ze in de laatste maand last hebben gehad van een periode van persisterende sombere stemming die ten minste twee weken duurde. Vervolgens worden alle 500 geïnterviewd door een team van psychiaters om vast te stellen wie voldoet aan ICD-10-criteria voor depressie. Het doel van de exercitie is om uit te vinden of de *E-mail questionnaire* een goede diagnostische test is voor een episode van klinische depressie. Van

de 500 geïnterviewden voldeden 30 (6%) aan de ICD-10-criteria van depressie (*pre-test probability* 6%). Van deze 30 hadden 25 in antwoord op de *E-mail questionnaire* geschreven dat ze een sombere stemming hadden. De e-mailtest had dus $25/30 = 83\%$ van de gevallen van depressie correct geïdentificeerd (sensitiviteit, of aantal ware positieven 83%). Van de 470 zonder depressie hadden 35 (7,5%) aangegeven een sombere stemming te hebben in de *E-mail questionnaire* (7,5% fout-positieven). De *Likelihood ratio for a positive test result* (LR+) van de test is het percentage ware positieven gedeeld door het percentage fout-positieven. In deze studie is de *likelihood ratio* dus $83\%/7,5\% = 11.1$. Een test met een LR+ van 10 of meer wordt geacht goede diagnostische eigenschappen te bezitten. De e-mailtest lijkt dus geschikt voor verder gebruik. Van de mensen met een positief testresultaat ($25+35 = 60$) hadden 25 een depressie. De *post-test probability* was $25/60 = 41,7\%$, een aanzienlijke verbetering ten opzichte van de *pre-test probability* van 6%.

AUTEUR

J. VAN OS, psychiater-epidemioloog, vakgroep Psychiatrie en Neuropsychologie, European Graduate School of Neuroscience.

Correspondentieadres: Universiteit Maastricht, Postbus 616, 6200 MD Maastricht. Telefoon: (043) 3299773, fax: (043) 3299708, e-mail: j.vanos@sp.unimaas.nl.

Geen strijdige belangen meegedeeld.

Het artikel werd voor publicatie geaccepteerd op 21-9-1999.

SUMMARY Cochrane-speak: An introduction – J. van Os –

A number of practical examples is used to illustrate the terminology that can be encountered in the systematic reviews of the Cochrane Library. The terms odds ratio, Mantel-Haenszel odds ratio, confidence interval, funnel plot, numbers needed to treat, numbers needed to harm, likelihood ratio for a positive test result, pre-test probability and post-test probability are reviewed.

[TIJDSCHRIFT VOOR PSYCHIATRIE 42 (2000) 6, 423-428]

KEYWORDS Cochrane Library, epidemiology, evidence-based psychiatry