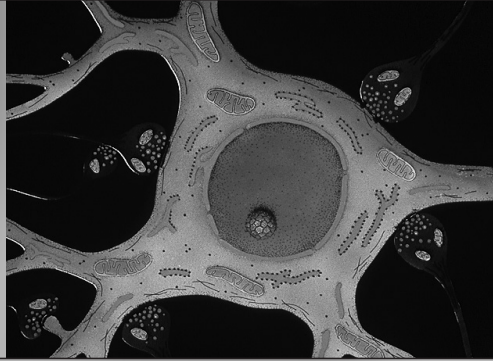
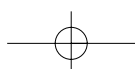
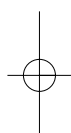
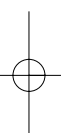
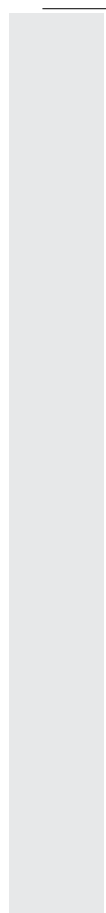
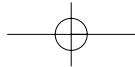


Supplement bij Neuron 2003; Vol 8 (Nr 9)



Extrapolatie van de resultaten van klinische studies naar de klinische praktijk

V Dubois
A Masson
B Gillain
E Stillemans
B Delatte
MA Domken
S Gallemaers
I Meire



Inleiding

Aangezien het aantal nieuwe geneesmiddelen gestaag toeneemt, zijn vergelijkende studies met die geneesmiddelen nodig om hun relatieve sterke en zwakke punten tegen elkaar af te wegen, om zo therapeutische beslissingen te leiden.

Goed uitgevoerde en correct geanalyseerde klinische studies artsen in een klinische praktijk waardevolle informatie bijbrengen.

Dit artikel toont aan waarom het zo belangrijk is dat men de methode begrijpt die in klinische studies wordt toegepast. Verschillen in methode kunnen vergelijkingen immers aanvechtbaar maken. Als men vertrouwd is met de verschillende studieprotocollen, kan men de resultaten van de klinische studies ook beter begrijpen en interpreteren. We zullen enkele vragen aansnijden die de resultaten van de klinische studies beïnvloeden, vragen die uiterst belangrijk en informatief zijn voor het extrapoleren van de resultaten. Bovendien spreekt het voor zich dat als studies met een verschillend design elkaar bevestigende resultaten opleveren, de conclusies ervan beter kunnen worden veralgemeend.

Atypische antipsychotica vormen momenteel de eerstelijnsbehandeling voor patiënten met schizofrenie, maar we beschikken over relatief weinig gegevens om een specifiek atypisch antipsychoticum te kiezen. Uitgaande van de bestaande studies met (oude en nieuwe) antipsychotica, zullen we uitleggen hoe de resultaten moeten worden geïnterpreteerd en hoe ze naar de klinische praktijk kunnen worden doorgetrokken.

De interpretatie van de resultaten van klinische studies hangt af van ons inzicht in de methode die werd gevolgd.

Wat is een klinische studie?

Nog niet zo lang geleden verstond men onder klinische studie vooral een gerandomiseerde klinische studie (RCT, *randomized controlled trial*). RCT-studies werden als enige referentie beschouwd. Nu wordt elke gestructureerde studie die het effect van één of meer behandelingen bij patiënten gaat meten, klinische studie genoemd. De kenmerken van de studies zijn erg wisselend, wat meteen mogelijkheden biedt tot extrapolatie op verschillende niveaus.

In dit hoofdstuk beschrijven we enkele aspecten die van belang zijn bij het uitvoeren en interpreteren van een studie. Om her overzichtelijk te houden, hebben we de structuur van de meeste gepubliceerde studies gevolgd, namelijk de formulering van de hypothese, de algemene methode, de kwantificering van de metingen, de statistische analyse en de interpretatie van de resultaten.

Formulering van de hypothese

Een groot deel van onze kennis in de klinische psychiatrie vloeit voort uit strikte observaties die hypothesen voortbrengen, hypothesen die door klinische studies worden getest.

De onderzoekers moeten trachten hun hypothese precies te formuleren alvorens ze beginnen, tenzij het een volledig verkennende studie betreft. De gepubliceerde resultaten van de studies stemmen dan ook niet steeds overeen met de vraag die in de hypothese werd geformuleerd, maar kunnen het verkennende gedeelte van de studie vormen.

De hypothese of de doelstelling moet bepalen waarom de studie wordt uitgevoerd en bepaalt mede het type metingen en de methode voor de statistische analyse. Over het algemeen wordt de hypothese geformuleerd in de inleiding van de publicatie van de studie. Maar vaak wordt de hypothese niet in de studie vermeld. Als er a priori geen hypothese wordt geformuleerd,

Een **hypothese** is een bewering die kan worden getest en die slaat op een veronderstelde relatie tussen twee of meer variabelen – een bewering die zal worden aanvaard of verworpen naargelang van de resultaten van de studie. Als een hypothese kan worden verworpen, zal een andere hypothese plausibeler worden geacht. Als een gegeven hypothese wordt bevestigd, worden bepaalde interpretaties dus waarschijnlijker en als de hypothese niet wordt bevestigd, worden andere interpretaties waarschijnlijker (1).

kunnen de lezers niet nagaan welke gegevens selectief min of meer in de verf werden gezet (2). Schooler (3) daarentegen vindt dat het niet nodig is een duidelijke hypothese te formuleren. Het verloop van de studie hangt af van de richting waarin de gegevens de onderzoekers sturen. Volgens Schooler zouden de studies niet steeds door een hypothese moeten worden gestuurd. In plaats van zich te concentreren op één enkele specifieke doelstelling die van tevoren werd vastgelegd, zou het beter zijn een zo goed mogelijk beeld te schetsen van de gevonden effecten van de onderzochte stoffen.

Methode

Algemeenheden

De verschillende protocollen

De doelstellingen (hypothesen) van de studie bepalen het studieprotocol. Het protocol hangt nauw samen met de doelstellingen van de studie en omgekeerd: de doelstellingen van een studie bepalen het studieprotocol.

Er bestaan verschillende soorten studies, maar vooraleer de meest voorkomende design te bespreken, zullen we eerst enkele concepten definiëren die vaak worden gebruikt.

Er is vooreerst het '**algemene concept**' van de studie. We kunnen twee soorten studies onderscheiden: experimentele studies waarin men de blootstellingsfactor gaat manipuleren, en observationale studies die de natuur de vrije loop laten, waar dus niet wordt ingegrepen in het gedrag van de arts, maar waarbij men de acties van de arts en de klinische evolutie van de patiënt noteert.

De '**tijd**' is een tweede variabele. Men maakt een onderscheid tussen statische of transversale studies en dynamische of longitudinale studies. Statische studies worden zelden uitgevoerd, maar ze bestaan. In die studies worden de verschillen op een bepaald ogenblik vergeleken (bijvoorbeeld prevalentie van de gemiddelde doseringen). In dynamische studies, die veel vaker worden uitgevoerd, buigt de onderzoeker zich over de evolutie van de behandeling, en tijdens de studie kunnen de behandelingsgroepen meermaals worden vergeleken. Die studies kunnen prospectief of retrospectief zijn. In prospectieve studies moeten de gegevens nog worden verzameld op het ogenblik dat de studie wordt opgezet. Doorgaans wordt een hypothese geformuleerd en worden de gegevens verzameld in functie van de hypothese. In retrospectieve studies daarentegen zijn de gegevens reeds verzameld als de doelstellingen van de studie worden gedefinieerd. Een retrospectieve studie biedt het voordeel dat gegevens uit de realiteit worden verzameld, en niet in het kunstmatige kader van een experimentele studie.

De studie kan al dan niet een hoge **graad van controle** inhouden. In een zeer gestructureerde studie, bijvoorbeeld de RCT, wordt de onderzochte populatie afgebakend door specifieke inclusie- en exclusiecriteria. De patiënten die aan de criteria voldoen, vormen de doelgroep en krijgen na randomisatie een of andere behandeling. De patiënten worden willekeurig in verschillende behandelingsgroepen ingedeeld om identieke groepen te verkrijgen, behalve wat het toegediende geneesmiddel betreft.

De toegediende behandeling wordt dan ook verantwoordelijk geacht voor een verschil tussen de groepen. Op die manier worden systematische fouten (bias) zo goed mogelijk gecontroleerd. Maar als het studieschema zeer gestructureerd is, zal de mogelijkheid tot veralgemening beperkt zijn.

Als randomisatie onmogelijk is doordat geen grote groepen kunnen worden gevormd (grote cognitieve testbatterijen, dure PET-metingen), kan een casus-controlestudie worden uitgevoerd om dat probleem op te lossen. Voor elke onderzochte casus neemt men een gezonde controlepersoon die zoveel mogelijk kenmerken gemeen heeft met de casus, bijvoorbeeld leeftijd, geslacht, opvoeding, gewoontes, socio-economische groep...

De **organisatie van de observaties** is een andere variabele. Er bestaan vier types: enkelblinde studies, dubbelblinde studies, open studies en gekruiste studies. Dubbelblinde evaluaties dragen bij tot een geldig experimenteel protocol. Men gaat ervan uit dat blinde evaluaties niet vertekend worden door kennis van de

onderzochte behandeling door de proefleider. De studies van Tran et al (4) en van Conley & Mahmoud (5) zijn twee voorbeelden van dubbelblinde studies. Die studies worden verder nog onderzocht en vergeleken. In open studies weet zowel de patiënt als de onderzoeker welk geneesmiddel de patiënt krijgt en is er dus meer kans op bias. De studie van Ho et al (6) en de QUEST-studie (7) zijn voorbeelden van open studies.

Bij een dergelijke studie moet rekening worden gehouden met **ethische aspecten**. Bijvoorbeeld, in welke omstandigheden is het ethisch verantwoord dat schizofrene patiënten, voor wie er een efficiënte behandeling bestaat, deelnemen aan een klinische studie? Psychofarmaca worden beschouwd als de hoeksteen van de behandeling. Is het dan ethisch verantwoord de patiënten een placebo te geven? Mogen patiënten die een bevredigende klinische respons vertonen op de behandeling, in de studie worden opgenomen, ook als dat hun klinische toestand zou kunnen verslechteren? Mag men zwaar zieke patiënten uitsluiten en hen zo blootstellen aan het risico van een klinische achteruitgang?

RCT:

- zeer gestructureerd
- de randomisatie vermindert de variabiliteit aan de basislijn
- een dubbelblinde studie controleert systematische bias bij de patiënten, bij de evaluaties door de onderzoeker en in de wijze van toediening van de behandeling
- doorgaans prospectief
- systematische fouten (bias) worden in de mate van het mogelijke uitgesloten
- de populatie is beperkt tot de patiënten die voldoen aan de inclusiecriteria en die erin toestemmen om deel te nemen aan een dubbelblinde klinische studie
- het zou kunnen dat het studieprotocol de behoeften van de patiënten in de reële praktijk niet weerspiegelt

Naturalistische studie:

- reële situaties
- representatieve populatie
- vaak retrospectief
- kan de artsen leiden in de klinische praktijk
- vatbaar voor bias vermits de patiënt en de arts weten welke behandeling wordt toegediend (wat later kan worden uitgesloten door analyse van de selectieprocedure)
- er wordt enkel een rapport gemaakt van de responders; de non-responders verlaten de studie doorgaans voortijdig
- vele omgevingsfactoren kunnen niet worden geïdentificeerd noch gecontroleerd

De rechten, de veiligheid en het welzijn van de proefpersonen worden uiteraard beschermd. De ethische principes die moeten worden gevolgd bij medisch onderzoek op mensen, staan bijvoorbeeld in de "Verklaring van Helsinki" van de *World Medical Association*. Er zijn echter verschillen tussen de Verenigde Staten en Europa. In de Verenigde Staten moet een placebo worden gebruikt, terwijl in Europa placebogecontroleerde studies verboden zijn.

De onderzoekers moeten scrupuleus nagaan of de patiënten die erin toestemmen om deel te nemen aan een studie, het studieprotocol en de risico's en voordelen van hun deelname echt begrijpen. Als de patiënt niet wil deelnemen of de studie op een of ander ogenblik wil stopzetten, moeten de onderzoekers die wens respecteren.

Doorgaans wordt aan de proefpersonen gevraagd te bevestigen of ze uit vrije wil toestemmen om aan een gegeven studie deel te nemen nadat ze op de hoogte werden gebracht van alle aspecten van de studie die pertinent zijn om een beslissing terzake te nemen, en dat door middel van een gedateerd en ondertekend informed-consentformulier.

Tegenstanders van psychiatrisch onderzoek argumenteren evenwel dat de huidige beschermende maatregelen zoals een informed consent onvoldoende zekerheid bieden om psychiatrische patiënten te beschermen tegen wetenschappelijke uitbuiting. Aangetoond werd dat de ernst van de symptomen en de cognitieve uitvalsverschijnselen bij schizofrenie een negatieve invloed hebben op het vermogen zijn toestemming te geven na voorlichting (8).

Een extrapolatie van de resultaten van RCT naar de praktijk wordt vaak beperkt door de hoge graad van controle. Daarom worden almaar vaker naturalistische studies uitgevoerd. Naturalistische studies, doorgaans retrospectieve studies, kunnen waardevolle aanvullende

informatie geven omdat ze een reële situatie weerspiegelen. De resultaten van naturalistische studies kunnen de artsen dan ook helpen in de klinische praktijk. Maar doordat geen inclusie- en exclusiecriteria gespecificeerd zijn en doordat er geen randomisatie werd doorgevoerd, kan een dergelijk studieprotocol meer bias of vertekende factoren met zich meebrengen. Bijvoorbeeld, aangezien poly-medicatie niet kan worden uitgesloten, zou het kunnen dat het verschil tussen de groepen niet enkel toe te schrijven is aan het verschil tussen de onderzochte geneesmiddelen.

Zowel gestructureerde als flexibele/naturalistische studies zijn nodig om een volledig beeld te geven van de wijze van gebruik van nieuwe en bestaande geneesmiddelen. Als één type studie bevestigd wordt door een ander type, verkrijgt men een hogere graad van zekerheid.

Validiteit van het design

Onderzoek van de verschillende vormen van validiteit kan bijzonder nuttig zijn om de resultaten van een klinische studie te interpreteren en de gevolgtrekkingen te bepalen, en op grond daarvan kunnen dan besluiten worden getrokken.

Davidson (9) somt enkele frequente factoren op die de validiteit in het gedrang kunnen brengen:

- klein aantal & vooraf en achteraf gedefinieerde metingen,
- te veel drop-outs (proefpersonen die de studie verlaten),
- de behandeling waarmee wordt vergeleken, is irrelevant of niet geschikt,
- korte duur van de studie,
- gebrekkig blind studieprotocol, slechte registratie van de gegevens,
- ...

Interne validiteit: Verwijst naar de doeltreffendheid van het protocol, de procedures en de methoden van een studie om de mogelijke vertekende variabelen onder controle te houden als men concludeert dat de experimentele interventie verantwoordelijk is voor de waargenomen verschillen tussen de controlegroep en de experimentele groep. De interne validiteit is hoog in RCT-studies (strikte inclusiecriteria) en laag in naturalistische studies (ruime studieparameters).

Bijvoorbeeld strikte/vage diagnose, duur van de ziekte, polymedicatie, enzovoort.

Externe validiteit: de mate waarin een studie niet-vertekende gevolgtrekkingen kan opleveren met betrekking tot de doelpopulatie, dus een bredere populatie dan de studiepopulatie (bijvoorbeeld verschillen tussen de gecontroleerde psychosociale omgeving en de reële situatie). De externe validiteit is laag in RCT en hoog in naturalistische studies.

Zo worden vrouwen in de meeste studies uitgesloten. De resultaten van die studies kunnen dus niet noodzakelijk op vrouwen worden toegepast (zie verder).

Stemt de studiepopulatie overeen met de populatie die in de praktijk wordt gezien?

Conceptuele validiteit: verwijst naar de geschiktheid van de gebruikte protocols, methodes en metingen om de belangrijkste hypothesen te testen met een conceptuele integriteit. Is het instrument geschikt om te meten wat men wil weten?

Statistische validiteit: verwijst naar het nut van een aanpak voor gegevensanalyse om significante verschillen te detecteren zonder kunstmatige verschillen te creëren.

Bijvoorbeeld correctie voor de verschillen in leeftijd tussen de studiegroepen

Specifieke problemen

Verschillende niveaus van vergelijking

Het eerste niveau is dat van **rechtstreekse** (*face to face*) vergelijking van de nieuwe antipsychotica. Er bestaan twee soorten studies, namelijk RCT en naturalistische studies, waarover we het reeds hebben gehad. In het volgende hoofdstuk onderzoeken we enkele voorbeelden van rechtstreeks vergelijkende studies (4-6). Het is belangrijk de resultaten van RCT te vergelijken met die van naturalistische studies (10-12).

Buiten rechtstreeks vergelijkende studies kan men het antipsychoticum ook **vergelijken met eenzelfde ander product** zoals een placebo of haloperidol. Die studies zijn doorgaans studies voor de FDA (*Food & Drug Administration*), wat betekent dat in al die studies vergelijkbare criteria worden opgelegd. Een studie die 2 geneesmiddelen rechtstreeks met elkaar vergelijkt, geeft meer rechtstreekse informatie dan een vergelijking van de resultaten van elk geneesmiddel met die van eenzelfde ander product, maar toch is die laatste benadering nuttig om na te gaan of bepaalde resultaten van rechtstreeks vergelijkende studies niet worden tegengesproken als ze afkomstig zijn van één enkele bron of gewoonweg als die gegevens niet voorhanden zijn. De voorbeelden die dat illustreren, zijn de studies van Beasley et al (13) (olanzapine versus haloperidol), Marder et al (14) (risperidone versus haloperidol), Tollefson et al (15) (olanzapine versus haloperidol) en Peuskens & Link (16) (quetiapine versus chloorpromazine). Men kan ook een vergelijking doorvoeren gebaseerd op klinische ervaringen, **subjectieve observaties of indrukken** van de clinicus. Zo kan de doeltreffendheid bijvoorbeeld worden beoordeeld aan de beschrijving van de klinische evolutie door de arts in het dossier van de patiënt.

Selectie van de patiënten

De onderzoekers moeten de patiënten die aan een studie deelnemen, precies beschrijven, want de kenmerken van de patiënten kunnen een

invloed hebben op de observaties. Een zorgvuldig beschreven populatie maakt een juiste interpretatie van de resultaten mogelijk.

De patiëntselectie heeft vooral tot doel een representatief deel van de patiëntenpopulatie in kwestie te verkrijgen. Goed gedefinieerde inclusie- en exclusiecriteria moeten zorgvuldig worden opgesteld vooraleer het onderzoek te starten, ongeacht het studieprotocol. De inclusiecriteria definiëren de patiënten met de desbetreffende diagnose binnen vooraf bepaalde limieten zoals de leeftijd en andere factoren (17).

De **diagnose** van de patiënten die in een studie worden opgenomen, moet worden genoteerd en gedocumenteerd (18). Aan de meeste studies mogen ook patiënten met stoornissen die verwant zijn met schizofrenie deelnemen. Een score van de **ernst van de symptomen** op grond van de PANSS geeft een correcte en eenvoudige beschrijving van de patiëntengroep. Beperkingen qua **geslacht** en **leeftijd** kunnen tot gevolg hebben dat de studieresultaten die bij de gerandomiseerde patiënten werden behaald, moeilijk kunnen worden doorgetrokken naar een bredere populatie. Schizofrenie is een heterogene aandoening. Schizofrene patiënten vertonen allerhande symptomen die mettertijd kunnen veranderen. Bijvoorbeeld, als we de leeftijd van de onderzochte populatie niet kennen, verliezen de resultaten hun waarde, want we kunnen ze niet toepassen op onze individuele patiënten. Bijvoorbeeld, klinische studies worden vaker uitgevoerd bij mannen omdat de ziekte bij mannen vroeger begint, omdat mannen minder goed reageren op antipsychotica en omdat hun prognose slechter is (19). Vrouwen kunnen bovendien enkel in een klinische studie worden opgenomen als een zwangerschap wordt uitgesloten en voorkomen.

De **fase van de ziekte** waarin de patiënten zich bevinden, kan ook een invloed hebben op de resultaten van de klinische studies. Opgenomen patiënten lijden vaker aan een ernstige ziekte, bijvoorbeeld een eerste episode

van psychose of een psychose die chronisch recidiveert, dan ambulante patiënten die beter op het geneesmiddel reageren en nog vrij goed kunnen functioneren in de maatschappij. Anderzijds kan het ziekenhuispersoneel toezicht uitoefenen op de inname van de medicatie, wat de therapietrouw van opgenomen patiënten kan verbeteren.

Doorgaans wordt aangenomen dat een **vroegere behandeling met antipsychotica** de interpretatie van een aantal biologische stellingen in het onderzoek naar schizofrenie kan bemoeilijken (20). Men kan bijvoorbeeld een farmacologische tolerantie veronderstellen na een eerdere langdurige behandeling.

Daarom wordt de studie voorafgegaan door een inlooperperiode zonder geneesmiddel of een wash-outperiode. Maar zelfs na volledige stopzetting van het geneesmiddel kan het meerdere maanden duren voordat de patiënt een recidive vertoont. De incidentie van extrapiramidiaal syndroom in de placebogroepen bijvoorbeeld, kan worden toegeschreven aan persisterende neveneffecten na stopzetting van een antipsychoticum.

Kortom, een studie wordt bepaald door allerhande criteria die leiden tot selectie van de patiënten. Welke zijn de belangrijkste selectiecriteria?

- de diagnose van de patiënten die in de studie worden opgenomen, moet worden genoteerd en gedocumenteerd,
- leeftijd & geslacht,
- comorbiditeit,
- polymedicatie en abusief gebruik van substanties,
- klinische kenmerken:
- ernst van de symptomen (PANSS),
- ziektefase:
- chronische patiënten of eerste episode?
- acute exacerbatie of stabiele populatie?
- therapieresistente patiënten?
- vroegere behandeling?

Grootte van het monster

De proefpersonen die aan de inclusiecriteria voldoen, vormen de doelgroep en zijn afkomstig uit de bronpopulatie. Derhalve kan slechts een klein deel van de doelgroep echt worden gerandomiseerd. Gewoonlijk worden die gegevens niet geregistreerd noch gepubliceerd. Als de selectiecriteria gedefinieerd zijn en de patiënten naar de behandeling zijn gerandomiseerd, mogen er tussen de groepen geen statistische verschillen zijn of als dat zo zou zijn, dan moeten daar statistische correcties voor worden aangebracht.

Aantal patiënten dat de studie verlaat

De grootte van de doelgroep kan tijdens de studie verkleinen doordat patiënten de studie verlaten. Gewoonlijk gaat het om patiënten die om de een of andere reden beslissen van behandeling te veranderen, bijvoorbeeld wegens onvoldoende doeltreffendheid, neveneffecten of slechte therapietrouw. Bij inclusie in de studie moeten de patiënten doorgaans een *informed consent* ondertekenen. Zo weten ze dat ze aan een studie deelnemen en dat de kans bestaat dat ze een placebo zullen krijgen. Het zou kunnen dat die kennis hen ertoe brengt hun engagement in de studie te verbreken zodra er een probleempje rijst of als het effect op zich laat wachten. Het zou ook kunnen dat de neveneffecten minder goed worden verdragen. Het aantal drop-outs blijkt doorgaans vrij groot te zijn bij schizofrene patiënten.

Die factoren bemoeilijken sterk de interpretatie van het effect van de drop-outs op de resultaten.

Een adequate behandeling

De keuze van de dosering is belangrijk bij het vergelijken en onderzoeken van geneesmiddelen, maar de criteria voor selectie van de dosis zijn slecht gedefinieerd. We zouden graag steeds dezelfde dosis van een antipsychoticum voorschrijven, maar we weten dat één enkele dosis niet geschikt is voor alle

patiënten. Patiënten met een eerste episode van schizofrenie reageren vaak op een lagere dosis dan patiënten die reeds vele jaren ziek zijn. Bejaarde patiënten hebben een lagere dosis nodig dan jonge patiënten. Factoren zoals de leeftijd, het ziektestadium, de ernst van de ziekte en de vroegere behandeling hebben een weerslag op alle initiële psychopathologische metingen. Een geschikt aanpassingsschema, mogelijke medicamenteuze interacties en de nawerking van vroegere geneesmiddelen zullen een invloed hebben op de respons (21).

Sommige symptomen reageren minder snel op het geneesmiddel dan andere. Meting van de doeltreffendheid op alle aspecten van de schizofrenie is een groot probleem. Het zou naïef zijn te denken dat één enkele dosis van één enkel geneesmiddel optimaal zal werken op alle symptomen. Ook is het uiterst moeilijk een dosis-responsrelatie vast te stellen in een brede waaier van neveneffecten. Het bindingsprofiel van de nieuwe antipsychotica heeft een invloed op verschillende receptoren, wat het evalueren van de dosis-responsrelatie nog moeilijker maakt (21). Men moet onder meer weten hoe de therapietrouw wordt geëvalueerd.

Het studieprotocol en de validiteit ervan zijn belangrijk, maar dat geldt ook voor de effecten, de patiëntselectie, de keuze van de dosering — rekening houdend met de onderzochte populaties.

Kwantificering van de metingen

In psychiatrisch onderzoek is het moeilijk veranderingen in de symptomen of de klinische toestand exact te meten. Daarom worden zorgvuldig gevalideerde schalen gebruikt om zich een zo precies mogelijk idee te vormen van de aspecten die men wil meten. De onderstaande tabel toont schalen die frequent worden gebruikt om het resultaat te meten.

Aan de hand van de meetresultaten kan de hypothese worden getoetst aan de uiteindelijke conclusie van een studie. De meetresultaten worden gebruikt om de hypothese te toetsen en om de resultaten te beschrijven. Ze moeten de klinische realiteit weerspiegelen en de geformuleerde hypothese toetsen.

De *validiteit* van een evaluatieschaal is de mate waarin de schaal evalueert wat hij zou moeten evalueren. Bijvoorbeeld, een zeer hoge

Tabel 1: Frequent gebruikte schalen.

- Schalen die een totaalindruk van de klinische toestand geven:
 - CGI (*Clinical Global Impression*)
 - GAF (*Global Assessment of Functioning scale*)
- Schalen die een gedetailleerder beeld van de klinische toestand schetsen:
 - PANSS (*Positive And Negative Syndrome Scale*)
 - BPRS (*Brief Psychiatric Rating Scale*)
 - SANS (*Scale for Assessment of Negative Symptoms*)
- Schalen die de neveneffecten meten:
 - ESRS (*Extrapyramidal Symptom Rating Scale*)
 - UKU (*Side Effect Rating Scale*)
 - BAS (*Barnes Akathisia Scale*)
 - AIMS (*Abnormal Involuntary Movement Scale*) (dyskinesie)
 - SAS (*Simpson and Angus Scale*) (Parkinsonisme)
- Schaal die alle vernoemde aspecten meet:
 - PECC (*Psychosis Evaluation tool for Common use by Caregivers*)
- Schalen die de levenskwaliteit meten:
 - SF-36 (*Health Survey*)
 - SQLS (*Quality of Life Scale*)

BPRS-score zou moeten overeenstemmen met een zeer intense psychose bij een grondige psychiatrische evaluatie door een ervaren psychiater. De betrouwbaarheid van een schaal weerspiegelt zijn vermogen om reproduceerbare coherente informatie te leveren. De inter-observer-betrouwbaarheid is de mate van concordantie tussen evaluaties die op hetzelfde ogenblik door verschillende personen worden uitgevoerd. De test-retest-betrouwbaarheid meet de concordantie tussen evaluaties die door een zelfde onderzoeker op verschillende tijdstippen werden uitgevoerd. De betrouwbaarheid van een schaal wordt weergegeven door een *coëfficiënt*. 0 is niet betrouwbaar en 1 is perfect betrouwbaar. Een betrouwbaarheid van meer dan 0,85 wordt doorgaans als adequaat beschouwd.

In veel studies worden de resultaten van de behandeling niet beschreven vanuit het oogpunt van de patiënt.

We hebben metingen nodig die de belangrijke klinische symptomen definiëren en tellen. Tellen van een symptomen en meten van de tijd tot optreden van een evenement zijn mogelijke strategieën die verder reiken dan de klassieke variatie in percentage op een evaluatieschaal van het beginpunt tot het eindpunt (*endpoint* of laatste evaluatie van de resultaten).

De klinische studies evalueren de doeltreffendheid van een interventie, terwijl de klinische praktijk vaak tracht een negatieve afloop te voorkomen of een verblijf in het ziekenhuis te voorkomen, liever dan resultaten te behalen die gericht zijn op de patiënt, waarbij de nadruk wordt gelegd op de doeltreffendheid van het geneesmiddel (22).

Het blijft moeilijk de resultaten te definiëren en te meten. Schizofrenie is een heterogene ziekte met verschillende outcomedomeinen. De laatste jaren is er een conflict gerezen

tussen de wens om meerdere domeinen te meten (gezien de complexiteit van de ziekte en het aantal verschillende domeinen van functioneren die worden aangetast) en de verwachting dat één of twee studieprotocollen staat zijn interpreteerbare en betekenisvolle antwoorden te geven op al die vragen.

Geen enkele evaluatieschaal is zo gevoelig dat men zich in een klinische studie met een atypisch antipsychoticum alleen tot die schaal kan beperken. Om dat te illustreren, vergelijken we in onderstaande kader 2 schalen die vaak worden gebruikt.

In een artikel bespreken Bell et al (24) de betrouwbaarheid, de vergelijkbaarheid en de predictieve validiteit van de PANSS en de BPRS. Beide instrumenten hebben een hoge inter-observer-betrouwbaarheid, maar de PANSS blijkt betrouwbaarder te zijn wat de gemeenschappelijke items en de aanvullende items van de PANSS betreft. Dit suggereert dat de symptomen beter worden beschreven in de PANSS dan in de BPRS.

Bovendien maakt de PANSS een onderscheid tussen de ernst en de frequentie van optreden van een symptoom – elementen die in de BPRS in één enkel punt worden gecombineerd. De PANSS geeft echter weinig informatie over twee belangrijke domeinen bij schizofrenie, namelijk de gemoedsstemming en de cognitieve functies.

De interne coherentie van de syndroomschalen van de PANSS, de algemene schaal en het instrument in zijn geheel blijkt beter te zijn dan die van de BPRS. Dat suggereert dat de aanvullende items van de PANSS en de wijziging van de gemeenschappelijke items de homogeniteit van de syndroomschalen verhoogd hebben.

Bij onderzoek van de vergelijkbaarheid van de gemeenschappelijke items blijkt dat de aanpassingen de betekenis van de items dermate gewijzigd hebben dat de meeste onderling niet kunnen worden uitgewisseld.

De PANSS heeft een betere voorspellende waarde wat de arbeidsprestaties betreft.

De auteurs besluiten dat hun resultaten pleiten voor het verder gebruik van de PANSS bij het onderzoek naar schizofrenie. De

De BPRS en de PANSS

BPRS

De BPRS omvat 18 (oorspronkelijk 16) groepen symptomen en van elke groep wordt de ernst beoordeeld op een schaal van 7 punten. De BPRS werd in het begin van de jaren '60 ontwikkeld om ziekenhuispatiënten snel en goed te kunnen evalueren, zowel bij het evalueren van geneesmiddelen als in de gewone klinische context. De BPRS is een van de oudste schalen die veel wordt gebruikt. De resultaten van studies die de BPRS gebruiken, kunnen vaak worden vergeleken met de resultaten van vroegere studies. Maar de BPRS heeft een belangrijke beperking: hij werd niet ontworpen om negatieve symptomen te meten. En die negatieve symptomen zijn zeer belangrijk bij het evalueren van atypische antipsychotica. Het CPMP (*Committee for proprietary medicinal products*) raadt aan de BPRS niet meer te gebruiken om patiënten met negatieve symptomen te evalueren.

PANSS

De PANSS werd ontworpen als een operationeel instrument dat gevoelig is voor geneesmiddelen, de positieve en de negatieve symptomen evenwichtig weergeeft en hun onderlinge relatie en hun relatie met de hele psychopathologie evalueert. De PANSS werd ontwikkeld door Kay et al (23) en bestaat uit schalen die 7 positieve items en 7 negatieve items meten, en een schaal van 16 items van algemene psychopathologie (zie lijst). In totaal werden 18 items van de PANSS overgenomen van de BPRS en 12 items zijn aanpassingen van de *Psychopathology Rating Schedule* (PRS) van Singh & Kay. De schaal van algemene psychopathologie telt items die 'het gebrek aan medewerking', 'het actieve sociale vermijdingsgedrag' en 'het gebrek aan aandacht' omvatten die betrekking hebben op een individuele functionaliteit, maar de PANSS meet niet rechtstreeks het algemene vermogen om te functioneren.

PANSS is een instrument met een bijzondere psychometrische stabiliteit en heeft een hoge voorspellende waarde. Om dat te illustreren, geven we het voorbeeld van hallucinatie dat we hebben gehaald uit het evaluatieboek van de PANSS (Figuur 1).

De keuze van de schalen levert zeer uiteenlopende resultaten op, gaande van een totale of gedetailleerde klinische eva-

luatie tot een observatie van de neveneffecten of het geheel van de genoemde elementen.

Non-inferioriteitsstudies

Een non-inferioriteitsstudie gaat uit van de veronderstelling dat een zeker verschil aanvaard wordt als een niet-significant/niet-relevant verschil. Als het verschil kleiner is dan de vooraf

bepaalde drempel, wordt het product beschouwd als niet minder goed dan het andere. De vooraf bepaalde drempel is dus zeer belangrijk: hij moet zo klein mogelijk zijn, maar moet redelijk blijven. De primaire conclusie kan enkel zijn dat A beter is dan B of dat A niet minder goed is dan en dus gelijk is aan B. Enkel A kan beter zijn dan B en niet omgekeerd.

p-waarde

In die studie gingen Tran et al (4) dus uit van de hypothese dat olanzapine beter was dan risperidone. Non-inferioriteitsstudies worden nu meer gebruikt.

Ook al tonen de gegevens van een klinische studie aan dat er een verschil is tussen de twee onderzochte groepen (de nulhypothese wordt verworpen), dan zou het toch nog kunnen dat het waargenomen resultaat toe te schrijven is aan het toeval. Met andere woorden, het zou kunnen dat er eigenlijk geen verschil is, hoewel in de studie een verschil wordt vastgesteld. Die waarschijnlijkheid wordt uitgedrukt door de p-waarde. Hoe kleiner de p-waarde, hoe groter de waarschijnlijkheid dat de nulhypothese terecht kan worden verworpen en dat er een echt verschil is. Een p-waarde van 0,05 betekent dat de waarschijnlijkheid dat het waargenomen verschil toevallig is of enkel toe te schrijven is aan het toeval, 5% bedraagt.

Figuur 1: Sample item on "Hallucinatory behaviour" from the PANSS Rating Manual.

Hallucinatory behavior: Verbal report indicating perceptions which are not generated by external stimuli. These may occur in the auditory, visual, olfactory, or somatic realms. Basis for rating: verbal report and physical manifestations during the course of interview as well as report of behavior by primary care workers or family.

1. Absent – definition does not apply.
2. Minimal – questionable pathology; may be at the upper extreme of normal limits.
3. Mild – one or two clearly formed but infrequent hallucinations, or else a number of vague abnormal perceptions which do not result in distortions of thinking or behavior.
4. Moderate – hallucinations occur frequently but not continuously, and the patient's thinking and behavior are affected only to a minor extent.
5. Moderate severe – hallucinations are frequent, may involve more than one sensory modality, and tend to distort thinking and/or disrupt behavior. Patient may have a delusional interpretation of these experiences and respond to them emotionally and, on occasion, verbally as well.
6. Severe – hallucinations are present almost continuously, causing major disruption of thinking and behavior. Patient treats these as real perceptions, and functioning is impeded by frequent emotional and verbal responses to them.
7. Extreme – patient is almost totally preoccupied with hallucinations, which virtually dominate thinking and behavior. Hallucinations are provided a rigid delusional interpretation and provoke verbal and behavioral responses, including obedience to command hallucinations.

bepaalde drempel, wordt het product beschouwd als niet minder goed dan het andere. De vooraf bepaalde drempel is dus zeer belangrijk: hij moet zo klein mogelijk zijn, maar moet redelijk blijven. De primaire conclusie kan enkel zijn dat A beter is dan B of dat A niet minder goed is dan en dus gelijk is aan B. Enkel A kan beter zijn dan B en niet omgekeerd.

Statistische analyse

Welke hypothesen?

Superioriteitsstudies

Superioriteitsstudies vormen de referentie bij het vergelijken van twee of meer behandelingen en gaan daarbij uit van de nulhypothese, de hypothese die stelt dat er geen verschil is tussen de 2 (of meer) onderzochte groepen. De nulhypothese wordt getoetst aan de alternatieve hypothese, de hypothese die zegt dat er wel een verschil is tussen de 2 (of meer) onderzochte groepen. Het

Dat wordt mooi geïllustreerd door enkele citaten uit de studie van Tran et al (4), die gebaseerd is op een non-inferioriteitsstudie met olanzapine versus risperidone: "A lower bound > 0 indicated that olanzapine was statistically **superior** to risperidone." "A lower bound ≥ -4 and ≤ 0 indicated there was **no difference** between the two treatments." "A lower bound < -4 indicated **no conclusion** could be drawn."

De p-waarde wordt vergeleken met een significantiedrempel, α -drempel genoemd, die in het begin van het onderzoek doorgaans arbitrair op 0,05 wordt vastgelegd. Als $p < 0,05$, dan kan men besluiten dat er tussen de groepen een statistisch significant verschil is wat de gemiddelde respons betreft. De α -drempel is de waarschijnlijkheid dat ten onrechte wordt besloten dat de nulhypothese verkeerd is (dus een type-I-fout, dat wil zeggen de mogelijkheid dat de observaties te wijten zijn aan het toeval). De significantiedrempel van de p-waarde is een arbitraire beslissing, maar wordt gewoonlijk vastgelegd op 5%.

α -drempel of significantiedrempel

Die tests analyseren de afwijkingen ten opzichte van de hypothese. Een bilaterale test kan dergelijke afwijkingen in de ene of andere zin analyseren. Een bilaterale

Uni- en bilaterale tests

Die tests analyseren de afwijkingen ten opzichte van de hypothese.

Een bilaterale test kan dergelijke afwijkingen in de ene of andere zin analyseren. Een bilaterale

test heeft een grotere validiteit dan een unilaterale test en vereist een groter verschil om de drempel van statistisch significantie te bereiken. Een unilaterale test is enkel geschikt als aangenomen wordt dat de zin van het verschil tussen de 2 groepen op voorhand met zekerheid bekend is, of als waargenomen verschillen in tegengestelde zin niet belangrijk of onmogelijk zijn. Een unilaterale test toetst enkel de verschillen in één enkele vooraf bepaalde zin en houdt geen rekening met verschillen in de andere zin.

Unilaterale tests worden doorgaans enkel gebruikt in een non-inferioriteitsstudie (zie hoger). Een dergelijke studie veronderstelt immers dat een louter unidirectionele superioriteit het enige mogelijke primaire besluit is.

De studie van Tran et al (4) is een voorbeeld van een unilaterale non-inferioriteitsstudie en de studie van Conley & Mahmoud (5) is een voorbeeld van een bilaterale superioriteitsstudie.

Interpretatie van de resultaten

Selectie van de patiënten – aantal drop-outs

De selectie van de patiënten kan een invloed hebben op de resultaten van klinische studies. Als een te klein aantal patiënten aan de inclusiecriteria voldoet, wordt het moeilijk de resultaten later te veralgemenen. Bij de selectie moet worden gespecificeerd hoeveel patiënten in beschouwing worden genomen om te weten of de doelgroep representatief is voor de betrokken patiëntenpopulatie en in welke mate de resultaten kunnen worden veralgemeend. Beperkingen wat het geslacht en de leeftijd betreft, kunnen tot gevolg hebben dat de resultaten afkomstig van gerandomiseerde patiënten slechts in beperkte mate kunnen worden doorgetrokken naar een bredere populatie – bijvoorbeeld uitsluiting van middelenmisbruik in studies over schizofrenie. Bij de evaluatie van de resultaten van klinische studies moet ook rekening worden gehouden met het aantal drop-outs (aantal patiënten dat de

studie verlaat). Als de doelgroep te klein wordt, kan dat ernstige methodologische problemen met zich meebrengen (18). Het aantal drop-outs in placebogecontroleerde studies is klassiek hoog: de meeste patiënten die een placebo krijgen, zullen immers niet reageren. De patiënten zijn doorgaans ook weinig geneigd om een geneesmiddel dat niet werkt of neveneffecten veroorzaakt, verder in te nemen. Het begrip drop-out moet vooraf zorgvuldig worden gepreciseerd en kan als afhankelijke variabele worden gebruikt bij het vergelijken van de verschillende behandelingen (18).

Dosis

Als 2 geneesmiddelen in klinische studies worden vergeleken, kan het percentage respons op elk geneesmiddel variëren naargelang van de dosis. Het zou kunnen dat de patiënt verbetert met een gegeven dosis, maar niet met andere doses. Omgekeerd kan een dosis die hoger is dan de adequate dosis van een geneesmiddel, een significante verbetering met de adequate dosis maskeren. We moeten dan ook voor ogen houden dat, na verloop van tijd, niet alle patiënten op dezelfde manier reageren op dezelfde dosis van het geneesmiddel (21).

De dosering als dusdanig is niet belangrijk, maar er moet rekening mee worden gehouden in relatie met de geselecteerde patiëntenpopulatie. Het studieprotocol kan een invloed hebben op de eventuele dosis.

In gecontroleerde studies worden de onderzochte geneesmiddelen steeds in monotherapie gebruikt binnen hun . Zo zal men in studies met antipsychotica nooit meerdere antipsychotica tegelijk geven. Anders zou men geen conclusies kunnen trekken. Daar moet rekening mee worden gehouden bij het extrapoleren van de resultaten van de studie naar de klinische praktijk: in de klinische praktijk worden immers zeer vaak meerdere geneesmiddelen tegelijk gegeven (polymedicatie) (25).

In niet-gecontroleerde studies daarentegen, kan betrouwbare informatie worden verza-

meld over de gemiddelde dosering bij het dagelijkse gebruik.

Statistische significantie

Statistische significantie is niet hetzelfde als *klinische significantie*. De statistische significantie weerspiegelt de invloed van het toeval op het resultaat. Als we kunnen besluiten dat de resultaten statistisch significant zijn, dan betekent dat dat de verschillen die tussen de groepen worden waargenomen met die specifieke metingen, niet toe te schrijven zijn aan het toeval; niets meer en niets minder. De waarschijnlijkheid dat het verschil te wijten is aan het toeval, is kleiner dan 5%. Als het verschil niet statistisch significant is, betekent dat nog niet dat de 2 geneesmiddelen identiek zijn. Het betekent alleen dat bij een specifieke meting in een specifieke groep geen statistisch verschil werd vastgesteld (3).

De klinische significantie weerspiegelt de *klinische waarde* van het resultaat. Statistische resultaten zijn afkomstig van groepen individuen, terwijl artsen aparte individuen behandelen.

Bij het toetsen van de hypothese wordt steeds de nadruk gelegd op de p-waarde en de significantie, maar men zou dat kunnen aanvullen met andere methoden van rapportering van de resultaten. Bij de evaluatie van een klinische studie wordt echter nog steeds het meeste gebruik gemaakt van de p-waarde. Toch kan geen enkel studieprotocol een antwoord geven op alle pertinente vragen over een geneesmiddel. Het geheel van de gegevens die worden verkregen uit meerdere soorten studies, levert de beste informatie op voor de geneeskundige praktijk. Elke studie kan een zeer beperkte hoeveelheid informatie geven, namelijk de informatie die de studie zou moeten meten. Dat kan beperkingen met zich meebrengen op intern vlak (de variabelen kunnen een effect hebben dat niet wordt herkend, gecontroleerd of gemeten) of op extern vlak (individuele observaties kunnen slechts in beperkte mate worden geëxtrapoleerd door variabelen zoals de leeftijd, het geslacht en het ziektestadium). De lezer moet die beperkingen begrijpen. Klinische beslissingen

worden immers vaak gebaseerd op besluiten van klinische studies (26). Artsen hebben baat als ze kunnen beschikken over informatie uit meerdere bronnen. Dan kunnen ze het werkingsmechanisme beter begrijpen van de geneesmiddelen die worden ingezet om de pijn en het lijden te verlichten in geval van een acute en chronische psychose (27).

Analoge resultaten uit gecontroleerde en naturalistische studies

Als de hypothese niet wordt geformuleerd, moet de lezer weten dat het een verkennend onderzoek betreft. Als die informatie ontbreekt, moet de lezer voorzichtig zijn bij het interpreteren van de resultaten. De observaties mogen zeker niet als definitief worden beschouwd, maar onder voorbehoud van bevestiging.

We hebben immers gegevens nodig die worden gestuurd door een hypothese, en observationele gegevens van naturalistische studies om ten volle te kunnen begrijpen hoe we de nieuwe geneesmiddelen moeten gebruiken. Als de resultaten van de twee soorten studies elkaar bevestigen, worden de bewijzen in het voordeel van de hypothese steviger. Met andere woorden, men kan moeilijk besluiten trekken uit één enkele studie; vergelijkbare observaties in meerdere soorten studies zijn noodzakelijk.

Last observation carried forward (LOCF)

Bij de interpretatie van de resultaten van een klinische studie moet men nagaan of de drop-outs werden meegeteld.

Nemen we bijvoorbeeld een populatie van 100 patiënten die kunnen worden behandeld (ITT of *intention to treat*), maar slechts 80 patiënten werden effectief in de studie opgenomen en werden na de beginsituatie minstens eenmaal geëvalueerd. 30 van de 80 patiënten hebben de studie voortijdig stopgezet. Op het einde van de studie blijven er dan nog maar 50 patiënten over. Als de analyse enkel wordt uitgevoerd op de patiënten die op de behandeling gereageerd hebben (50 patiënten), mag men een positiever resultaat

verwachten dan als ook de drop-outs worden meegeteld (80 patiënten). Men kan dus 2 soorten analyses uitvoeren: een analyse van de geobserveerde gevallen (*observed case analysis*), een analyse van de patiënten die op de behandeling gereageerd hebben, en een LOCF analyse, waarbij de laatste (en soms de enige) meting die werd uitgevoerd bij de patiënten die de studie verlaten hebben, gebruikt wordt als eindpunt (*endpoint*). Een LOCF-analyse geeft dus ook informatie over de patiënten die niet op de behandeling hebben gereageerd. Een LOCF-analyse geeft dus een betere raming van het effect dat de behandeling zal hebben op een willekeurige volgende patiënt.

Als de twee analyses gelijkaardige resultaten opleveren, kan men er zeker van zijn dat de resultaten van de analyse van de geobserveerde gevallen niet afhangt af van de drop-outs. Hoe langer de studie duurt, hoe groter de waarschijnlijkheid dat de observaties in beide analyses zullen verschillen. In een studie die langer dan een jaar duurt, zullen de geobserveerde gevallen een fijner geselecteerde groep vormen dan in een kortere studie.

Neveneffecten

Bij het evalueren van de neveneffecten dient men enkele aspecten voor ogen te houden. Zo moet precies worden gemeld of de neveneffecten spontaan door de patiënt werden gerapporteerd dan wel of ze met instrumenten werden gemeten. Een andere vraag is hoe de neveneffecten in de placebogroep moeten worden geïnterpreteerd. Die neveneffecten, die men duidelijk moet onderscheiden van de klinische symptomatologie, zijn soms te wijten aan de vorige behandeling die abrupt werd onderbroken, met ontweningsverschijnselen als gevolg. Zo kunnen zelfs bij meting van de neveneffecten voor inclusie van de patiënten in de studie extrapiramidale neveneffecten door vroegere toediening van antipsychotica worden gezien.

Men moet rekening houden met de grootte van de doelgroep, bijvoorbeeld in verhouding tot het aantal drop-outs, en ook met de toegediende dosis, de significantie van de resultaten en het feit of die al dan niet in andere studies werden bevestigd.

Enkele voorbeelden van kritische lectuur van klinische studies

Tot dusver hebben maar enkele studies atypische antipsychotica onderling rechtstreeks vergeleken. In deze paragraaf bespreken we twee vergelijkende studies met olanzapine en risperidone (4, 5). De eerste studie werd uitgevoerd met de steun van *Eli Lilly and Company*, de tweede met de steun van de *Janssen Research Foundation*. Een andere studie (6) werd voor

Samenvatting Tran et al (4) van Rogers (28)

De eerste vergelijking tussen olanzapine en risperidone werd gepubliceerd in 1997 (4). Die multicentrische prospectieve dubbelblinde studie van 28 weken werd uitgevoerd bij 339 patiënten met schizofrenie, een schizofreniforme stoornis of een schizoaffectieve stoornis. De studie werd gestart vooraleer olanzapine in de Verenigde Staten op de markt werd gebracht. Vijfendertig onderzoekers uit negen landen hebben de patiënten gerandomiseerd naar olanzapine (n = 172) of risperidone (n = 167). Olanzapine werd gestart in een dosering van 15mg/dag gedurende zeven dagen. Daarna konden de artsen de dosis stapsgewijze verhogen of verlagen met 5mg om de zeven dagen, 10 tot 20mg/dag. Risperidone werd gestart in een dosering van 1mg tweemaal per dag op dag 1, 2mg tweemaal per dag op dag 2 en 3mg tweemaal per dag van dag 3 tot dag 7. Daarna mochten de onderzoekers de dosering stapsgewijze verhogen of verlagen met 2mg/dag om de zeven dagen, maar ze moesten binnen een bereik van 4 tot 12mg/dag blijven. De gemiddelde dosis van olanzapine bedroeg 17,2mg/dag en die van risperidone 7,2mg/dag. De doeltreffendheid werd gemeten met de totale PANSS en zijn subschalen, het item depressie van de PANSS, de BPRS met 18 items, de CGI-S en de SANS.

een deel gesteund door het NIMH (*National Institute of Mental Health*).

Kritische commentaren

Er werden meerdere kritieken op die studie gepubliceerd. Volgens Kasper (12) benadeelt het studieprotocol risperidone wat de dosering betreft. De dosering die voor risperidone werd gekozen, was gebaseerd op de richtlijnen die van kracht waren op het ogenblik dat de studie op touw werd gezet, namelijk tot een maximumdosis van 12mg/dag. Die dosering stemt niet meer overeen met de huidige klinische praktijk. Dat punt wordt ook vermeld in een artikel van Gheuens & Grebb (29).

Een post-hoc-analyse van Kasper suggereert voorts dat er een significant verschil in een aantal items was tussen de groepen. De initiële beknopte SANS-score was significant hoger in de olanzapinegroep dan in de risperidonegroep. Dat roept twijfels op over de statistische significantie van het verschil in SANS-score op het einde van de studie, verschil dat werd berekend zonder correctie (30). De besluiten betreffende andere items blijven evenwel geldig.

In de studie van Tran et al (31) hebben de auteurs een unilaterale non-inferioriteitstest gebruikt zonder statistische correctie voor multiële vergelijkingen door te voeren (32). De test onderzocht de hypothese dat olanzapine beter is dan risperidone, en niet een nul-hypothese (d.w.z. geen verschil tussen

beide geneesmiddelen). Die methode mag worden gehanteerd voor een kleinere doelgroep, maar een bilaterale test met correcties zou beter geweest zijn en zou een groter gewicht hebben gegeven aan de studie. Het is immers de relatieve doeltreffendheid van de behandelingen die ons interesseert en niet zozeer de zogenaamde superioriteit van een van beide.

De auteurs zeggen ook dat er geen statistisch significant verschil was tussen de 2 behandelingen wat het aantal ziekenhuisopnames voor de studie betreft, en dat er een verschil is ontstaan in het aantal ziekenhuisopnames na de studie. Maar ze zeggen niet of dat verschil statistisch significant was.

Rogers (28) besluit dat zowel olanzapine als risperidone de ernst van de totale psychotische symptomen vermindert. Maar slechts twee van de acht berekende p-waarden die de gemiddelde variatie ten opzichte van het begin tussen de behandelingsgroepen vergeleken (d.w.z. risperidone versus olanzapine), waren statistisch significant, namelijk het item depressie van de PANSS (daling met resp. -1,1 met olanzapine en -0,7 met risperidone) en de beknopte SANS-score (daling met resp. -4,3 met olanzapine en -2,9 met risperidone), waarbij er evenwel reeds in het begin een verschil was. In de risperidonegroep werd een hogere incidentie van extrapiramidale symptomen, hyperprolactinemie en seksuele stoornissen vastgesteld. Dat kan

worden toegeschreven aan de hoge gemiddelde dosis van 7,2mg/dag. Wat de andere neveneffecten betreft, was er tussen de groepen enkel een statistisch significant verschil in nausea (in de risperidonegroep) en gewichtstoename (in de olanzapinegroep).

Rogers (28) geeft de volgende commentaar:

We moeten twee punten belichten: de duur van de studie (8 weken) was korter dan in de

Abstract Ho et al (6) van Rogers (28)

Ho et al (6) stelden een open studie voor, waarin oorspronkelijk 42 schizofrene patiënten werden opgenomen (21 olanzapine, 21 risperidone). De patiënten werden in totaal 6 maanden gevolgd. Slechts 2/3 van het oorspronkelijke aantal patiënten was beschikbaar voor follow-up. Twee patiënten konden op het einde van de studie niet opnieuw worden geëvalueerd.

De gemiddelde dosis bij ontslag bedroeg 13,7mg/dag (SD = 4,85) voor olanzapine en 6,0mg/dag (SD = 1,15) voor risperidone. De neveneffecten werden gemeten met de Simpson-Angusschaal en de Barnes-Acathiaschaal. Tijdens de acute behandeling veroorzaakten beide geneesmiddelen weinig parkinsonisme, maar risperidone veroorzaakte vaker acathisie. Er was geen verschil in extrapiramidale symptomen tussen de verschillende behandelingsgroepen, zelfs als rekening wordt gehouden met het feit dat er meer anticholinergica werden gebruikt in de risperidonegroep.

De doeltreffendheid werd gemeten met de SANS, de SAPS, de BPRS, de GAS en een schaal voor meting van de levenskwaliteit. Tijdens de follow-up verbeterde de psychopathologie significant in beide groepen.

De negatieve symptomen verminderden in beide groepen, maar de verschillen waren niet significant lager dan in het begin van de studie. De psychotische en gedesorganiseerde symptomen verminderden significant meer in de risperidonegroep. Daaruit leiden de onderzoekers af dat risperidone beter zou zijn bij patiënten met ernstige wanen en hallucinaties. Dat effect werd zelfs gezien als de gemiddelde dosis daalde van 5,7mg/dag bij ontslag naar 4,5mg/dag na 6 maanden follow-up.

Tot dusver werd geen kritiek gepubliceerd op het artikel van Ho et al (6).

Abstract Conley & Mahmoud (5) van Rogers (28)

Conley en Mahmoud (5) hebben de resultaten gerapporteerd van hun dubbelblinde gerandomiseerde studie die de veiligheid en de doeltreffendheid van olanzapine en risperidone onderzocht bij 377 patiënten met schizofrenie of een schizoaffectieve stoornis. De patiënten werden gerandomiseerd naar risperidone (n = 188) of olanzapine (n = 189). De dosering van risperidone bedroeg 2mg/dag op dag 1 en 2, 2mg/dag tot 4mg/dag van dag 3 tot 7, 2mg/dag tot 6mg/dag van dag 8 tot 14 en 2mg/dag tot 6mg/dag van dag 15 tot 56. De dosering van olanzapine bedroeg 10mg/dag op dag 1 en 2, 5mg/dag tot 10mg/dag van dag 3 tot 7, 5mg/dag tot 15mg/dag van dag 8 tot 14 en 5mg/dag tot 20mg/dag van dag 15 tot 56. De gemiddelde dosering van olanzapine bedroeg 12,4mg/dag (SD = 4,6), de gemiddelde dosering op het einde van de studie bedroeg 13,1mg/dag (SD = 5,1).

Tot dusver werd geen enkele kritiek gepubliceerd op het artikel van Conley en Mahmoud (5).

andere twee, en de gemiddelde dosis van risperidone in die studie (4,8mg/dag; SD = 1,2) was significant lager dan de gemiddelde dosis van 7,2mg/dag in de studie van Tran et al (4). De dosis van risperidone in de studie van Conley en Mahmoud (5) stemt beter overeen met de dosis die nu in de klinische praktijk wordt gegeven. De gemiddelde dosis van risperidone op het einde van de studie bedroeg 4,7mg/dag (SD = 1,4).

De doeltreffendheid werd beoordeeld aan de totale PANSS-score, aan vijf factoren van de PANSS-score en aan de CGI-schaal. De auteurs besluiten dat in beide behandelingsgroepen een significante verbetering (gedefinieerd als een daling van de totale PANSS-score > 0%) werd waargenomen. De positieve symptomen en de symptomen van angst/depressie waren na 8 weken meer verbeterd met risperidone ($p < 0,05$). Olanzapine is er bij geen enkele meting beter uitgekomen dan risperidone.

De onderzoekers rapporteerden echter meer ernstige neveneffecten in de olanzapinearm dan in de risperidonearm (11,6% versus 8,0%). Bovendien werd een gewichtstoename met 7% waargenomen bij 27,3% van de patiënten in de olanzapinegroep en bij 11,6% van de patiënten die met risperidone werden behandeld.

Tegenstrijdige resultaten of niet?

Uitgaande van de gegevens van Conley & Mahmoud (5) besluiten de onderzoekers dat beide geneesmiddelen goed worden verdragen en doeltreffend zijn. Risperidone had na 8 weken een sterker effect op de positieve en de negatieve symptomen (angst en depressie). In de studie van Tran et al (4) werd geen verschil waargenomen tussen risperidone en olanzapine wat de positieve symptomen betreft, maar de SANS-score van de negatieve symptomen was significant beter in de olanzapinegroep. De auteurs

merkten op dat de initiële SANS-score hoger was in de olanzapinegroep. Daarom kan het waargenomen verschil een sterkere toevallige regressie weerspiegelen. De studie van Ho et al (6) treedt de bevindingen van Conley & Mahmoud (5) gedeeltelijk bij, namelijk dat risperidone na 6 maanden behandeling een sterker effect heeft op de negatieve symptomen dan olanzapine.

In de studie van Conley & Mahmoud (5) werd geen significant verschil in extrapiramidale symptomen vastgesteld in tegenstelling tot de studie van Tran et al (4). Zoals reeds werd gezegd, gebruikten Tran et al (4) een heel wat hogere dosis risperidone dan de dosis die de laatste jaren in de klinische praktijk wordt gegeven. Hogere doses risperidone veroorzaken vaker extrapiramidale neveneffecten. De resultaten van de studie van Tran et al zijn dan ook niet verrassend. De studie van Ho et al (6) suggereert dat lagere doses risperidone even doeltreffend zijn als die van olanzapine zonder meer parkinsonisme te veroorzaken. Olanzapine veroorzaakte echter minder acathisie.

Olanzapine kan gezondheidsrisico's inhouden: in beide studies steeg het lichaamsgewicht immers significant meer met olanzapine.

Overtuigende of tegenstrijdige bewijzen uitgaande van studies waarin met een zelfde product wordt vergeleken?

Meta-analyses kunnen gestandaardiseerde metingen geven van de omvang van de effecten van de behandeling zodat men nieuwe geneesmiddelen gemakkelijker onrechtstreeks met elkaar kan vergelijken. Een meta-analyse van Leucht et al (33) vat de doeltreffendheid en de tolerantie van een aantal recente antipsychotica samen in vergelijking met de placebo en met de klassieke antipsychotica. Risperidone, olanzapine, sertindol en quetiapine zijn alle 4 matig efficiënter dan de placebo. Risperidone en olanzapine zijn iets efficiënter dan haloperidol in de behandeling van de totale schizofrene symptomatologie. Alle nieuwe

antipsychotica hebben een sterker effect op de negatieve symptomen dan de placebo. Risperidone en olanzapine zijn significant efficiënter dan haloperidol. Tot slot, de patiënten die nieuwe antipsychotica nemen, hebben minder antiparkinsonmiddelen nodig dan de patiënten die haloperidol krijgen.

De afwijkende resultaten van Tran en Conley worden niet tegengesproken door analyses van studies die de twee geneesmiddelen vergeleken met een placebo of haloperidol.

Overtuigende of tegenstrijdige bewijzen uitgaande van naturalistische studies?

Doorgaans tonen de resultaten van naturalistische studies ook aan dat zowel risperidone als olanzapine doeltreffend zijn en goed worden verdragen bij schizofrenie. Maar in een aantal studies is risperidone statistisch significant beter dan olanzapine (10, 11, 34).

Conclusie

De vraag moet duidelijk worden gesteld vooraleer een antwoord kan worden gegeven. De onderzoekers hebben de verantwoordelijkheid de hypothesen en de doelstellingen van de studie te specificeren. Zowel het studieprotocol, de selectie van de meetinstrumenten, de patiëntenpopulatie als de dosering moeten op die doelstellingen gericht zijn. De analyses en de interpretaties van de gegevens moeten hiërarchisch worden geordend om een antwoord te kunnen geven op de primaire hypothese. De lezers moeten die criteria controleren bij het lezen van de resultaten.

Aangezien klinische studies van vitaal belang zijn voor een efficiënt en veilig gebruik van nieuwe geneesmiddelen, moet men de problemen begrijpen die door de studies worden aangesneden de onderzochte populatie en de vragen waarop de studie een antwoord geeft. Men moet voor ogen houden dat geen enkel studieprotocol een antwoord kan geven op alle pertinente vragen over de effecten

Figuur 2: Vergelijking van 3 vergelijkende studies met risperidone en olanzapine afkomstig van 3 verschillende bronnen (2 farmaceutische bedrijven en 1 universitaire instelling).

	Tran et al (4)	Conley & Mahmoud (5)	Ho et al (6)
Doelstelling	Aantonen dat OLA veiliger en/of efficiënter is dan RIS	Vergelijking tussen RIS en OLA van het percentage extrapiramidale neveneffecten en de doeltreffendheid	Vergelijking tussen RIS en OLA van de extrapiramidale neveneffecten en de doeltreffendheid
Metingen van het resultaat	PANSS, BPRS, CGI, SANS, Simpson-Angus, Barnes-Akathisia, AIMS, QoL	PANSS, ESRS, lichaamsgewicht, body mass index	BPRS, SANS, SAPS, GAS, QoL (Psych-Base), Simpson-Angus, Barnes-Akathisia
Statistische methode	Unilaterale t-test Non-inferioriteitsstudie	Bilaterale t-test Superioriteitsstudie	Bilaterale t-test Superioriteitsstudie
α -drempel	0,05	0,05	0,05
Evaluatie	Blind	Blind	Open
Rekrutering	Gerandomiseerd	Gerandomiseerd	Alternatief naturalistisch
Dosering	RIS 4-12mg/dag (7,2 \pm 2,7) OLA 10-20mg/dag (17,2 \pm 3,6)	RIS 2-6mg/dag (4,8mg) OLA 10-20mg/dag (12,4mg)	Bij ontslag uit het ziekenhuis: RIS: 5,7mg/dag OLA: 14,4mg/dag Bij follow-up: RIS: 4,5mg/dag OLA: 13,8mg/dag
Diagnose	81,7% schizofrene patiënten DSM	86,2% schizofrene patiënten DSM	100% schizofrene patiënten DSM
Leeftijd	Gemiddeld 36,21	Gemiddeld 40,0 de OLA-groep (38,9) was wat jonger dan de RIS-groep (41,0)	OLA-groep: gemiddeld 33,5 RIS-groep: gemiddeld 29,6
Geslacht	64,9% mannen	72,7% mannen	76,2% mannen
Patiënten	Ziekenhuispatiënten (100%)	Ambulante (79%) en ziekenhuispatiënten (21%)	Ziekenhuispatiënten (100%)
Ernst van de symptomen	Min. 42 op de BPRS	Initiële totale PANSS-score \geq 60 en \leq 120	Totale BPRS RIS: 46,3 OLA: 43,9
Drop-outs	42,4% in de OLA-groep 52,7% in de RIS-groep	23% in de OLA-groep 28% in de RIS-groep	
Vroegere behandeling	Niet vermeld	\pm 50% antipsychotica 15% klassieke antipsychotica	- geen neuroleptica - stopzetting vroegere behandeling - instemming met stopzetting
Conclusies	veiligheid: RIS = OLA doeltreffendheid: OLA > RIS neg. sympt. SANS/ sympt. depressie PANSS) neveneffecten: RIS > OLA gewichtstoename: OLA > RIS abowaarden: prolactine: OLA beter dan RIS transaminasen OLA hoger dan RIS, maar niet op einde studie	veiligheid: RIS = OLA doeltreffendheid: RIS > OLA op pos. sympt. en angst/depr (week 8) extrapiramidale neven- effecten: RIS = OLA gewichtstoename: OLA > RIS labowaarden: profiel levertransaminasen en serumlipiden: OLA slechter dan RIS	doeltreffendheid: RIS = OLA op neg., pos. en desorg. sympt. tijdens de acute behandeling, RIS > OLA op pos. sympt. na 6 maanden acathisie: OLA beter dan RIS prolactine: RIS slechter dan OLA

van antipsychotica, ook al gaat het om een goede en correcte studie, waarin alle vertekende variabelen werden uitgesloten of gecontroleerd. In dergelijke studies zal het uiteindelijke aantal patiënten vaak beperkt zijn in vergelijking met de bronpopulatie. Het is een combinatie van meerdere studies die de artsen het best informeert. Hoewel het concept van naturalistische studies minder goed gecontroleerd is, kunnen dergelijke studies toch nuttig zijn. De verschillende soorten studies moeten elkaar bevestigen.

Richtlijnen: belangrijkste aspecten bij het evalueren of interpreteren van een studie:

1. Selectie van de patiënten:
 - Lijken de geselecteerde patiënten op de patiënten die u in uw praktijk ziet (diagnose, leeftijd, geslacht, ambulante of ziekenhuispatiënt)?
 - Zo nee, waarin verschillen ze?
2. Dosering:
 - Is de dosering die in de studie werd gebruikt, pertinent voor uw patiëntenpopulatie?
 - Zo ja, stroken de resultaten van de studie met uw ervaring?
3. Resultaten:
 - Statistische significantie
 - Klinische significantie
4. Veralgemening
 - Waarom kunnen de resultaten niet goed worden veralgemeend?
 - Polymedicatie kan in de klinische praktijk leiden tot ervaringen die niet stroken met de studieresultaten. De meeste studies worden immers uitgevoerd in monotherapie.
 - De extrapolatie van individuele observaties wordt beperkt door variabelen zoals de leeftijd, het geslacht, het ziektestadium, middelenmisbruik...
5. Bevestiging door andere gecontroleerde en naturalistische studies
6. Adequate methode

Referenties

1. Lang TA, Secic M. How to report statistics in Medicine: Annotated guidelines for authors, editors and reviewers. Philadelphia: American College of Physicians; 1997.
2. Lieberman JA. Hypothesis and hypothesis testing in clinical trial. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):5-10.
3. Schooler NR. The statistical comparison of clinical trials. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):35-9.
4. Tran PV, Hamilton SH, Kuntz AJ, et al. Double-blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders. *Journal of Clinical Psychopharmacology* 1997;17(5):407-18.
5. Conley RR, Mahmoud R. A randomized double-blind study of risperidone and olanzapine in the treatment of schizophrenia or schizoaffective disorder. *American Journal of Psychiatry* 2001;158(5):765-74.
6. Ho BC, Miller D, Nopoulos P, Andreasen NC. A comparative effectiveness study of risperidone and olanzapine in the treatment of schizophrenia. *Journal of Clinical Psychiatry* 1999;60(10):658-63.
7. Mullen J, Reinstein M, Bari M, et al. Quetiapine and risperidone in outpatients with psychotic disorders: results of the QUEST trial. Presented at the International Congress on Schizophrenia Research Biennial Meeting. 1999.
8. Roberts, L. W. Ethics of psychiatric research: conceptual issues and empirical findings. *Comprehensive Psychiatry* 1998;39:99-110.
9. Davidson. *Communication Personelle*. 1999.
10. Procyshyn RM, Zerjav S. Drug utilization patterns and outcomes associated with in-hospital treatment with risperidone of olanzapine. *Clinical Therapeutics* 1998;20(6):1203-17.
11. Snatser M, Welch R. A retrospective, naturalistic review comparing clinical outcomes of in-hospital treatment with risperidone and olanzapine. *Clin Drug Invest* 2000;20(3):159-64.
12. Kasper S. Comments on "double blind comparison of olanzapine versus risperidone in the treatment of schizophrenia and other psychotic disorders" by Tran and Associates. *Journal of Clinical Psychopharmacology* 1998;18(4):353-4.
13. Beasley CM, Tollefson G, Tran P, Satterlee W, Sanger T, Hamilton S. Olanzapine versus placebo and haloperidol: acute phase results of the North American double-blind olanzapine trial. *Neuropsychopharmacology*. 1996a;14(2):111-23.
14. Marder SR, Davis JM, Chouinard G. The Effects of Risperidone on the Five Dimensions of Schizophrenia Derived by Factor Analysis: Combined Results of the North American Trials. *Journal of Clinical Psychiatry* 1997;58:538-46.
15. Tollefson GD, Beasley CM Jr, Tran PV, et al. Olanzapine versus haloperidol in the treatment of schizophrenia and schizoaffective and schizophreniform disorders: results of an international collaborative trial. *American Journal of Psychiatry* 1997;154:457-65.
16. Peuskens J, Link CGG. A comparison of quetiapine and chlorpromazine in the treatment of schizophrenia. *Acta Psychiatrica Scandinavica* 1997;96:265-73.
17. Licht RW, Gouliaev G, Vestergaard P, et al. Generalizability of results from randomized drug trials: a trial on antimanic treatment. *British Journal of Psychiatry* 1997;170:264-7.
18. Angst J, Bech P, Bobon D, et al. Report on the Third Consensus Conference on the Methodology of Clinical Trials With Antipsychotic Drugs. *Pharmacopsychiatry* 1991;24:149-52.
19. Lindamer LA, Lohr JB, Harris MJ, et al. Gender-related clinical differences in older patients with schizophrenia. *Journal of Clinical Psychiatry* 1999;60:61-7.
20. Kirch DG, Keith SJ, Matthews S. M. Research on first episode psychosis: report on a National Institute of Mental Health Workshop. *Schizophrenia Bulletin*. 1992; 18:179-84.
21. Kane JM. Dose selection and comparator drugs in schizophrenia research. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):29-34.
22. Conley RR. Evaluating Clinical trial data: outcome measures. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):23-8.
23. Kay SR, Opler LA, Lindenmayer JP. The positive and negative syndrome scale (PANSS): rationale and standardisation. *British Journal of Psychiatry* 1989;155(7):59.
24. Bell M, Milstein R, Beam-Goulet J, Lysaker P, Cicchetti D. The positive and negative syndrome scale and the brief psychiatric rating scale. *Journal of Nervous and Mental Disease*. 1992;180(11):723-8.
25. De Hert M. Psychose schizophrénique et abus de substances. Dom G, editor. *Troubles psychiques et accoutumance. Comment les traiter ensemble?* EPO 2000. p. 221-44.
26. Docherty JP. Interpretations and conclusions in the clinical trial. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):40-5.
27. Casey DE. Clinical trial design issues in schizophrenic research. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):17-22.
28. Rogers D. Comparing atypical antipsychotics. *Psychiatric Times* 2002;19(1).
29. Gheuens JA, Grebb J. Comments on article by Tran and associates, 'Double-blind comparison of olanzapine versus risperidone in treatment of schizophrenia and other psychotic disorders'. *Journal of Clinical Psychopharmacology* 1998;18:176-9.
30. Keith SJ. Evaluating Characteristics of patient selection and dropout rates. *Journal of Clinical Psychiatry* 2001;62(Suppl.9):11-6.
31. Tran PV, Dellva MA, Tollefson GD, Wentley AL, Beasley CM. Oral olanzapine versus oral haloperidol in the maintenance treatment of schizophrenia and related psychosis. *British Journal of Psychiatry* 1998;172:499-505.
32. Schooler N. Comments on Article by Tran and Colleagues, "double blind comparison of olanzapine versus risperidone in treatment of schizophrenia and other psychotic disorders". *Journal of Clinical Psychopharmacology* 1998;18(2):174-5.
33. Leucht S, Pitschel-Walz G, Abraham D, Kissling W. Efficacy and extrapyramidal side-effects of the new antipsychotics olanzapine, quetiapine, risperidone, and sertindole compared to conventional antipsychotics and placebo. A meta-analysis of randomized controlled trials. *Schizophrenia Research* 1999;35(1):51-68.
34. Kasper S, Rosillon D, Duchesne I. Risperidone olanzapine drug outcomes studies in schizophrenia (RODOS): efficacy and tolerability results of an international naturalistic study. *International Clinical Psychopharmacology* 2001;16:179-87.

The work of the Belgian Discussion Board on Efficacy of Antipsychotic Treatment is supported by an unrestricted educational grant from Janssen-Cilag & Organon.