

Een betrouwbaarheidsonderzoek van de ICD-9-CM codering

door J. Dutilh en P. M. A. J. Dingemans

Inleiding

In Nederland worden in het merendeel van alle intramuraal behandelde psychiatrische patiënten gegevens verzameld door het Centrale Patiënten Register Psychiatrische Instituten. De diagnostische gegevens worden volgens de negende revisie van de International Classification of Diseases (ICD-9-CM) geclassificeerd en gecodeerd. De combinatie van demografische gegevens, opnameduur, behandelverloop en diagnostische gegevens biedt een unieke mogelijkheid tot wetenschappelijk onderzoek waarvan onder andere door Giel en zijn medewerkers sinds 1974 gebruik gemaakt wordt. Zij publiceerden sinds 1974 in vele afleveringen de 'Mededelingen uit het Centrale Patiënten Register' (Van Weerden-Dijkstra en Giel, 1974).

In de psychiatrische kliniek van het Academisch Ziekenhuis bij de Universiteit van Amsterdam (AZUA) wordt sinds 1970 volgens een van het CHAM-systeem (Silbermann, 1971) afgeleid systeem geclassificeerd en gecodeerd volgens de International Classification of Diseases. Aanvankelijk codeerde één van de zenuwartsen, later werd dit door codeurs overgenomen.

Tien jaar later, in 1980, nam de belangstelling voor diagnostische codering opnieuw toe, toen er door de introductie van DSM-III een beter classificatie-systeem ter beschikking leek te zijn gekomen. Om met beide systemen vertrouwd te raken werden zowel DSM-III als de ICD-9-CM in het onderwijs opgenomen. Begin 1981 namen de assistent-psychiaters het coderen voor de 'blauwe kaart' van het Centrale Patiënten Register over van de codeurs in dit ziekenhuis. Hiervoor gold als argument dat artsen, omdat ze meer gegevens tot hun beschikking hebben, tot een betere ICD-codering in staat zouden zijn. Hiernaar

Schrijvers zijn psychiater, respectievelijk psycholoog in de kliniek voor volwassenenpsychiatrie van het Academisch Ziekenhuis bij de Universiteit van Amsterdam (AZUA).

wordt onderzoek verricht (Dingemans, Hooijer en Dutilh, nog te publiceren).

De International Classification of Diseases werd samengesteld onder auspiciën van de World Health Organization. Bij het maken van dit systeem werd beoogd dat het in zoveel mogelijk landen geaccepteerd zou moeten kunnen worden, ook al zouden deze zeer verschillende diagnostische tradities kunnen hebben. Over het bereikte compromis was men in de Verenigde Staten in elk geval weinig tevreden zodat er een aangepaste versie verscheen: ICD-9-Clinical Manual. Van deze versie is een Nederlandse vertaling verschenen welke in onze kliniek gebruikt wordt (Classificatie van Ziekten, 1980).

Het psychiatrische deel van dit systeem bestaat uit een lijst van ongeveer driehonderd syndroombenamingen welke zijn ingedeeld in 29 klassen. Er is een verklarende woordenlijst toegevoegd, waarin voor een aantal, maar niet voor alle ziektebeelden of termen, een karakterisering gegeven wordt.

Kritiek op de ICD-9 – Op dit systeem is veel kritiek gebleken. Zo schreef Lit in 1981 dat ‘... het met de ICD-9 onmogelijk is de ontstaansvoorwaarden in diagnosen te rubriceren, en vervolgens te coderen. Een ernstig feit als men beseft dat de patiëntenregistratie onder meer is opgezet om een inzicht te krijgen in de ontstaansvoorwaarden van psychiatrisch dysfunctioneren’. Slooff e.a. (1982) schreven onlangs dat ‘... het classificatie-principe (van ICD-9) op onsystematische wijze berust op fenomenologische, beloops- en aetiologische criteria. Beslissingsregels voor de hantering en de weging van deze criteria ontbreken. Dit heeft tot gevolg dat de resultaten met dit systeem erg afhankelijk worden van de individuele interpretatie van de diagnosticus. De betrouwbaarheid is dan ook matig’. Wij willen door middel van een vijftal voorbeelden laten zien waar deze kritiek op berust.

A. Een depressie kan door middel van verschillende categorieën geclassificeerd worden. De klasse van affectieve psychosen (ICD: 296) bevat de categorie ‘Ernstig Depressieve Stoornis, Enkele episode’ (ICD: 296.2). Deze categorie is niet te onderscheiden van de ‘Depressieve Vorm van Psychose’ (ICD: 298.0) welke tot de klasse ‘Overige, niet-Organische Psychosen’ (ICD: 298) behoort.

B. Het is eveneens onduidelijk waar precies het onderscheid gemaakt moet worden tussen ‘Neurotische Depressie’ (ICD: 330.4) en ‘Ernstig Depressieve Stoornis’ (ICD: 296.2). Onder deze laatste categorie worden onder meer als voorbeeld genoemd: ‘Endogene Depressie’, ‘Psychotische Depressie’, ‘Unipolaire Depressie’. In deze categorie kan het dus om een psychotische patiënt gaan, maar het hoeft niet.

C. Het classificeren van paranoïde stoornissen geeft een probleem doordat voor de beloopduur van het ziektebeeld geen beslissingscriteria worden gegeven om te onderscheiden tussen: ‘Paranoïde vorm van Schizofrenie’ (ICD: 295.3), ‘Paranoïde Toestanden’ (ICD: 297.3) en

PSYCHOFARMACA

'Acute Paranoïde Reactie' (ICD: 298.3).

D. Verschillende principes worden voor het classificeren door elkaar heen gebruikt. Dit wordt duidelijk door het volgende voorbeeld van vijf opvolgende diagnostische klassen. 'Drugpsychosen' (ICD: 292): hier geldt een aetiologisch criterium. 'Passagère Organisch Psychotische Stoornissen' (ICD: 293), 'Overige Organische Psychosen, Chronisch' (ICD: 294): bij deze twee klassen geldt de beloopsvorm en de aetiologie als criterium. 'Schizofrene Psychosen' (ICD: 295), 'Affectieve Psychosen' (ICD: 296): bij deze twee diagnoses gaat het niet om aetiologische criteria (alhoewel hier misschien nog meningsverschil over kan bestaan) maar om descriptieve en beloopfactoren.

E. Paradoxaal is de diagnose in het volgende voorbeeld. Bij de 'Affectieve Psychosen' (ICD: 296) kunnen een aantal subklassen gebruikt worden. Deze worden uitgedrukt door middel van een vierde cijfer. Hiernaast moet nog de mate van ernst van het ziektebeeld worden aangeduid door middel van een vijfde cijfer. Als we subklasse 2 kiezen is dit: 'Ernstig Depressieve Stoornis Enkele Episode'. Als we als mate van ernst de 5 kiezen is dit: 'Ernstig, Zonder vermelding van psychotisch gedrag'. Als we nu de combinatie van al deze mogelijkheden aan elkaar verbinden wordt de diagnose (ICD: 296.25): 'Affectieve Psychose, Ernstig Depressieve Stoornis, Enkele Episode, Ernstig, Zonder vermelding van psychotisch gedrag'.

Voor het epidemiologisch onderzoek dat kan worden uitgevoerd met alle gegevens die door het Centrale Patiënten Register verzameld worden is het van belang dat er zo betrouwbaar mogelijke gegevens worden ingestuurd. Onderzoek naar de diagnostische betrouwbaarheid met ICD-9 is in Nederland tweemaal gedaan. Gielen Bogaarts publiceerden in 1980 een onderzoek naar de betrouwbaarheid van het coderen van diagnostische omschrijvingen. Een situatie waarvoor de beroepscoders zich gesteld zien. 30 codeurs, 17 artsen, 11 administratieve medewerkers, 1 psycholoog en 1 verpleger werkten aan dit onderzoek mee. De overeenstemming bleek zeer verschillend per diagnostische categorie: 27 tot 97%. Uit de in dit onderzoek vermelde gegevens berekenden wij de gemiddelde overeenstemming: 53%. Door Giel e.a. (1981) werd een internationaal betrouwbaarheidsonderzoek verricht met de ICD-8, in het kader van de WHO-Collaborative Study on Strategies for Extending Mental Health Care.

26 beoordelaars beoordeelden in dit onderzoek 10 gevals vignetten. De diagnostische overeenstemming per gevalsbeschrijving varieerde tussen 46 en 92% met een door ons berekend gemiddelde van 71%.

Tegelijk met de introductie van DSM-III, werden twee betrouwbaarheidsonderzoeken met dit classificatie-systeem gepubliceerd (Spitzer, 1979; DSM-III, 1980, Appendix F, pag. 467). Beide onderzoeken zijn zogenaamde 'Field Trials'. Het eerste onderzoek werd met een voorlopige versie van de DSM-III verricht, het tweede met de definitieve

versie. In deze onderzoeken zagen daartoe aangezochte psychiaters al of niet gelijktijdig dezelfde patiënt, over wie zij hun diagnostische bevindingen onafhankelijk van elkaar dienden vast te leggen. Deze beoordelaars konden dus niet van gestandaardiseerde gegevens uitgaan, toch bleek de interbeoordelaarsbetrouwbaarheid redelijk: .72 respectievelijk .72 in beide onderzoeken. De betrouwbaarheid werd volgens een methode met correctie voor kans op overeenstemming berekend, zodat de resultaten niet vergelijkbaar zijn met de eerder genoemde onderzoeken (Giel en Bogaarts, 1980; Giel e.a., 1981).

Door de ervaring in onze kliniek, van steeds weerkerende verwarring rond het coderen met de ICD-9-CM, zijn wij gaan twijfelen aan het mogelijk zijn van een redelijke betrouwbaarheid. Om de waarde te toetsen van de gegevens welke vanuit deze kliniek worden ingestuurd, hebben wij een betrouwbaarheidsonderzoek gedaan.

Methode

Aan zeven artsen, zes assistent-psychiaters en één psychiater, werd gevraagd hun oordeel te geven over dezelfde twintig gestandaardiseerde ontslagbrieven. Deze brieven werden at random gekozen uit de groep van patiënten die in 1981 in onze kliniek werden opgenomen en ontslagen. De brieven zijn volgens een standaardprotocol opgesteld zodat ze allemaal een aantal vaste gegevens bevatten: reden van opname, klinische toestand bij opname volgens het CHAM-systeem (Silbermann, 1971), psychiatrisch onderzoek, lichamelijk onderzoek, biografische anamnese, hetero-anamnese, verloop van de behandeling met de gebruikte behandeltechnieken en aetiologische gegevens.

De beoordelaars hadden elk minstens drie maanden ervaring met de ICD-9-CM. Als hulpmiddel werd door hen de Nederlandse vertalingen de toegevoegde woordenlijst van de ICD-9-CM gebruikt. Hen werd gevraagd om net als in de dagelijkse praktijk de diagnostische categorieën en codenummers te kiezen, zoals de 'blauwe kaart' voorschrijft: een hoofddiagnose met al of niet één of twee bijdiagnosen.

Om vergelijking met ander onderzoek (Giel e.a., 1980; Giel e.a., 1981; Spitzer e.a., 1979; DSM-III, 1980, Appendix F, pag. 467) mogelijk te maken werd de diagnostische overeenstemming gezien als overeenstemming binnen een diagnostische klasse, zodat aangenomen kan worden dat de feitelijke overeenstemming veel lager is. Zo werden de categorie 'Bipolaire Affectieve Stoornis, Depressief' (ICD: 296.5) en de categorie 'Ernstig Depressieve Stoornis, Enkele Episode' (ICD: 296.2) gezien als overeenstemmend omdat ze zich beide bevinden in de klasse van 'Affectieve Psychosen' (ICD: 296).

De overeenstemming werd berekend volgens twee methoden, eveneens om vergelijking met andere onderzoeken mogelijk te maken.

De procenten-methode – Deze methode berekent het aantal beoordelaars dat het eens is als verhouding van het totaal aantal beoordelaars. Dit getal wordt dan in procenten uitgedrukt. Deze methode is eenvoudig en wordt ook veel gebruikt. Een nadeel is echter dat er niet gecorrigeerd wordt voor de kans op overeenstemming (Spitzer, 1967) zodat de uitkomst een groter getal oplevert dan wanneer dat wel wordt gedaan. Een tweede nadeel is dat meervoudige diagnoses niet vergeleken kunnen worden. De codeurs moeten aangeven wat hun hoofddiagnose is. Deze worden dan vergeleken.

De Kappa-methode – Bij veel onderzoekers is de methode volgens Cohen (1960) in gebruik geraakt. De coëfficiënt voor diagnostische overeenstemming is:

$$\text{Kappa} = \frac{P_O - P_c}{1 - P_c}$$

P_O is de geobserveerde overeenstemming, P_c is de kans op overeenstemming. De coëfficiënt voor diagnostische overeenstemming is de geobserveerde overeenstemming verminderd met de kans op overeenstemming, dit in verhouding tot de maximaal mogelijke overeenstemming (=1) verminderd met de kans op overeenstemming.

Deze methode is bruikbaar in het geval twee beoordelaars elk één diagnostische categorie kiezen. Door Mezzich e.a. (1981) werden drie procedures ontwikkeld voor de situatie, dat meer beoordelaars voor elke diagnose méér dan één categorie kiezen. Omdat deze volgens ons het meest elegant is kozen we de proportionele overlap methode. De volgorde van de diagnostische categorieën, dus de hiërarchie tussen hoofd- en bijdiagnosen, telt hier niet mee.

Volgens deze procedure worden alle mogelijke paren van diagnoses van één patiënt vergeleken. Dus als er vier beoordelaars zijn: A, B, C, D, zijn alle mogelijke paren: AB, AC, AD, BC, BD en CD. Van elk diagnosepaar worden het aantal overeenstemmende categorieën geteld en in verhouding gebracht met het totaal aantal gebruikte categorieën. Dit getal heet de proportionele overlap: P_O . Voorbeeld:

A heeft gekozen: a, b, c;

B heeft gekozen: b, c, d, en

C heeft gekozen: d, e.

Paar AB heeft nu twee overeenstemmende categorieën, b en c, ten opzichte van in totaal vier gebruikte categorieën: a, b, c en d. Hieruit volgt dat $P_O 2/4$ is, = 0,5. Paar BC heeft één overeenstemming over vier gebruikte categorieën: $P_O = 0,25$. Paar AC heeft geen overeenstemming: $P_O = 0$. Per beoordeelde patiënt wordt de gemiddelde proportionele overlap berekend (in dit voorbeeld: $P_O \text{ gem.} = 0,25$) en eveneens van alle in het onderzoek betrokken patiënten, zodat Kappa wordt:

$$\text{Kappa} = \frac{\frac{\sum P_O}{N} - P_c}{1 - P_c}$$

N is het aantal diagnoseparen. P_c wordt berekend volgens dezelfde procedure als de proportionele overlap, echter met het verschil dat alle mogelijke paren van alle in het onderzoek gestelde diagnoses, van alle diagnostici met elkaar vergeleken worden. Het gemiddelde van al deze getallen is P_c .

Resultaten

Er werden door zeven beoordelaars ontslagbrieven van in totaal 20 patiënten beoordeeld, waarvan wel de ICD-9-CM code en/of DSM-III code verwijderd was, echter niet de klinische diagnose volgens het systeem dat in onze kliniek gebruikt wordt. In 71 gevallen werd één diagnostische categorie gekozen, 51 maal twee categorieën en 18 maal drie categorieën.

ICD-9-CM bestaat uit 29 diagnose klassen welke door de drie eerste cijfers van de code worden weergegeven. Vier klassen betreffen aandoeningen op kinderleeftijd. Deze konden voor dit onderzoek dus niet gebruikt worden. Van de 25 overige klassen werden er 17 wel gebruikt. Volgens de procenten-methode is de overeenstemming tussen de zeven beoordelaars 71%. Volgens de Kappa-methode is de overeenstemming: $\text{Kappa} = .39$, met een t-waarde van 1.2; bij 19 vrijheidsgraden geeft dit een p-waarde van $> .20$. Hieruit blijkt al dat de correctie voor kans tot een aanmerkelijk lager getal leidt. In dit onderzoek is de betrouwbaarheid, gecorrigeerd voor kans op overeenkomst, dus veel te laag. Slechts van twee patiënten leidde het classificeren tot een betrouwbaar resultaat. Bij beiden was men het over de neurotische stoornis eens (ICD: 300): de waarden van de proportionele overlap waren respectievelijk .81 en .86. De 18 overige patiënten werden met een wisselende mate van onbetrouwbaarheid beoordeeld: een proportionele overlap tussen .18 en .65 (tabel 1).

Vergelijking met ander betrouwbaarheidsonderzoek – De diagnostische betrouwbaarheid, berekend volgens de procenten-methode kan met twee andere onderzoeken vergeleken worden (tabel 2). Giel en Bogaarts (1980) gaven hun beoordelaars niet méér dan diagnostische omschrijvingen, omdat in een aantal ziekenhuizen codeurs deze omschrijvingen in een ICD-code hebben om te zetten. Zij vonden een overeenstemming van 53%. Dit is lager dan de 71%, welke werd gevonden door Giel e.a. (1981), waarbij de beoordelaars gevals vignetten codeerden. Treffend is de overeenkomst met de procentuele betrouwbaarheid berekend in ons onderzoek, eveneens 71%. Het lijkt of de beschikking over méér patiëntengegevens tot een grotere diagnosti-

Tabel 1: Diagnostische overeenstemming van zeven beoordelaars van twintig ontslagbrieven

Diagnostische gegevens van de ontslagbrief	ICD-9-CM-codes		% overeenstemming	Proportionele Overlap, P _o
	hoofd-diagnose	Bijdiagnosen in volgorde van frequentie		
Maniform syndroom bij zwakbegaafdheid en epilepsie	296	295, 294, 298, 317	43	.19
Paranoïsch-hallucinatoir syndroom, involutieproblemen, paranoïde persoonlijkheidsstoornis	297	295, 301, 298, 317	43	.21
Recidief paranoïsch syndroom	297	295, 298, 303	43	.22
Paranoïsch syndroom, chronische psychose	297	295, 296	57	.33
Depressief syndroom, anorexia nervosa	300	307, 306, 296	71	.38
Amentia, hysterische karakterneurose	298	300, 296, 301	71	.38
Geagiteerd depressief syndroom, hysterische karakterneurose	300	301, 309	71	.39
Gedragstoornissen, geagiteerd depressief syndr., zwakbegaafdheid	300	317, 309, 301	43	.40
Melancholisch syndroom, dementie	296	290, 300	100	.45
Paranoïsch-hallucinatoir syndroom, chronische psychose	295	297, 298, 302, 301	71	.46
Anorexia nervosa, Dulcolax-verslaving	307	305, 304, 295, 300	71	.46
Depressief syndroom, psychasthenie	300	304, 301, 296	86	.48
Paranoïsch-hallucinatoir syndroom, karakterneurose	295	298, 301, 309	86	.49
Stupor, depressief syndroom, dementie	290	295, 298, 300, 311	71	.54
Agitatie syndroom, cyclische psychose	296	295, 300	71	.60
Paranoïsch-hallucinatoir syndroom, chronische psychose	295	297, 300	71	.61
Amentieel syndroom, passief afhankelijk karakter	298	295, 301	71	.61
Depressief syndroom, fobie, hysterische karakterneurose	300	309, 301	100	.65
Depressief syndroom, hysterische karakterneurose	300	309, 301	86	.81
Angstsyndroom, hysterische karakterneurose	300	301	100	.86

Methode	Giel en Bogaarts 1980	Giel e.a. 1981	Dit onderzoek	Spitzer 1979	DSM, appendix F 1980
	diagnostische omschrijvingen	gevals-vignetten	ontslag-brieven	interviews met patiënten	
Classificatie-systeem	ICD-9	ICD-8	ICD-9-CM	DSM-III-draft	DSM-III
Aantal patiënten	10	10	20	281	670
Aantal beoordelaars	30	26	7	Twee per patiënt	Twee per patiënt
Gemiddelde overeenstemming	53%	71%	% Kappa 71% .39	Kappa .72	Kappa .72
<i>Enkele diagnose klassen</i>					
syndromen (ICD 290)	97%	—	71% .39	.79	.76
Schizofrene syndromen (ICD 295)	—	76%	76% .37	.82	.81
Paranoïde syndromen (ICD 297)	66%	—	48% .19	—	.75
Affectieve psychoses (ICD 296)					
Major affective disorders-DSM-III	60%	70%	71% .26	.68	.80
Neurotische depressie (ICD 300.4)	—	85%	76% .37	.47	.69

sche betrouwbaarheid leidt. Hierover zal in een vervolgonderzoek bericht worden (Dingemans, Hooijer en Dutilh, 1983).

Gebruik van de Kappa-methode biedt de gelegenheid te vergelijken met Amerikaans betrouwbaarheidsonderzoek met de DSM-III. In beide 'Field Trials' (Spitzer 1979; DSM-III, 1980, Appendix F, pag. 467) werden patiënten gezien. Er werd dus niet van gestandaardiseerde gegevens uitgegaan. Deze factor zou tot meer onbetrouwbaarheid tussen beoordelaars kunnen leiden. Tegenover de Kappa, in dit onderzoek .37, is de gemeten interbeoordelaars-betrouwbaarheid van de DSM-III overtuigend beter: .72 respectievelijk .72 (tabel 2). Spitzer e.a. (1979) menen dat een aantal innovaties in de DSM-III hieraan hebben bijgedragen. Veranderingen in de classificatie zelf (zoals het hergroeperen van ziektebeelden), de scheiding van ziektebeelden en persoonlijkheidsstoornissen op verschillende assen, en de invoering van diagnostische criteria. Als methodologisch bezwaar kan echter aangevoerd worden dat er waarschijnlijk geen goede controle was op het uitwisselen van ideeën tussen de beoordelaars.

Naast de vergelijking van de betrouwbaarheid in vijf onderzoeken toont de tabel de betrouwbaarheid opgesplitst voor enkele diagnostische klassen, welke in de meeste onderzoeken voorkwamen. De procentuele overeenstemming is voor drie diagnose klassen, tussen Giels onderzoek uit 1981 en ons onderzoek opvallend overeenkomstig. Des te meer is opvallend dat de Kappa's in de 'Field Trials' met de DSM-III zoveel hoger zijn dan in dit onderzoek.

Conclusies en aanbevelingen

(1) Wij menen dat de diagnostische codes uit onze kliniek voor epidemiologisch onderzoek een te geringe waarde hebben om er wetenschappelijk onderzoek mee te kunnen doen.

(2) De procenten-methode suggereert een redelijke overeenstemming, daar waar na correctie op kans een veel te lage diagnostische overeenstemming blijkt te zijn. Omdat de procedure met de procenten-methode in ons onderzoek identiek is met die van Giels beide onderzoeken, vermoeden we dat ook die onderzoeken eenzelfde lage Kappa-coëfficiënt zouden hebben opgeleverd, als die berekend was. De suggestie ontstaat dan dat met de ICD-9 überhaupt geen betrouwbaarheid bereikt kan worden.

(3) Opvallend is het verschil met de betrouwbaarheid welke met de DSM-III door matig geoefende gebruikers bereikt werd. Als goed opgezet betrouwbaarheidsonderzoek met de DSM-III ook in ons land tot betere resultaten leidt dan met de ICD-9, zou de ICD-9 voor de Centrale Patiënten Registratie beter kunnen worden vervangen door de DSM-III.

Samenvatting

Gesignaleerd wordt dat de ICD-9-CM als classificatiesysteem voor psychiatrische diagnoses veel tekortkomingen heeft. Ter berekening van de diagnostische betrouwbaarheid van ICD-9-CM wordt een eenvoudig uitvoerbare methode met kanscorrectie beschreven en toegepast. Zeven artsen kozen de diagnostische categorieën bij twintig gestandaardiseerde ontslagbrieven. De Kappa-coëfficiënt voor overeenstemming is .37. Na vergelijking met eerder verricht onderzoek wordt geconcludeerd dat de suggestie ontstaat dat met ICD-9 geen aanvaardbare diagnostische overeenstemming wordt bereikt. Aanbevolen wordt voor de landelijke registratie van psychiatrische ziektes op de DSM-III over te gaan indien in ons land verricht onderzoek bevestigt, dat dit systeem tot een betere betrouwbaarheid aanleiding geeft.

Literatuur

- American Psychiatric Association, *Diagnostic and statistical manual of mental disorders*, third edition (DSM-III), Washington, 1980.
- Classificatie van Ziekten*, Centrum voor Informatieverwerking voor de Nederlandse Ziekenhuizen, Stichting Medische Registratie, Utrecht (1980).
- Cohen, J. (1960) A coefficient of agreement for nominal scales. *Educational and psychological measurement* 20, 37-46.
- Dingemans, P. M. A. J., C. Hooijer en J. Dutilh (1983) *ICD-codering in de psychiatrie*. In bewerking.
- Giel, R. en F. Bogaarts (1980) Diagnostische classificatie en codering in de psychiatrie. *T. v. Psychiatrie* 22, 327-335.
- Giel, R., E. d'Arrigo Busnello, C. E. Climent, A. S. E. D. Elhakim, H. H. A. Ibrahim, L. Ladrigo-Ignacio en N. N. Wig (1981) The Classification of Psychiatric Disorders, A Reliability Study in the WHO-Collaborative Study on Strategies for Extending Mental Health Care. *Acta Psychiat. Scand.* (1981) 63, 61-74.
- Lit, A. C. (1981) Over het classificeren in de psychiatrie. *T. v. Psychiatrie* 23, 5-20.
- Mezzich, J. E., H. C. Kraemer, D. R. L. Worthington en G. A. Coffman (1981) Assessment of Agreement among several Raters formulating multiple diagnoses. *J. Psychiat. Res.* 16, 29-39.
- Silbermann, R. M. (1971) *CHAM; a classification of psychiatric states*. Excerpta Medica, Amsterdam.
- Slooff, C. J., R. Giel, A. de Jong en D. Wiersma (1982) De classificatie van de functionele niet-affectieve psychosen. *T. v. Psychiatrie* 24, 454-469.
- Spitzer, R. L., J. Cohen, J. L. Fleiss en J. Endicott (1967) Quantification of Agreement in Psychiatric diagnosis, a new Approach. *Arch. Gen. Psychiat.* 17, 83-87.
- Spitzer, R. L., J. B. W. Forman en J. Nee (1979) DSM-III Field Trials: I. Initial Interrater Diagnostic Reliability. *Am. J. Psychiat.* 136, 815-817.
- Weerden-Dijkstra, J. R. van en R. Giel (1974) Mededelingen uit het centrale patiënten-register voor psychiatrische ziekenhuizen en universiteitsklinieken. *T. v. Psychiatrie* 16, 453-461.