KM 18(1985)
pag 27 - 40

BEPALING VAN SENSITIVITEIT EN SPECIFICITEIT VAN DIAGNOSTISCHE TEST-PROCEDURES ZONDER "GOUDEN STANDAARD".

G.A.M.Eilers(*)
R.van Strik(*)

SAMENVATTING

Bepaling van sensitiviteit en specificiteit van een diagnostische test uit empirische gegevens is in principe alleen mogelijk indien de correcte diagnose met 100% zekerheid bekend is ("gouden standaard"). In dit artikel wordt de methode van Hui en Walter (1980), waarmee sensitiviteit en specificiteit van twee verschillende diagnostische testen kunnen worden geschat in het geval er geen "gouden standaard" voorhanden is, nader bekeken. De eerste voorwaarde voor het toepassen van deze methode is dat het waarnemingsmateriaal is te splitsen in twee subgroepen waarin de prevalenties duidelijk verschillend zijn.
Daarnaast moet aan nog twee voorwaarden worden voldaan, welke niet aan de resultaten kunnen worden gecontroleerd. De methode van Hui en Walter is beoordeeld op de gevoeligheid voor afwijkingen van deze drie voorwaarden / aannamen en toegepast op een praktijkvoorbeeld.

(*) Instituut voor Biostatistica
Erasmus Universiteit Rotterdam
Postbus 1738, 3000 DR Rotterdam
Tel. 010 - 634136 / 4149

1. INLEIDING

In de praktijk van het medisch handelen wordt vaak op grond van één of meer testen een diagnose gesteld. Verschillende testen zullen echter niet steeds tot dezelfde diagnose leiden. In het geval er ook geen "absoluut zekere test" ofwel "gouden standaard" voorhanden is, rijst er een probleem.

Stel dat bij 1000 personen twee testen, 1 en 2, zijn verricht met de onderstaande resultaten:

		TEST 2 +	TEST 2 −	
TEST 1	+	223	117	340
	−	87	573	660
		310	690	1000

De wijze waarop dergelijke gegevens over twee testen doorgaans worden vertaald in termen van sensitiviteit en specificiteit is de volgende (tussen haakjes zijn steeds schattingen voor de standaardfout gegeven):

Als TEST 1 wordt beschouwd als de "gouden standaard", dan geldt:

prevalentie = 340/1000 = .34 (.015)

en voor TEST 2: sensitiviteit = 223/340 = .66 (.026)

specificiteit = 573/660 = .87 (.013)

Als TEST 2 wordt beschouwd als de "gouden standaard", dan geldt:

prevalentie = 310/1000 = .31 (.015)

en voor TEST 1: sensitiviteit = 223/310 = .72 (.026)

specificiteit = 573/690 = .83 (.014)

De werkelijkheid zal wel ergens in het midden liggen, zodat de hierboven gevonden waarden voor de sensitiviteit en specificiteit kunnen worden beschouwd als benedengrenzen voor de werkelijke waarden van deze grootheden. De beide waarden voor de prevalentie kunnen daarbij worden gezien als grenzen voor de werkelijke prevalentie in de totale groep van personen.

In Staquet et al.(1981) worden drie situaties besproken waarin van tenminste één van beide testen de sensitiviteit en/of specificiteit bekend is (1. referentietest met bekende sensitiviteit en specificiteit, 2. referentietest en nieuwe test beiden specificiteit 100%, 3. referentietest met specificiteit 100%). Hui en Walter (1980) beschrijven een methode zonder deze restricties waarmee de sensitiviteit en specificiteit van twee verschillende testen kan worden geschat, mits de totale groep waarop de twee testen zijn toegepast, kan worden opgesplitst in twee subgroepen waarin de prevalenties duidelijk verschillen. Daarbij wordt allereerst verondersteld dat de twee testen

onafhankelijk fouten maken zowel binnen de groep werkelijk gezonden als de werkelijk zieken. Op de tweede plaats moet voldaan worden aan de voorwaarde dat de sensitiviteit en specificiteit voor beide testen konstant is in de twee subgroepen. Naast de schattingen voor de sensitiviteit en specificiteit van de beide testen geeft deze methode tevens schattingen voor de prevalentie in de twee verschillende subgroepen en (asymptotische) schattingen voor de bijbehorende standaardfouten. Met behulp hiervan kunnen we de twee testen vergelijken en verschillen ertussen toetsen.

Voor het toepassen van de methode van Hui en Walter is een computerprogramma geschreven in Commodore-basic. De uitvoer met dit programma voor het voorbeeld in het artikel van Hui en Walter (1980) is gegeven in Appendix A1. De methode is toegepast op gesimuleerde gegevens, waarbij de robuustheid voor bepaalde aannamen is onderzocht. Daarnaast is deze methode toegepast op enkele praktijksituaties. Op grond van onze bevindingen zouden wij gaarne over meer praktijkgegevens beschikken alvorens tot algemene propaganda van deze methode over te gaan.

2. CONSTRUCTIE VAN GESIMULEERDE DATA EN TERUGSCHATTING

Stel: er zijn twee diagnostische testen waarvan het volgende is gegeven:

TEST 1: sensitiviteit = .9 TEST 2: sensitiviteit = .8
specificiteit = .9 specificiteit = .9

en verder twee verschillende subgroepen, waarvoor geldt:

SUBGROEP 1 : 500 personen, prevalentie = .2
SUBGROEP 2 : 500 personen, prevalentie = .4

Op grond van deze gegevens en onder de aanname dat de twee testen geheel onafhankelijk van elkaar fouten maken binnen de twee subgroepen kunnen we hieruit het volgende berekenen:

SUBGROEP 1: prev. = .2, dus 500 personen = 100 zieken + 400 gezonden

100 zieken + 400 gezonden = 500 personen

T1 \ T2	+	−	
+	72	18	90
−	8	2	10
	80	20	100

+

T1 \ T2	+	−	
+	4	36	40
−	36	324	360
	40	360	400

=

T1 \ T2	+	−	
+	76	54	130
−	44	326	370
	120	380	500

SUBGROEP 2: prev.= .4, dus 500 personen = 200 zieken + 300 gezonden.

200 zieken + 300 gezonden = 500 personen

T1 \ T2	+	−	
+	144	36	180
−	16	4	20
	160	40	200

+

T1 \ T2	+	−	
+	3	27	30
−	27	243	270
	30	270	300

=

T1 \ T2	+	−	
+	147	63	210
−	43	247	290
	190	310	500

Zodat voor de totale groep van 1000 personen geldt:

T1 \ T2	+	−	
+	223	117	340
−	87	573	670
	310	690	1000

Dit zijn precies de aantallen uit het voorbeeld in de inleiding.
De resultaten van de methode Hui en Walter geven de ingevoerde gegevens exact terug (zie appendix A2), namelijk:

SUBGR 1: prevalentie = .200 (.044)
SUBGR 2: prevalentie = .400 (.057)

TEST 1: sensitiviteit = .900 (.080)
specificiteit = .900 (.037)

TEST 2: sensitiviteit = .800 (.076)
specificiteit = .900 (.032)

De schattingen in paragraaf 1 zijn dus inderdaad benedengrenzen voor de werkelijke waarden. Opvallend in de correlatiematrix van de schattingen (zie appendix A2) is de hoge negatieve correlatie tussen de sensitiviteit van de ene test en de specificiteit van de andere, terwijl de sensitiviteit en specificiteit van dezelfde test nauwelijks gecorreleerd zijn. Het vergelijken van (eventueel toetsen van het verschil tussen) de sensitiviteiten resp. specificiteiten van de twee testen verloopt door middel van betrouwbaarheidsintervallen voor het verschil.
Een puntschatting voor het verschil (sens1-sens2) is in dit geval .9−.8=.1 en de schatting voor de variantie hiervan is:

$$\begin{aligned} VAR(sens1-sens2) &= VAR(sens1) + VAR(sens2) - 2*COV(sens1,sens2) \\ &= .00646 + .00573 - 2*.00091 \\ &= .01037 \end{aligned}$$

zodat een schatting voor de standaardfout in (sens1-sens2) bij benadering gelijk is aan $\sqrt{.01037} = .10$ en het 95% betrouwbaarheidsinterval voor het verschil (sens1-sens2) bij benadering:

95% B.I. voor (sens1-sens2): $.10 \pm 2*.10$, ofwel (−.10,+.30)

Analoog volgt voor het verschil (spec1-spec2) in dit voorbeeld:

95% B.I. voor (spec1-spec2): $0 \pm 2*.045$, ofwel (−.09,+.09)

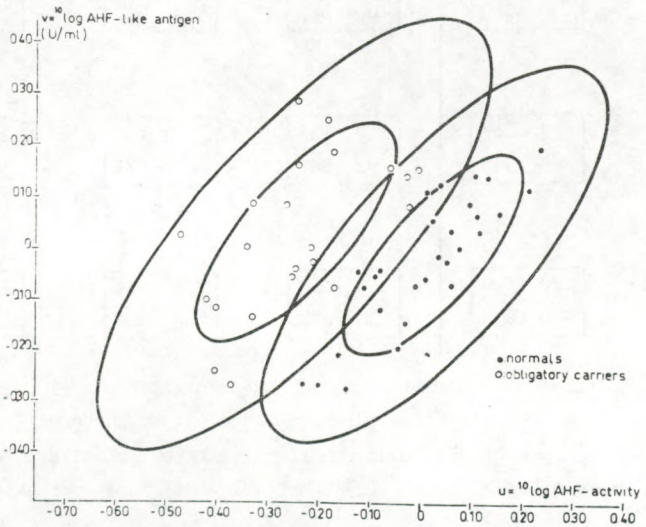
Deze resultaten wekken vertrouwen in de methode van Hui en Walter. Alle schattingen zijn schattingen van kansen welke alleen nauwkeurig kunnen zijn voor aantallen van enige honderden. De standaardfouten geven in dat geval een goede indruk van de nauwkeurigheid van deze schatters, mits aan alle voorwaarden / aannamen wordt voldaan. Deze aannamen zijn:

1. onafhankelijkheid van het maken van fouten binnen de werkelijk zieken resp. gezonden
2. sensitiviteit en specificiteit konstant in de twee subgroepen
3. prevalenties duidelijk verschillend in de twee subgroepen

In de praktijk zal niet altijd aan deze drie aannamen worden voldaan, zodat we tenminste de robuustheid van deze methode voor afwijkingen van deze aannamen zullen moeten nagaan.

3. ROBUUSTHEID VOOR DE AANNAME: ONAFHANKELIJKHEID VAN HET MAKEN VAN FOUTEN BINNEN WERKELIJK ZIEKEN EN GEZONDEN

De aanname in de methode van Hui en Walter dat de twee testen onafhankelijk van elkaar fouten maken binnen de werkelijk zieken en gezonden betekent in het bovenstaande geconstrueerde voorbeeld dat van de 90 door test 1 als juist gediagnostiseerde zieken er 80% door test 2 ziek worden bevonden, doch ook 80% van de 10 door test 1 verkeerd beoordeelde zieken (en omgekeerd). Ofwel in statistisch jargon: de twee testen zijn ongecorreleerd binnen werkelijk zieken en gezonden. Dit lijkt een voor de praktijk niet erg reële aanname, hetgeen we tevens kunnen illustreren met het volgende plaatje uit Habbema en Hermans (1978, pag.159):



Om tot een beter model te komen bekijken we eerst nog eens het gesimuleerde voorbeeld uit de eerste paragraaf. De totale groep van 1000 personen kunnen we opsplitsen in 300 zieken en 700 gezonden waarvoor geldt:

WERKELIJKHEID		TEST 1 +	TEST 1 −	TEST 2 +	TEST 2 −
ziek (+)	300	270	30	240	60
gezond(−)	700	70	630	70	630

In geval van een extreem positieve correlatie tussen de beide testen wat betreft sensitiviteit zulen de 240 door test 2 als juist gediagnostiseerde zieken ook allen door test 1 als ziek worden bevonden. In geval van een extreem negatieve correlatie tussen de twee testen voor sensitiviteit zullen de 30 door test 1 onjuist geklassificeerde zieken allemaal door test 2 juist worden gediagnostiseerd. We kunnen voor de afhankelijkheid van de foutenklassificaties de volgende tabel opstellen:

700 gezonden \ 300 zieken		+	−	EXTR.NEG. +	EXTR.NEG. −	ONAFH. +	ONAFH. −	EXTR.POS. +	EXTR.POS. −
				+ 210	60	+ 216	54	+ 240	30
				− 30	0	− 24	6	− 0	30
EXTREEM NEG.	+	0	70	+ 210	130	+ 216	124	+ 240	100
	−	70	560	− 100	560	− 94	566	− 70	590
ONAFH.	+	7	63	+ 217	123	+ 223	117	+ 247	93
	−	63	567	− 93	567	− 87	573	− 63	597
EXTREEM POS.	+	70	0	+ 280	60	+ 286	54	+ 310	30
	−	0	630	− 30	630	− 24	636	− 0	660

In deze tabel bevat de middelste cel de aantallen die al eerder gesimuleerd werden. Deze cel ligt veel dichter bij de linkerbovencel (extreem negatieve correlaties) dan bij de rechterondercel (extreem positieve correlaties). In plaats van onafhankelijkheid lijkt het volgende model voor de praktijk reëler.

De sensitiviteit van test 1 is 0.9, hetgeen betekent dat test 1 10% van alle zieken verkeerd beoordeelt. Deze 10% bestaat waarschijnlijk uit moeilijke gevallen welke eveneens onder de 90% juist beoordeelde zieken voorkomen. Stel dat deze 90% ook voor 10% bestaat uit moeilijke gevallen, terwijl de overige 80% door test 1 zeker juist beoordeeld worden. Voor de totaal 20% moeilijk te beoordelen zieken veronderstellen we nu dat de diagnose wordt gesteld door het gooien van een zuiver muntstuk: kop = ziek en munt = gezond! Hierdoor verkrijgen we precies de veronderstelde sensitiviteit van 0.9. Als we dezelfde redenering ophangen voor de sensitiviteit van test 2, dan bestaan de zieken uit 60% door test 2 als zeker juist te beoordelen en 40% door kop/munt. Schematisch kunnen we dit als volgt weergeven:

ZIEKEN: sensitiviteit test 1 = 0.9 sensitiviteit test 2 = 0.8

0.8 zeker + 0.2 kop/munt 0.6 zeker + 0.4 kop/munt

zitten daarbij

Als we nu veronderstellen dat de 60% door test 2 als zeker juist beoordeelde zieken ook allemaal door test 1 zeker juist worden beoordeeld, dan geldt voor de zieken:

ZIEKEN: 60% zeker + 40% kop/munt = 100% totaal

		T2 +	T2 −
T1	+	60	0
	−	0	0

+

		T2 +	T2 −
T1	+	15	15
	−	5	5

=

		T2 +	T2 −
T1	+	75	15
	−	5	5

Bij ongecorreleerdheid is het totaal:

		T2 +	T2 −
T1	+	72	18
	−	8	2

Doordat de testen nu gecorreleerd zijn zullen zij het in meer gevallen met elkaar eens zijn dan onder onafhankelijkheid, zodat er meer op de diagonalen komt te liggen. Voor de specificiteit wordt de analoge redenering:

GEZONDEN: specificiteit test 1 = 0.9 specificiteit test 2 = 0.9

0.8 zeker + 0.2 kop/munt 0.8 zeker + 0.2 kop/munt

zelfden

Als we nu veronderstellen dat de 80% door test 1 als zeker juist beoordeelde gezonden ook allemaal door test 2 zeker juist worden beoordeeld, dan geldt voor de gezonden:

GEZONDEN: 80% zeker + 20% kop/munt = 100% totaal

T1 \ T2	+	−
+	80	0
−	0	0

+

T1 \ T2	+	−
+	5	5
−	5	5

=

T1 \ T2	+	−
+	85	5
−	5	5

Bij ongecorreleerdheid is het totaal:

T1 \ T2	+	−
+	81	9
−	9	1

Als we nu deze percentages laten gelden voor het voorbeeld, dan krijgen we:

SUBGR.1: 100 zieken + 400 gezonden = 500 personen

T1 \ T2	+	−	
+	75	15	90
−	5	5	10
	80	20	100

+

T1 \ T2	+	−	
+	20	20	40
−	20	340	360
	40	360	400

=

T1 \ T2	+	−	
+	95	35	130
−	25	345	370
	120	380	500

SUBGR.2: 200 zieken + 300 gezonden = 500 personen

T1 \ T2	+	−	
+	150	30	180
−	10	10	20
	160	40	200

+

T1 \ T2	+	−	
+	15	15	30
−	15	255	270
	30	270	300

=

T1 \ T2	+	−	
+	165	45	210
−	25	265	290
	190	310	500

Zodat voor de totale groep van 1000 personen geldt:

T1 \ T2	+	−	
+	260	80	340
−	50	610	660
	310	690	1000

Het totaal is precies het gemiddelde van de 4 hoekcellen in de tabel met extreem positieve of negatieve correlaties. De schattingen met de methode van Hui en Walter worden:

SUBGR 1: prevalentie = .24 (.04)
SUBGR 2: prevalentie = .42 (.05)

TEST 1: sensitiviteit = .94 (.06)
specificiteit = .95 (.04)

TEST 2: sensitiviteit = .84 (.07)
specificiteit = .95 (.03)

95% B.I. voor (sens1-sens2): .10 $\pm$ 2*.09 ofwel (-.08,+.28)
95% B.I. voor (spec1-spec2): .01 $\pm$ 2*.04 ofwel (-.07,+.09)

Zoals valt te verwachten worden de schattingen voor de sensitiviteit en specificiteit voor beide testen hoger; het relatieve verschil ertussen blijft echter ongeveer gelijk. De methode kan wellicht nog steeds een redelijke uitspraak doen over het verschil tussen de beide testen, doch de absolute schattingen van de sensitiviteit en specificiteit blijken overtrokken hoog uit te vallen.

4. ROBUUSTHEID VOOR DE AANNAME: KONSTANTE SENSITIVITEIT EN SPECIFICITEIT IN DE TWEE SUBGROEPEN

Om de robuustheid van de methode van Hui en Walter te onderzoeken met betrekking tot konstantheid van sensitiviteit en specificiteit in de twee subgroepen, hebben we in het voorbeeld uit paragraaf 2 de sensitiviteit voor test 2 in subgroep 2 veranderd van .8 in .9. Voor deze subgroep worden de aantallen dan:

SUBGR 2: 200 zieken + 300 zieken = 500 personen

T1 \ T2	+	−	
+	162	18	180
−	18	2	20
	180	20	200

+

T1 \ T2	+	−	
+	3	27	30
−	27	243	270
	30	270	300

=

T1 \ T2	+	−	
+	165	45	210
−	45	245	290
	210	290	500

De schattingen met de methode van Hui en Walter worden nu:

SUBGR 1: prevalentie = .163 (.032)

SUBGR 2: prevalentie = .372(.050)

TEST 1: sensitiviteit = .900 (.066)
specificiteit = .864 (.029)

TEST 2: sensitiviteit = .960 (.081)
specificiteit = .900 (.029)

95% B.I. voor (sens1-sens2): $-.06 \pm 2*.10$ ofwel (-.26,+.14)

95% B.I. voor (spec1-spec2): $-.04 \pm 2*.04$ ofwel (-.12,+.04)

De invloed van deze verandering op de schattingen is vrij drastisch. Intuitief valt te verwachten dat de schatting voor de sensitiviteit van test 2 komt te liggen tussen .8 en .9, doch deze valt hierbuiten. Hiermee gaat tevens weer gepaard een flinke verandering in de specificiteit van test 1. Een waarschuwing is dus op zijn plaats: de methode van Hui en Walter blijkt erg gevoelig voor de aanname dat de sensitiviteit en specificiteit van de beide testen konstant zijn in de twee subgroepen. Dit houdt voor de praktijk in dat we het kriterium waarop we de totale groep in twee subgroepen splitsen met zorg moeten kiezen. Een voor de hand liggende splitsing als bijvoorbeeld in mannen en vrouwen kan wel eens een slechte keuze hiervoor zijn. Indien mogelijk lijkt het in ieder geval aan te bevelen om de totale groep op meerdere kriteria in tweeën te

splitsen en de uitkomsten voor de verschillende opsplitsingen met elkaar te vergelijken.

5. ROBUUSTHEID VOOR DE AANNAME: PREVALENTIES IN DE TWEE SUBGROEPEN DUIDELIJK VERSCHILLEND

De robuustheid voor de aanname dat de prevalenties in de twee subgroepen duidelijk verschillend moeten zijn is niet met gesimuleerde data te onderzoeken. In paragraaf 2 hebben we namelijk laten zien dat voor gesimuleerde data de ingevoerde parameters exact worden teruggeschat, hetgeen ook gebeurt indien de prevalenties in de twee subgroepen dicht bij elkaar liggen. Er moet daarom naast een klein verschil in de prevalenties ook een toevallige variatie in de getallen worden aangebracht. Voor meerdere praktijkgegevens liet de methode van Hui en Walter het evenwel afweten, onder andere voor de volgende gegevens (voor beschrijving zie paragraaf 6):

Responsie voor materiaal uit de urethra

SUBGR 1: geb.jaar 1928-1955

		KWEEK +	KWEEK −	
GONOZYME	+	8	4	12
	−	2	287	289
		10	291	301

SUBGR 2: geb.jaar 1956-1967

		KWEEK +	KWEEK −	
GONOZYME	+	17	2	19
	−	8	281	289
		25	283	308

De schattingen met de methode van Hui en Walter worden:

SUBGR 1 : prevalentie = .04 (.02)
SUBGR 2 : prevalentie = .07 (.03)

GONOZYME: sensitiviteit = .60 (.18)
specificiteit = .98 (.01)

KWEEK : sensitiviteit =1.27 (.48) !!
specificiteit =1.01 (.02) !!

De schattingen voor de prevalenties in de twee subgroepen liggen vrij dicht bij elkaar, hetgeen wellicht de oorzaak is dat de schattingen voor de sensitiviteit en specificiteit van de kweek buiten het toegelaten gebied [0,1] komen. Dit is voor ons aanleiding om in gevallen als deze te twijfelen of wel aan alle aannamen wordt voldaan; Hui en Walter zoeken in dit geval toch een geforceerde oplossing binnen de toegelaten ruimte. Het is dus oppassen geblazen indien, zoals in dit voorbeeld, de prevalenties in de twee subgroepen dicht bij elkaar liggen. Een kontrôle op deze aanname is alleen mogelijk via de schattingen die

in dit geval waarschijnlijk eveneens niet erg betrouwbaar zijn, zodat het toepassen van de methode van Hui en Walter hier een dubieuze zaak wordt.

6. EEN VOORBEELD UIT DE PRAKTIJK: VERGELIJKING VAN TWEE VERSCHILLENDE TESTEN TER BEPALING VAN GONORRHOE

Voor kontrôle op gonorrhoe zijn bij 609 vrouwelijke prostituees welke tussen 5/7/83 en 2/11/83 de polikliniek venerologie van het Dijkzigt ziekenhuis bezochten twee testen uitgevoerd, namelijk een gonozyme-test en een kweek op gonorrhoe. De groep is naar leeftijd gesplitst in twee ongeveer evengrote subgroepen; de volgende resultaten werden gemeten:

Responsie voor materiaal uit de cervix

SUBGR 1: geb.jaar 1928-1955

		KWEEK +	KWEEK −	
GONOZYME	+	13	3	16
	−	2	283	285
		15	286	301

SUBGR 2: geb.jaar 1956-1967

		KWEEK +	KWEEK −	
GONOZYME	+	31	3	34
	−	4	270	274
		35	273	308

De schattingen met de methode van Hui en Walter worden:

SUBGR 1 : prevalentie = .048 (.017)
SUBGR 2 : prevalentie = .113 (.025)

GONOZYME: sensitiviteit = .899 (.110)
specificiteit = .990 (.011)

KWEEK : sensitiviteit = .993 (.128)
specificiteit = .998 (.010)

95% B.I. voor (sens1-sens2): $-.09 \pm 2*.17$ ofwel (-.43,+.25)
95% B.I. voor (spec1-spec2): $-.01 \pm 2*.015$ ofwel (-.04,+.02)

Om op grond van deze gegevens een echt verschil tussen de twee testen op sensitiviteit te kunnen aantonen moet dus een nog groter aantal waarnemingen worden verricht. Wat betreft de specificiteit zijn beide testen ongeveer aan elkaar gelijk.

7. DISCUSSIE EN KONKLUSIES

De methode van Hui en Walter lijkt op het eerste gezicht de vondst waarmee we in het geval dat er geen "gouden standaard" beschikbaar is toch de mogelijkheid hebben om twee (of meer) verschillende diagnostische testen met elkaar te vergelijken. Om deze methode te mogen toepassen moet echter aan een tweetal aannamen worden voldaan, welke niet met de gegevens getoetst kunnen worden. Zodoende blijft er bij het toepassen van deze methode altijd enige onzekerheid bestaan en blijft de vraag wat de kwaliteit is van de gevonden schattingen, ook al worden hiervoor benaderende standaardfouten geproduceerd. Bij meerdere voorbeelden bleken sommige schattingen buiten het toegelaten gebied [0,1] uit te komen. Dit is aanleiding om in zulke gevallen te twijfelen of wel aan alle aannamen wordt voldaan; Hui en Walter zoeken in dit geval toch geforceerde oplossingen binnen de toegelaten ruimte. De methode lijkt alleen verantwoord toe te passen als een kontrôle op de aannamen mogelijk is. Hiervoor doen we de volgende suggesties:

1. Het waarnemingsmateriaal in meer dan 2 groepen splitsen op één kriterium, bijvoorbeeld naar leeftijd in 3 categorieën waarvoor wel weer de prevalenties duidelijk verschillend moeten zijn.
2. Het waarnemingsmateriaal op verschillende kriteria in twee groepen splitsen waarvoor de prevalenties duidelijk verschillend zijn, bijvoorbeeld naar leeftijd resp. geslacht.
3. Simultaan meer dan twee testen uitvoeren.

Op (een van) deze manieren kunnen min of meer onafhankelijke schattingen worden verkregen, zodat tenminste kan worden nagegaan of aan de gemaakte aannamen wel voldaan wordt. In Walter (1984) wordt de uitbreiding naar drie testen (waarnemers) beschreven. Voor ons is het duidelijk dat de methode van Hui en Walter met de nodige reserve moet worden toegepast, temeer daar de resultaten door hun eenvoud zo aanspreken.

8. LITERATUUR

Hui,S.L. and Walter,S.D.(1980). Estimating the Error Rates of Diagnostic Tests. Biometrics, vol.36,167-171.

Habbema,J.D.F. and Hermans,J.(1978). Statistical Methods for Clinical Decision Making. Proefschrift Leiden

Staquet,M. et al.(1981). Methodology for the assessment of new dichotomous diagnostic tests. J.Chron.Dis.,vol.34,599-610.

Walter,S.D.(1984). Measuring the reliability of clinical data: the case for using three observers. Rev.Epid.et Santé Publ.,vol.32,206-211.

Ontvangen: 3-4-1985

APPENDIX A1. Voorbeeld uit Hui en Walter

SUBGROUP 1			
		TEST 2	TOTAL
		+	-	
TEST 1	+	14	4	18
	-	9	528	537
TOTAL		23	532	555

IF TEST 2 STANDARD: PREVALENCE = .0414
TEST 1: SENSITIVITY = .6087 (.1018)
SPECIFICITY = .9925 (3.7E-03)
IF TEST 1 STANDARD: PREVALENCE = .0324
TEST 2: SENSITIVITY = .7778 (.098)
SPECIFICITY = .9832 (5.5E-03)

SUBGROUP 2			
		TEST 2	TOTAL
		+	-	
TEST 1	+	887	31	918
	-	37	367	404
TOTAL		924	398	1322

IF TEST 2 STANDARD: PREVALENCE = .6989
TEST 1: SENSITIVITY = .96 (6.4E-03)
SPECIFICITY = .9221 (.0134)
IF TEST 1 STANDARD: PREVALENCE = .6944
TEST 2: SENSITIVITY = .9662 (6E-03)
SPECIFICITY = .9084 (.0144)

SUBGROUP 1 + SUBGROUP 2			
		TEST 2	TOTAL
		+	-	
TEST 1	+	901	35	936
	-	46	895	941
TOTAL		947	930	1877

IF TEST 2 STANDARD: PREVALENCE = .5045
TEST 1: SENSITIVITY = .9514 (7E-03)
SPECIFICITY = .9624 (6.2E-03)
IF TEST 1 STANDARD: PREVALENCE = .4987
TEST 2: SENSITIVITY = .9626 (6.2E-03)
SPECIFICITY = .9511 (7E-03)

ESTIMATES (STANDARD ERROR):

SUBGR 1: PREVALENCE = .0268 (.0071)
SUBGR 2: PREVALENCE = .7168 (.0128)

TEST 1 : SENSITIVITY = .9661 (.0069)
SPECIFICITY = .9933 (.0038)

TEST 2 : SENSITIVITY = .9688 (.0062)
SPECIFICITY = .9841 (.0056)

CORRELATION\VARIANCE-COVARIANCE MATRIX :

	SPEC1	SENS1	SPEC2	SENS2	PREV1	PREV2
SPEC1	+.0000144	-.0000001	+.0000002	-.0000066	-.0000020	-.0000044
SENS1	-.0053560	+.0000482	-.0000137	+.0000012	+.0000021	+.0000100
SPEC2	+.0127300	-.3516221	+.0000315	-.0000001	-.0000023	-.0000096
SENS2	-.2819878	+.0290443	-.0039906	+.0000387	+.0000018	+.0000049
PREV1	-.0763156	+.0438308	-.0589944	+.0419857	+.0000508	+.0000015
PREV2	-.0910847	+.1125809	-.1338043	+.0625708	+.0171143	+.0001636

APPENDIX A2. Gesimuleerd voorbeeld

SUBGROUP 1			
	TEST 2 +	TEST 2 −	TOTAL
TEST 1 +	76	54	130
TEST 1 −	44	326	370
TOTAL	120	380	500

IF TEST 2 STANDARD: PREVALENCE = .24
TEST 1: SENSITIVITY = .6333 (.044)
SPECIFICITY = .8579 (.0179)
IF TEST 1 STANDARD: PREVALENCE = .26
TEST 2: SENSITIVITY = .5846 (.0432)
SPECIFICITY = .8811 (.0168)

SUBGROUP 2			
	TEST 2 +	TEST 2 −	TOTAL
TEST 1 +	147	63	210
TEST 1 −	43	247	290
TOTAL	190	310	500

IF TEST 2 STANDARD: PREVALENCE = .38
TEST 1: SENSITIVITY = .7737 (.0304)
SPECIFICITY = .7968 (.0229)
IF TEST 1 STANDARD: PREVALENCE = .42
TEST 2: SENSITIVITY = .7 (.0316)
SPECIFICITY = .8517 (.0209)

SUBGROUP 1 + SUBGROUP 2			
	TEST 2 +	TEST 2 −	TOTAL
TEST 1 +	223	117	340
TEST 1 −	87	573	660
TOTAL	310	690	1000

IF TEST 2 STANDARD: PREVALENCE = .31
TEST 1: SENSITIVITY = .7194 (.0255)
SPECIFICITY = .8304 (.0143)
IF TEST 1 STANDARD: PREVALENCE = .34
TEST 2: SENSITIVITY = .6559 (.0258)
SPECIFICITY = .8682 (.0132)

ESTIMATES (STANDARD ERROR):

SUBGR 1: PREVALENCE = .2000 (.0440)
SUBGR 2: PREVALENCE = .4000 (.0574)

TEST 1 : SENSITIVITY = .9000 (.0804)
SPECIFICITY = .9000 (.0372)

TEST 2 : SENSITIVITY = .8000 (.0757)
SPECIFICITY = .9000 (.0316)

CORRELATION\VARIANCE-COVARIANCE MATRIX :

	SPEC1	SENS1	SPEC2	SENS2	PREV1	PREV2
SPEC1	+.0013817	-.0004026	+.0001716	-.0025102	-.0011223	-.0013725
SENS1	-.1347278	+.0064631	-.0022744	+.0009077	+.0021533	+.0031735
SPEC2	+.1460533	-.8948429	+.0009995	-.0003082	-.0008189	-.0011871
SENS2	-.8918752	+.1491141	-.1287621	+.0057334	+.0022647	+.0028090
PREV1	-.6862862	+.6088090	-.5887871	+.6798026	+.0019357	+.0019634
PREV2	-.6429259	+.6873026	-.6538020	+.6459241	+.7770199	+.0032986