

Sensitiviteit, Specificiteit, Aannemelijkheidsverhouding en
Voorspellende Waarden van Diagnostische Testen. (1)

Paul A.R. Koopman (2)

Samenvatting

De waarde van een diagnostische test kan uitgedrukt worden met begrippen als sensitiviteit, specificiteit, aannemelijkheidsverhoudingen en voorspellende waarden. Voor deze indices, alsmede voor het verschil in voorspellende waarde tussen twee diagnostische testen, worden de proefopzet, puntschatters en betrouwbaarheidsgebieden besproken.

De indeling van het artikel is zodanig dat het geschikt is om slechts kennis te nemen van de 'receptuur' : de wiskundige verantwoording is zoveel mogelijk naar de laatste paragraaf verschoven.

(1) Dit onderzoek is uitgevoerd in de periode dat de auteur werkzaam was bij de afdeling Medische Statistiek der Vrije Universiteit.

(2) Huidig adres: Duphar B.V., afdeling W.I.S., postbus 2, 1380 AA Weesp.

Inhoudsopgave:

- §1. Inleiding
- §2. Sensitiviteit en specificiteit
 - 2.1. Proefopzet en model
 - 2.2. Betrouwbaarheidsgebieden
 - 2.3. Voorbeeld
- §3. Aannemelijkheidsverhoudingen (λ en n) en voorspellende waarden (ρ_+ en ρ_-)
 - 3.1. Proefopzet en model
 - 3.2. Schatters voor λ en ρ_+
 - 3.3. Betrouwbaarheidsgebieden voor λ en ρ_+
 - 3.4. Voorbeeld
- §4. Het verschil (κ_+ of κ_-) tussen voorspellende waarden van twee diagnostische testen
 - 4.1. Proefopzet en model
 - 4.2. Een schatter voor κ_+
 - 4.3. Een betrouwbaarheidsinterval voor κ_+
 - 4.4. Een betrouwbaarheidsgebied voor (κ_+ , κ_-)
 - 4.5. Voorbeeld
- §5. Verantwoording
 - 5.1. Het simultane betrouwbaarheidsgebied voor de sensitiviteit en de specificiteit
 - 5.2. De schatter en het betrouwbaarheidsinterval voor het quotient van twee binomiale kansen
 - 5.3. Het betrouwbaarheidsinterval voor κ_+
 - 5.3.1. Drie methodes
 - 5.3.2. Een schatting voor de variantie in $\underline{k}_+$
 - 5.3.3. Een vergelijking tussen de drie methodes

§1. Inleiding

Het klinische beslissingsproces wordt mede gebaseerd op diagnostische testen zoals bijv. de bepaling van het calciumgehalte in het bloed of een waarneming op een röntgenfoto. Voor een juist gebruik van zo'n resultaat dient uiteraard de waarde van de test voor de diagnostiek bekend te zijn.

Voor de situatie dat de test slechts twee mogelijke uitslagen kent, nl. positief (T_+) of negatief (T_-) zijn in de literatuur begrippen geïntroduceerd die de waarde van de test aangeven. Uiteraard zijn die begrippen ook van toepassing op testen met kwantitatieve uitkomsten die na keuze van een discriminatiegrens gedichotomiseerd worden. Een aantal van deze begrippen zal de revue passeren. Het al dan niet hebben van de ziekte in kwestie wordt daarbij aangeduid met D_+ resp. D_- . In 1947 introduceerde Yerushalmy (9) de nosologische begrippen sensitiviteit en specificiteit van een diagnostische test. Sensitiviteit is de fractie patiënten met de ziekte D die als zodanig door de test wordt herkend, in wiskundige notatie $\Pr (T_+ \mid D_+)$. Specificiteit is de fractie personen die terecht door de test wordt aangeduid als vrij van ziekte D , genoteerd met $\Pr (T_- \mid D_-)$. In de medische literatuur volstaat men vrijwel altijd met het geven van puntschattingen voor de sensitiviteit en specificiteit. Het spreekt echter voor zich dat betrouwbaarheidsgebieden meer informatie verschaffen. Dergelijke gebieden worden in §2 beschreven.

In aanvulling op de sensitiviteit en specificiteit introduceerde Vecchio (7) in 1966 een tweetal diagnostische begrippen: de voorspellende waarde van een positieve testuitslag (ρ_+) en de voorspellende waarde van een negatieve testuitslag (ρ_-). Hierbij zijn $\rho_+ = \Pr (D_+ \mid T_+)$ en $\rho_- = \Pr (D_- \mid T_-)$ dan de kansen om de feitelijke toestand van de onderzocht persoon te voorspellen. Het ligt voor de hand dat deze kansen afhankelijk zijn van de prevalentie, i.e. de fractie personen in de populatie die de ziekte heeft.

Dit verschijnsel formaliseren we als volgt : noteer de prevalentie $\Pr (D_+)$ als γ en de aannemelijkheidsverhoudingen

$$\frac{\Pr (T_+ \mid D_+)}{\Pr (T_+ \mid D_-)} \quad \text{en} \quad \frac{\Pr (T_- \mid D_+)}{\Pr (T_- \mid D_-)} \quad \text{als } \lambda \text{ respectievelijk } \eta.$$

Dan volgt op elementaire wijze uit het theorema van Bayes dat

$$\rho_+ = \frac{\gamma\lambda}{\gamma\lambda + 1 - \gamma} \quad \text{en} \quad \rho_- = \frac{(1-\gamma)\eta}{(1-\gamma)\eta + \gamma}$$

De aannemelijkheidsverhoudingen λ en η zijn hierbij functie's van de sensitiviteit en specificiteit, en als zodanig ook bruikbaar bij de beoordeling van diagnostische testen. In §3 worden voor λ, η, ρ_+ en ρ_- schatters en betrouwbaarheidsgebieden gegeven.

In §4 vergelijken we 2 diagnostische testen met elkaar. Daarbij beperken we ons tot voorspellende waarden omdat deze in feite de belangrijkste parameters zijn bij het diagnosticeren. Er worden schatters en betrouwbaarheidsgebieden voor verschillen in de voorspellende waarden voorgesteld.

Ter illustratie zal een (fictief) voorbeeld als een rode draad door het artikel lopen.

Terwille van de leesbaarheid zijn in §§2 t/m 4 mathematische formules zoveel mogelijk vermeden. De opzet van die §§ is nl. om een 'receptuur' te geven, los van een wiskundige verantwoording. Deze komt aan de orde in §5.

§2. Sensitiviteit en specificiteit

2.1. Proefopzet en model

Men kan op twee manieren gegevens verzamelen om de sensitiviteit en specificiteit te schatten.

- (1) Trek een aselechte steekproef van m zieken en een aselechte steekproef van $n-m$ gezonden. Bepaal voor elk van die personen de testuitslag.
- (2) Trek een aselechte steekproef van n personen. Bepaal voor ieder persoon de testuitslag en de feitelijke gezondheidstoestand (onaf-

hankelijk van elkaar).

We nemen hierbij impliciet aan dat het mogelijk is om de feitelijke gezondheid foutloos vast te stellen.

In beide gevallen geven we de resultaten weer in een 2x2- tabel.

	T_+	T_-	
D_+	a	b	m
D_-	c	d	n-m
			n

Bij methode (1) heeft $\underline{a}$ (stochasten worden onderstreept) een binomiale verdeling $B(m, p_a)$ waarbij de sensitiviteit gelijk is aan p_a en heeft $\underline{d}$ een binomiale verdeling $B(n-m, p_d)$ waarbij de specificiteit gelijk is aan p_d .

Bij methode (2) heeft $(\underline{a}, \underline{b}, \underline{c}, \underline{d})$ een multinomiale verdeling $M(n: p_1, p_2, p_3, p_4)$ waarbij de sensitiviteit en de specificiteit gelijk zijn aan $p_1/(p_1+p_2)$ resp. $p_4/(p_3+p_4)$.

Dit laatste oogt in vergelijking met methode (1) niet erg aantrekkelijk als het er om gaat betrouwbaarheidsgebieden voor de sensitiviteit en specificiteit af te leiden. Omdat we echter betrouwbaarheidsgebieden voor een parameter θ laten bestaan uit alle waarden ε waarvoor $H: \theta = \varepsilon$ niet verworpen worden ten gunste van het alternatief $K_\varepsilon: \theta \neq \varepsilon$, valt het nog wel mee. Immers standaard theorie voor exponentiele familie's zoals beschreven in Lehmann (5) leert ons dat we voor dergelijke toetsingsproblemen 'similar' toetsen met onbetrouwbaarheid α mogen conditioneren op de 'ancillary statistic' $\underline{a} + \underline{b}$ zonder de onbetrouwbaarheid te veranderen. Dit betekent dat we ook bij methode (2) het statistische model van de eerste methode kunnen hanteren.

2.2. Betrouwbaarheidsgebieden

Betrouwbaarheidsintervallen voor de sensitiviteit en specificiteit, als binomiale kansen, zijn in vrijwel elk elementair boek op statistisch gebied te vinden.

Het kan echter ook zinvol zijn om deze beide parameters van de diagnostische test gezamenlijk te beoordelen.

In dat geval geeft een simultaan betrouwbaarheidsgebied uitkomst. Een twee-dimensionaal (benaderend) $(1-\alpha)$ betrouwbaarheidsgebied wordt gegeven door de ovaal

$$\{ (p_a, p_d) : \frac{(a - mp_a)^2}{mp_a(1-p_a)} + \frac{(d - (n-m)p_d)^2}{(n-m)p_d(1-p_d)} < \chi^2_{2, 1-\alpha} \}$$

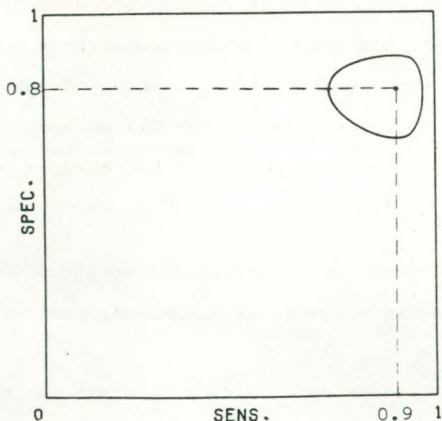
waarbij $\chi^2_{2, 1-\alpha}$ het $(1-\alpha)$ fractiel is van een χ^2 -verdeling met twee vrijheidsgraden.

Wanneer het aantal waarnemingen gering is, valt een continuïteitscorrectie te overwegen.

2.3. Voorbeeld

Ter evaluatie van een diagnostische test wordt bij 40 personen die de ziekte in kwestie onder de leden hebben, de testuitslag bepaald. Hiervan zijn er 36 positief. Van de 80 gezonde personen die getest werden, waren er 64 negatief.

De geschatte sensitiviteit is dan 0.90 en een 95% betrouwbaarheidsinterval is (0.763, 0.972). De geschatte specificiteit is 0.80 en een 95% interval is (0.696, 0.881). In figuur 1 is een twee-dimensionaal ovaal 95% betrouwbaarheidsgebied gegeven. We merken volledigheidshalve op dat de betrouwbaarheidsintervallen niet gelijk zijn aan de projectie's van de ovaal op de assen.



Figuur 1: Puntschatter en 95% betrouwbaarheidsgebied voor sens. en spec.

§3. Aannemelijkheidsverhoudingen en voorspellende waarden

3.1. Proefopzet en model

In §1 is reeds opgemerkt dat de voorspellende waarden afhankelijk zijn van de prevalentie γ . Deze γ is de kans op de betreffende ziekte in de doelgroep.

Wanneer de steekproef uit deze doelgroep genomen wordt, is het met een multinomiale proefopzet (zie 2.1.) mogelijk om niet alleen de aannemelijkheidsverhoudingen λ en n , maar ook de prevalentie ($\gamma = p_1 + p_2$) te schatten. Tevens leidt conditioneren op $\underline{a+c} = a+c$ er dan toe dat de voorspellende waarden overgaan in binomiale kansen zodat schatters en betrouwbaarheidsgebieden op simpele wijze verkregen worden.

In de praktijk echter zal het vaak voorkomen dat de doelgroep (o.a. qua prevalentie) verschilt van de populatie die gebruikt is om de waarden van de diverse test-indices te schatten. Zo is gevoeglijk aan te nemen dat bijvoorbeeld de prevalentie van een ziekte in een huisartsenpraktijk anders zal zijn dan in een specialistenpraktijk. In het vervolg zullen we ons dan ook beperken tot een dergelijke situatie.

Uit de resultaten van de diagnostische test kan dan alleen informatie over de aannemelijkheidsverhoudingen verkregen worden. Immers in §1 hebben we gezien dat deze verhoudingen functie's zijn van de sensitiviteit en specificiteit. Deze op hun beurt kunnen, zoals Feinstein (2) betoogt, algemeen geldig worden geacht voor elke groep personen. Wanneer men toch de voorspellende waarde van een testuitslag wil betrekken bij het stellen van een diagnose, moet γ anderszins bekend zijn. Soms is dan een subjectieve keuze nodig. Deze problematiek wordt door Weinstein et al. (8) besproken. In de onderhavige studie zal γ verder bekend worden verondersteld.

Zoals eerder kan het statistische model van methode (1) uit 2.1. gebruikt worden m.b.t. de aannemelijkheidsverhoudingen en voorspellende

waarden. Dit impliceert dat λ en n als quotiënten van 2 binomiale kansen beschouwd worden, nl. $\lambda = p_a / (1 - p_d)$ en $n = p_d / (1 - p_a)$.

In 3.2. en 3.3. zullen slechts resultaten m.b.t. λ en ρ_+ gegeven worden: door herschikking van de 2x2-tabel zijn op analoge wijze formules m.b.t. n en ρ_- af te leiden.

3.2. Schatters voor λ en ρ_+

Voor de schatter $\hat{\lambda} = \{a \cdot (n-m+1)\} / \{(c+1) \cdot m\}$ geldt dat $E(\hat{\lambda} - \lambda) = -p_d^{n-m+1} \lambda$. Deze schatter van λ heeft een onzuiverheid die voor $n-m \rightarrow \infty$ exponentieel afneemt. Bovendien is vrij eenvoudig in te zien dat $\hat{\rho}_+ = \gamma \hat{\lambda} / (\gamma \hat{\lambda} + 1 - \gamma)$ als schatter van ρ_+ een onzuiverheid van dezelfde orde heeft.

3.3. Betrouwbaarheidsgebieden voor λ en ρ_+

Een (benaderend) $(1-\alpha)$ betrouwbaarheidsinterval voor λ wordt gegeven

$$\text{door } \{\lambda_0 : X^2(\lambda_0) = \frac{(a - m\hat{p}_a)^2}{m\hat{p}_a(1 - \hat{p}_a)} + \frac{(d - (n-m)\hat{p}_d)^2}{(n-m)\hat{p}_d(1 - \hat{p}_d)} < \chi_{1, 1-\alpha}^2\}$$

waarbij $\hat{p}_a$ en $\hat{p}_d$ de meest aannemelijke schattingen zijn onder de restrictie $\lambda = \lambda_0$. Na berekening blijkt dat

$$\hat{p}_a = \frac{\lambda_0(n-d) + a + n - m - [\{\lambda_0(n-d) + a + n - m\}^2 - 4\lambda_0 n(a + n - m - d)]^{\frac{1}{2}}}{2n} \quad \text{en} \quad \hat{p}_d = 1 - \frac{\hat{p}_a}{\lambda_0}$$

Omdat X^2 een convexe functie van λ_0 is, ontstaat een asymmetrisch interval (λ_l, λ_r) met $X^2(\lambda_l) = X^2(\lambda_r) = \chi_{1, 1-\alpha}^2$ en $\lambda_l < \lambda_r$.

De waarden voor λ_l en λ_r kunnen slechts numeriek verkregen worden, bijv. via een bisectiemethode.

Het $(1-\alpha)$ betrouwbaarheidsinterval voor ρ_+ wordt nu gegeven door

$$\left(\frac{\gamma \lambda_l}{\gamma \lambda_l + 1 - \gamma}, \frac{\gamma \lambda_r}{\gamma \lambda_r + 1 - \gamma} \right) \quad \text{omdat } \rho_+ \text{ een strict stijgende functie in } \lambda \text{ is.}$$

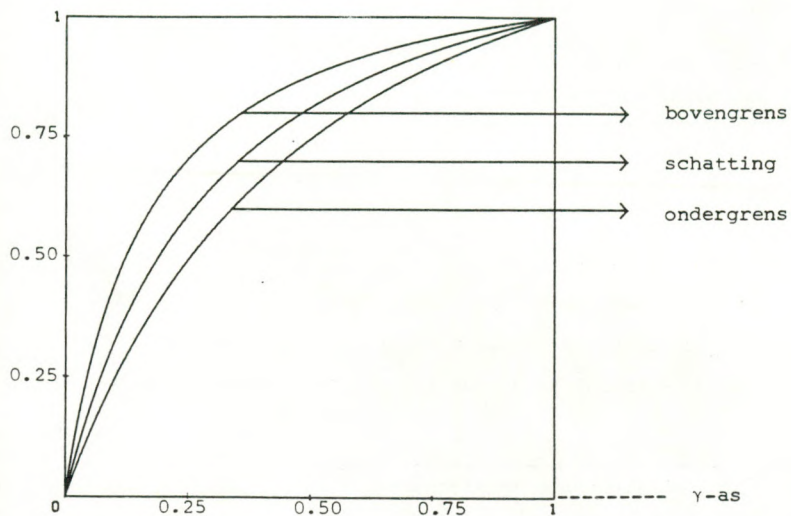
Afhankelijk van het doel der diagnostische testen, is soms de voorspellende waarde van de positieve testuitslag, soms die van de negatieve testuitslag de belangrijkste parameter (zie Feinstein (2)).

Indien beide parameters even relevant zijn, is een simultane beschouwing aangewezen. Simultane gebieden voor (λ, n) of (ρ_+, ρ_-) kunnen via

één-éénduidige transformatie verkregen worden uit het ovale betrouwbaarheidsgebied voor de sensitiviteit en specificiteit (2.2.). We werken dit niet verder uit.

3.4. Voorbeeld

Voortbordurend op de gegevens in 2.3. vinden we $\hat{\lambda}=4.29$ als schatter voor de aannemelijkheidsverhouding λ . Het 95% interval voor λ is (2.94, 7.17). In figuur 2 is bij elke prevalentie γ de schatting en een 95% betrouwbaarheidsinterval voor ρ_+ afgebeeld.



Figuur 2: Schatting en 95% interval voor de voorspellende waarde van een positieve testuitslag, als functie van de prevalentie γ , bij $\hat{\lambda}=4.29$.

§4. Het verschil tussen voorspellende waarden van twee diagnostische testen

4.1. Proefopzet en model

De twee testen T_1 en T_2 kunnen op één steekproef of op aparte steekproeven zijn uitgevoerd. We zullen ons hier beperken tot de situatie waarbij beide testen onafhankelijk bij dezelfde personen worden afgenomen. Immers als dit praktisch haalbaar is, is er een variatiebron minder in het spel. We veronderstellen daarbij dat de feitelijke gezondheidstoestand onafhankelijk van de onderhavige testen vastgesteld is.

De resultaten worden weergegeven in een 2×4 -tabel.

	T_{1+}, T_{2+}	T_{1+}, T_{2-}	T_{1-}, T_{2+}	T_{1-}, T_{2-}	
D_+	a	b	c	d	n_1
D_-	e	f	g	h	n_2

Ook nu veronderstellen we de prevalentie γ bekend.

De 2×4 -tabel wordt beschouwd als een realisatie van de multinomiale verdelingen $M(n_1: p_a, p_b, p_c, p_d)$ en $M(n_2: p_e, p_f, p_g, p_h)$.

Het verschil tussen de voorspellende waarde van positieve testuitslagen is dan

$$\kappa_+ \text{ def } \rho_{1+} - \rho_{2+} = \frac{\gamma \lambda_1}{\gamma \lambda_1 + 1 - \gamma} - \frac{\gamma \lambda_2}{\gamma \lambda_2 + 1 - \gamma} \quad \text{waarbij}$$

$\lambda_1 = (p_a + p_b) / (p_e + p_f)$ en $\lambda_2 = (p_a + p_c) / (p_e + p_g)$ aannemelijkheidsverhoudingen zijn.

Een analoge uitdrukking voor $\kappa_- = \rho_{1-} - \rho_{2-}$ wordt verkregen door de indices a t/m h te vervangen door h t/m a, en γ door $1 - \gamma$. Derhalve wordt in 4.2. en 4.3. slechts κ_+ beschouwd.

4.2. Een schatter voor κ_+

Definieer $\hat{\lambda}_1 = \frac{a+b}{n_1} \cdot \frac{n_2+1}{e+f+1}$ en $\hat{\lambda}_2 = \frac{a+c}{n_1} \cdot \frac{n_2+1}{e+g+1}$, dan is analoog

aan 2.2. $\hat{\kappa}_+ = \{\gamma\lambda_1/(\gamma\lambda_1+1-\gamma)\} - \{\gamma\lambda_2/(\gamma\lambda_2+1-\gamma)\}$ een schatting voor κ_+ met $E(\hat{\kappa}_+ - \kappa_+) = \exp(-O(n_2))$.

4.3. Een betrouwbaarheidsinterval voor κ_+

Definieer $q_1 = \frac{a+b}{n_1} \cdot \frac{n_2}{e+f}$ en $q_2 = \frac{a+c}{n_1} \cdot \frac{n_2}{e+g}$ en zij $\kappa_+ = \frac{\gamma q_2}{\gamma q_1 + 1 - \gamma} - \frac{\gamma q_2}{\gamma q_2 + 1 - \gamma}$.

In 5.3. wordt een schatter $\hat{\sigma}^2$ voor de asymptotische variantie van $\hat{\kappa}_+$ gegeven.

Noteer ϕ en ξ voor de standaardnormale verdelingsfunctie en haar inverse. We memoreren dat $\operatorname{arctanh}(x) \stackrel{\text{def}}{=} 0.5 \ln \{(1+x)/(1-x)\}$.

Dan kan worden bewezen (Bishop et al. (1), section 14.6) dat voor

$$n_1 \rightarrow \infty \text{ en } n_2 \rightarrow \infty \Pr\left(\frac{\operatorname{arctanh}(\hat{\kappa}_+) - \operatorname{arctanh}(\kappa_+)}{\hat{\sigma}} \cdot (1 - \hat{\kappa}_+^2) < x\right) \rightarrow \phi(x).$$

Noteer $u \stackrel{\text{def}}{=} 2\xi_{1-0.5\alpha} \frac{\hat{\sigma}}{1 - \hat{\kappa}_+^2}$ waarbij $\hat{\sigma}$ de realisatie is van $(\hat{\sigma}^2)^{\frac{1}{2}}$,

dan is een benaderend $(1-\alpha)$ betrouwbaarheidsinterval voor κ_+

$$\left(\frac{(1+\hat{\kappa}_+)e^{-u} - (1-\hat{\kappa}_+)}{(1+\hat{\kappa}_+)e^{-u} + (1-\hat{\kappa}_+)} , \frac{(1+\hat{\kappa}_+)e^u - (1-\hat{\kappa}_+)}{(1+\hat{\kappa}_+)e^u + (1-\hat{\kappa}_+)} \right).$$

4.4. Een betrouwbaarheidsgebied voor (κ_+, κ_-)

Met een twee-dimensionale uitbreiding van de methode uit 4.3. verkrijgen

we een betrouwbaarheidsgebied voor (κ_+, κ_-) . Een schatting $\hat{\Sigma}_{\kappa}$ voor de asymptotische covariantie-matrix van $(\operatorname{arctanh} \hat{\kappa}_+, \operatorname{arctanh} \hat{\kappa}_-)$

wordt gegeven door

$$\hat{\Sigma}_{\kappa} = \left(\frac{1}{1-\hat{\kappa}_+^2} \cdot \frac{\partial \hat{\kappa}_+}{\partial t} , \frac{1}{1-\hat{\kappa}_-^2} \cdot \frac{\partial \hat{\kappa}_-}{\partial t} \right) \cdot \hat{\Sigma} \cdot \left(\frac{1}{1-\hat{\kappa}_+^2} \cdot \frac{\partial \hat{\kappa}_+}{\partial t} , \frac{1}{1-\hat{\kappa}_-^2} \cdot \frac{\partial \hat{\kappa}_-}{\partial t} \right)'$$

In 5.3.2. worden uitdrukkingen voor $\hat{\Sigma}$, $\frac{\partial \hat{\kappa}_+}{\partial t}$ en $\frac{\partial \hat{\kappa}_-}{\partial t}$ gegeven.

Definieer $V = \begin{pmatrix} \operatorname{arctanh}(\hat{\kappa}_+) - \operatorname{arctanh}(\kappa_+) \\ \operatorname{arctanh}(\hat{\kappa}_-) - \operatorname{arctanh}(\kappa_-) \end{pmatrix}$, dan is $\hat{Q}^2 \stackrel{\text{def}}{=} V'(\hat{\Sigma}_{\kappa})^{-1}V$

asymptotisch χ^2 -verdeeld met 2 vrijheidsgraden. Dit geeft een benaderend $(1-\alpha)$ betrouwbaarheidsgebied $\{(\kappa_+, \kappa_-): \hat{Q}^2 < \chi_{2, 1-\alpha}^2\}$.

4.5. Voorbeeld

Om de diagnostische test uit voorbeeld 2.3. en 3.4. te vergelijken met een tweede test, wordt ook de uitslag van deze laatste test bepaald bij de 120 personen uit de steekproef. Dit leidt tot

	T_{1+}, T_{2+}	T_{1+}, T_{2-}	T_{1-}, T_{2+}	T_{1-}, T_{2-}	
D_+	30	6	2	2	40
D_-	4	12	16	48	80

Veronderstel dat de testen bedoeld zijn voor een populatie met prevalentie $\gamma=0.20$.

Een schatting voor κ_+ is dan 0.082 en een 95% betrouwbaarheidsinterval is $(-0.065, 0.231)$ met $\hat{g}=0.076$.

Een schatting voor κ_- is 0.032 en een 95% interval is $(-0.008, 0.073)$ met $\hat{g}=0.021$.

We concluderen tevens dat, hoewel bij test 1 de schattingen van de voorspellende waarden hoger zijn, er geen verschil aangetoond is bij onbetrouwbaarheidsdrempels van 5%.

§5. Verantwoording

In deze paragraaf wordt een wiskundige verantwoording gegeven van hetgeen in §§ 2 t/m 4 aan de orde is geweest.

5.1. Het simultane betrouwbaarheidsgebied voor sensitiviteit en specificiteit

In 2.2. is een ovaal $(1-\alpha)$ betrouwbaarheidsgebied OV geïntroduceerd. Omdat $\underline{a}$ en $\underline{d}$ in de binomiale steekproefopzet onafhankelijk zijn, is via de $\sqrt{1-\alpha}$ betrouwbaarheidsintervallen voor de sensitiviteit resp. de specificiteit een rechthoekig $(1-\alpha)$ betrouwbaarheidsgebied RE te verkrijgen.

Om een keuze tussen OV en RE te maken, worden beide gebieden vergeleken in oppervlakte en 'accuracy'.

De oppervlakte van OV is ongeveer $\pi \cdot \frac{\sqrt{a(m-a)d(n-m-d)}}{m^3 (n-m)^3} \cdot \chi_{2,1-\alpha}^2$

en die van RE $4 \cdot \frac{\sqrt{a(m-a)d(n-m-d)}}{m^3 (n-m)^3} \cdot \chi_{1, \sqrt{1-\alpha}}^2$.

Omdat $(\pi \chi_{2,1-\alpha}^2) / (4 \chi_{1, \sqrt{1-\alpha}}^2)$ voor $\alpha=0.01$ gelijk is aan 0.92 en voor $\alpha=0.05$ gelijk is aan 0.94, heeft de ovaal een kleinere oppervlakte dan de rechthoek.

De 'accuracy' van de ovaal OV is voor p_a^x, p_d^x gedefinieerd als

$$\Pr(p_a^x, p_d^x) ((p_a, p_d) \notin OV) = \Pr(p_a^x, p_d^x) \left(\frac{(a - mp_a)^2}{mp_a(1-p_a)} + \frac{(d - (n-m)p_d)^2}{(n-m)p_d(1-p_d)} \geq \chi_{2,1-\alpha}^2 \right).$$

Dit is te schrijven als

$$\Pr(p_a^x, p_d^x) \left(\frac{p_a^x(1-p_a^x)}{p_a(1-p_a)} \left[\frac{w_1 - \frac{a - mp_a^x}{\sqrt{mp_a^x(1-p_a^x)}}}{\frac{a - mp_a^x}{\sqrt{mp_a^x(1-p_a^x)}}} \right]^2 + \frac{p_d^x(1-p_d^x)}{p_d(1-p_d)} \left[\frac{w_2 - \frac{(n-m)(p_d - p_d^x)}{\sqrt{(n-m)p_d^x(1-p_d^x)}}}{\frac{(n-m)(p_d - p_d^x)}{\sqrt{(n-m)p_d^x(1-p_d^x)}}} \right]^2 \geq \chi_{2,1-\alpha}^2 \right) \text{ waarbij } w_1 \text{ d\text{e}f } \frac{a - mp_a^x}{\sqrt{mp_a^x(1-p_a^x)}} \text{ en } w_2 \text{ d\text{e}f } \frac{d - (n-m)p_d^x}{\sqrt{(n-m)p_d^x(1-p_d^x)}}$$

onafhankelijk zijn en bij benadering $N(0,1)$ verdeeld zijn.

Met behulp van Ruben(6), die de verdelingsfunctie's geeft van stochasten met de vorm $\sum_{i=1}^n \lambda_i (w_i - w_i)^2$, waarbij λ_i en w_i bekende constanten zijn, is de 'accuracy' van OV te berekenen.

De accuracy van de rechthoek RE is $\Pr(p_a^x, p_d^x) ((p_a, p_d) \notin RE)$.

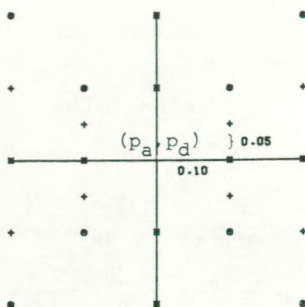
Dit is bij benadering gelijk aan

$$1 - \left[\Phi \left(\frac{\sqrt{m}(p_a - p_a^x) + \sqrt{p_a(1-p_a)} \xi_{(1+\sqrt{1-\alpha})/2}}{\sqrt{p_a^x(1-p_a^x)}} \right) - \Phi \left(\frac{\sqrt{m}(p_a - p_a^x) - \sqrt{p_a(1-p_a)} \xi_{(1+\sqrt{1-\alpha})/2}}{\sqrt{p_a^x(1-p_a^x)}} \right) \right] \\ \cdot \left[\Phi \left(\frac{\sqrt{n-m}(p_d - p_d^x) + \sqrt{p_d(1-p_d)} \xi_{(1+\sqrt{1-\alpha})/2}}{\sqrt{p_d^x(1-p_d^x)}} \right) - \Phi \left(\frac{\sqrt{n-m}(p_d - p_d^x) - \sqrt{p_d(1-p_d)} \xi_{(1+\sqrt{1-\alpha})/2}}{\sqrt{p_d^x(1-p_d^x)}} \right) \right]$$

Voor $n=50, 100$; $m=0.25n, 0.50n, 0.75n$; $p_a = 0.50$ (0.15) 0.95;

$p_d = 0.50$ (0.15) 0.95 en het schema voor (p_a^x, p_d^x) zoals in figuur 3

(voor zover binnen $(0,1) \times (0,1)$) zijn de 'accuracy' van OV en RE berekend.



Figuur 3: Het schema voor (p_a^*, p_d^*)

Omdat diagnostische testen met een sensitiviteit of specificiteit van minder dan 50% niet erg nuttig zijn, zijn slechts $p_a \geq 0.50$ en $p_d \geq 0.50$ gekozen.

Zowel voor $\alpha=0.01$ als voor $\alpha=0.05$ blijkt dat de accuracy van RE minus die van OV tussen -0.01 en 0.03 ligt voor alternatieven *, tussen -0.10 en 0.01 voor alternatieven +, en tussen -0.15 en 0.01 voor alternatieven o. Deze resultaten geven de indruk dat de ovaal in het algemeen een grotere accuracy heeft.

Op grond hiervan en ook vanwege het kleinere oppervlak, is in 2.2. de ovaal voorgesteld.

5.2. Het quotient van twee binomiale kansen

De in 3.3. beschreven formules worden besproken in Koopman(4). Daarin komt tevens een vergelijking aan de orde tussen het in 3.3. voorgestelde interval voor λ en een dienaangaande door Katz et al. (3) aanbevolen procedure.

Wat betreft de puntschatter, merken we eerst op dat er geen zuivere schatter bestaat voor het quotient van twee binomiale kansen.

De meest aannemelijke schatter $\{(\underline{a}/m)\}/\{(\underline{c}/(n-m))\}$ heeft zelfs een oneindige verwachting omdat er een positieve kans bestaat dat $\underline{c}$ de waarde 0 aanneemt. Indien het geval $\underline{c}=0$ omzeild wordt door dan $\underline{c}=0.5$

te kiezen, krijgen we een schatter met een onzuiverheid van de orde $(n-m)^{-1}$.

Voorgesteld is om $\{\hat{a}/m\} / \{(\hat{c}+1)/(n-m+1)\}$ als schatter voor het quotient te gebruiken. We verkrijgen dan een verdere reductie in de onzuiverheid want na enig rekenwerk blijkt de verwachting hiervan gelijk te zijn aan $\lambda\{1-(1-p_c)^{n-m+1}\}$.

5.3. Het betrouwbaarheidsinterval voor κ_+

5.3.1. Drie methodes

In deze paragraaf worden 3 methodes vergeleken, die allen een benaderend interval voor κ_+ geven en die verschillen in symmetrie en waardenbereik voor de gebruikte stochast. Voor het gemak laten we de index '+' weg. Zij $\underline{k}, \hat{\sigma}, \hat{\phi}$ en ξ als in 4.3 gedefinieerd.

METHODE 1: Beschouw k . Het waardenbereik van k is $[-1, 1]$ en voor $n_1 \rightarrow \infty$ en $n_2 \rightarrow \infty$ geldt, $\Pr\left(\frac{k - \kappa}{\hat{\sigma}} < x\right) \rightarrow \phi(x)$.

METHODE 2: Beschouw $\text{arctanh}(k)$. De asymmetrische functie arctanh heeft een waardenbereik $(-\infty, \infty)$. Met behulp van de zogenaamde delta-methode (Bishop et al. (1), section 14.6) is te bewijzen dat voor $n_1 \rightarrow \infty$ en $n_2 \rightarrow \infty$

$$\Pr\left(\frac{\text{arctanh}(\underline{k}) - \text{arctanh}(\kappa)}{\hat{\sigma}} \cdot (1 - \underline{k}^2) < x\right) \rightarrow \phi(x).$$

METHODE 3: Beschouw $\tan \frac{1}{2}\pi k$. De symmetrische functie $\tan$ heeft een waardenbereik $(-\infty, \infty)$. De deltamethode geeft dat voor $n_1 \rightarrow \infty$ en $n_2 \rightarrow \infty$, $\Pr\left(\frac{\tan \frac{1}{2}\pi \underline{k} - \tan \frac{1}{2}\pi \kappa}{\frac{1}{2}\pi \hat{\sigma}} \cdot \cos^2(\frac{1}{2}\pi \underline{k}) < x\right) \rightarrow \phi(x)$.

Vanwege numerieke problemen is een intervalconstructie, gebaseerd op de likelihood-ratio test niet aan te bevelen en derhalve hier buiten beschouwing gebleven.

5.3.2. Een schatting voor de variantie in k_+

Uitgaande van de 2x4-tabel in 4.1. leiden we een uitdrukking af voor de asymptotische variantie $\hat{\sigma}^2$ van k_+ .

Met behulp van de eerder genoemde delta-methode volgt dat $\hat{\sigma}^2 = \frac{\partial k}{\partial t} \hat{\Sigma} \frac{\partial k'}{\partial t}$.

Hierbij is $\hat{\Sigma}$ de geschatte 8x8-covariantiematrix van de twee multinomiale stochasten (a, b, c, d) en (e, f, g, h) en $\frac{\partial k}{\partial t} = d \vec{e} f \left(\frac{\partial k}{\partial a}, \dots, \frac{\partial k}{\partial h} \right)$.

Zij $q_1 = \frac{a+b}{n_1} \frac{n_2}{e+f}$ en $q_2 = \frac{a+c}{n_1} \frac{n_2}{e+g}$ als in 4.3. dan definiëren we eerst

$$\zeta_1 = \frac{\partial q_1}{\partial a} = \frac{\partial q_1}{\partial b} = \frac{1}{e+f} \cdot \frac{n_2}{n_1}$$

$$\zeta_2 = \frac{\partial q_1}{\partial e} = \frac{\partial q_1}{\partial f} = \frac{-(a+b)}{(e+f)^2} \cdot \frac{n_2}{n_1}$$

$$\zeta_3 = \frac{\partial q_2}{\partial a} = \frac{\partial q_2}{\partial c} = \frac{1}{e+g} \cdot \frac{n_2}{n_1}$$

$$\zeta_4 = \frac{\partial q_2}{\partial e} = \frac{\partial q_2}{\partial g} = \frac{-(a+c)}{(e+g)^2} \cdot \frac{n_2}{n_1}$$

Vervolgens definiëren we :

$$\omega_1 = (1-\gamma + \gamma q_1)^{-2} \quad \text{en} \quad \omega_2 = (1-\gamma + \gamma q_2)^{-2}.$$

Met behulp van deze definitie's en de prevalentie γ geldt nu :

$$\frac{\partial k}{\partial a} = \gamma(1-\gamma) (\omega_1 \zeta_1 - \omega_2 \zeta_3)$$

$$\frac{\partial k}{\partial e} = \gamma(1-\gamma) (\omega_1 \zeta_2 - \omega_2 \zeta_4)$$

$$\frac{\partial k}{\partial b} = \gamma(1-\gamma) \omega_1 \zeta_1$$

$$\frac{\partial k}{\partial f} = \gamma(1-\gamma) \omega_1 \zeta_2$$

$$\frac{\partial k}{\partial c} = -\gamma(1-\gamma) \omega_2 \zeta_3$$

$$\frac{\partial k}{\partial g} = -\gamma(1-\gamma) \omega_2 \zeta_4$$

$$\frac{\partial k}{\partial d} = 0$$

$$\frac{\partial k}{\partial h} = 0$$

5.3.3. Een vergelijking tussen de 3 methodes

De drie tweezijdige $(1-\alpha)$ betrouwbaarheidsintervallen voor κ worden vergeleken in feitelijke betrouwbaarheid en 'accuracy'. We vergelijken

$$\text{dus } \Pr_{\kappa_i} \left(\left\{ \frac{k - \kappa_0}{\hat{\sigma}} \right\}^2 \leq \chi_{1-0.5\alpha}^2 \right),$$

$$\Pr_{\kappa_i} \left(\left\{ \frac{\text{arctanh}(\underline{k}) - \text{arctanh}(\kappa_0)}{\hat{\sigma}} \cdot (1 - \underline{k}^2) \right\}^2 \leq \chi_{1-0.5\alpha}^2 \right) \quad \text{en}$$

$$\Pr_{\kappa_1} \left(\left(\frac{\tan \frac{1}{2} \pi \kappa}{\frac{1}{2} \pi \kappa_0} - \tan \frac{1}{2} \pi \kappa_0 \right) \cdot \cos^2 \left(\frac{1}{2} \pi \kappa \right) \right)^2 < \xi_{1-0.5\alpha}^2 \quad ,$$

waarbij voor $i=0$ en $\kappa_0 \in (-1,1)$ de betrouwbaarheid beschouwd wordt en voor $i=1$ met $\kappa_1 \neq \kappa_0$ een uitdrukking voor '1 minus accuracy' gegeven is.

Hiertoe bestuderen we eerst de functie's $f_1(k, \kappa) = (k - \kappa)^2$,

$$f_2(k, \kappa) = (\operatorname{arctanh}(k) - \operatorname{arctanh}(\kappa))^2 \cdot (1 - k^2)^2 \quad \text{en}$$

$$f_3(k, \kappa) = (\tan \frac{1}{2} \pi k - \tan \frac{1}{2} \pi \kappa)^2 \cdot \cos^4 \left(\frac{1}{2} \pi k \right) \cdot (4/\pi^2) \quad .$$

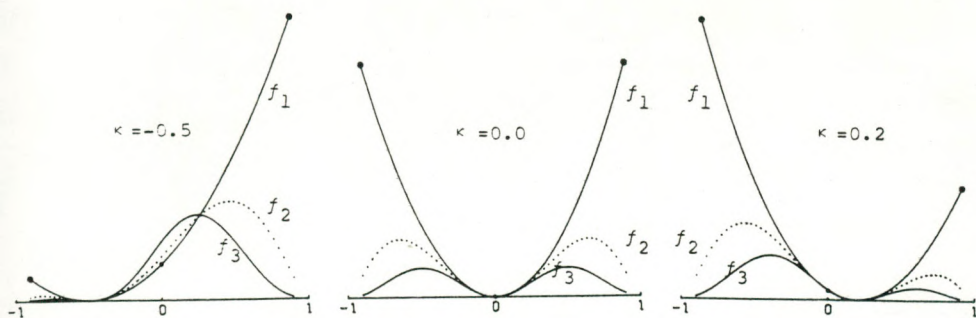
Voor bijna alle $(k, \kappa) \in (-1,1) \times (-1,1)$ blijkt nu te gelden dat hetzij

$$f_1(k, \kappa) \leq f_2(k, \kappa) \leq f_3(k, \kappa), \text{ hetzij } f_3(k, \kappa) \geq f_2(k, \kappa) \geq f_1(k, \kappa) \quad .$$

Dit wordt geïllustreerd in figuur 4 waar voor $\kappa = -0.5; 0.0; 0.2$ en

$k \in (-0.9, 0.9)$ f_1 gegeven wordt door de curve met omcirkelde eindpunten,

f_2 door de gestippelde curve en f_3 door de ononderbroken curve.



Figuur 4: De functie's f_1, f_2, f_3 .

Deze eigenschap suggereert dat de verdelingsfunctie van $f_2(\underline{k}, \kappa)$, een enkele uitzondering daargelaten, tussen die van $f_1(\underline{k}, \kappa)$ en $f_3(\underline{k}, \kappa)$ ligt, hetgeen doet vermoeden dat de betrouwbaarheid en 'accuracy' van methode 2 zich tussen de (onderling wisselende) waarden van de andere twee methoden zal bevinden.

Om dit verder te onderzoeken is de betrouwbaarheid en accuracy berekend voor een aantal combinatie's van relevante parameters. De resultaten, gebaseerd op 5000 simulatie's per parameter-combinatie, bevestigen het hierboven geuite vermoeden.

Daarbij is geconstateerd dat (1) de betrouwbaarheid nogal fluctueert om $1-\alpha$, ook voor methode 2 (zie tabel 1) en dat (2) de accuracy vaak weinig verschilt, maar soms toch een aanzienlijk verschil (0.15) tussen methode 1 en 3 laat zien.

Vanwege het meer stabiele gedrag in betrouwbaarheid en accuracy is methode 2 aanbevolen.

TABEL 1: De betrouwbaarheid in %, voor $n_1=n_2=50$, bij $\alpha=0.05$

(p_a, p_b, p_c, p_d) (p_e, p_f, p_g, p_h)	(0.50, 0.10, 0.35, 0.05)			(0.90, 0.05, 0.01, 0.04)		
	(0.05, 0.10, 0.45, 0.40)			(0.05, 0.30, 0.35, 0.30)		
γ	0.20	0.50	0.80	0.20	0.50	0.80
methode 1	92.94	89.60	85.76	96.50	93.22	91.22
methode 2	94.34	89.74	85.78	96.78	93.34	91.22
methode 3	95.34	90.48	85.80	97.16	93.68	91.28

Dankwoord

Voor de adviezen die de auteur bij de uitvoering en verslaglegging van dit onderzoek mocht ontvangen van dr.P.D.Bezemer, Professor dr.J.Oosterhoff en Professor dr.Chr.L.Rümke is hij hen zeer erkentelijk.

Literatuur

1. Bishop, Y.M.M., Fienberg, S.E., Holland, P.W. (1977) : Discrete Multivariate Analysis, the MIT press Cambridge, Massachusetts and London, England.
2. Feinstein, A.R. (1977) : Clinical Biostatistics, St. Louis, CV Mosby.
3. Katz, D., Baptista, J., Azen, S.P. and Pike, M.C. (1978) : Obtaining confidence intervals for the risk ratio in cohort studies, Biometrics 34, 469-474.
4. Koopman, P.A.R. : Confidence intervals for the ratio of two binomial proportions, (onlangs ter publicatie aangeboden).
5. Lehmann, E.L. (1959) : Testing Statistical Hypotheses, John Wiley & Sons, New York.
6. Ruben, H. (1962) : Probability content of regions under spherical normal distributions IV, AMS 33, 542-570.
7. Vecchio, T.J. (1966) : Predictive value of a single diagnostic test in unselected populations, New England Journal of Medicine 274, 1171-1173.
8. Weinstein, M.C., Fineberg, H.V. (1980) : Clinical Decision Analysis, Saunders Company, Philadelphia.
9. Yerushalmy, J. (1947) : Statistical problems in assessing methods of medical diagnosis, with special reference to X-ray techniques, Public Health Report 62, 1432-1449.