

INTERPRETATIE VAN RESULTATEN VAN SCREENINGSTESTEN TOEGEPAST BIJ HET OPSPOREN VAN RESIDUEN IN EETWAREN VAN DIERLIJKE OORSPRONG

L. Okerman, J. Van Hende, J. Van Hoof, H. De Brabander

Vakgroep Toezicht op Eetwaren, Faculteit Diergeneeskunde, Universiteit Gent, Salisburylaan 133, B-9820 Merelbeke

SAMENVATTING

Voedingsmiddelen kunnen geanalyseerd worden op aanwezigheid van residuen door middel van eenvoudige methoden. In dit overzicht worden de resultaten van dergelijke screeningstesten mathematisch toegelicht. Dit leidt tot de volgende vaststellingen.

1. Het is mogelijk dat resultaten van colorimetrische testen zoals *enzyme immuno assays* (EIA) en van biologische testen zoals bacteriële inhibitietesten vals-positief of vals-negatief zijn. Het aantal foutieve resultaten is afhankelijk van het gehalte van het analyt dat men wil aantonen en van de *cutoff*-waarde. De wijze van aflezen kan bij colorimetrische testen zo gekozen worden dat het aantal vals-positieve of het aantal vals-negatieve resultaten verwaarloosbaar klein wordt.

2. Bij een lage prevalentie geeft zelfs een beperkt aantal vals-positieve resultaten aanleiding tot een lage predictieve waarde; dit wil zeggen dat een groot aantal van de positieve resultaten vals-positief zijn.

3. Indien de positieve resultaten van een test met een relatief lage specificiteit aanleiding geven tot sancties, dan moeten de screeningstesten bevestigd worden, bijvoorbeeld door middel van chemische analyses.

4. Statistieken opgesteld aan de hand van screeningstesten met een suboptimale specificiteit en sensitiviteit zijn een weerspiegeling van een reële situatie. Ze zijn betrouwbaar zonder een chemische bevestiging van positieve resultaten, in tegenstelling tot de individuele testresultaten. De echte prevalentie kan wel afwijken van de schijnbare prevalentie, die rechtstreeks uit de analyseresultaten berekend wordt.

INLEIDING

De analysemethoden gebruikt voor het opsporen van residuen van wettelijk toegelaten geneesmiddelen en van illegale producten die gebruikt worden in de veehouderij kunnen grofweg ingedeeld worden in drie categorieën. De eerste categorie omvat de biologische methoden. De bacteriële inhibitietesten waarmee antibiotica-residuen worden gedetecteerd behoren tot die groep. Vroeger werden ook residuen van hormonen opgespoord door middel van biologische methoden, maar dit is nu niet meer gebruikelijk, omdat ze te weinig betrouwbaar en te weinig gevoelig zijn. Methoden die meerdere of alle leden van een groep scheikundig verwante stoffen aantonen behoren tot de tweede categorie. Immunologische testen zoals *Enzyme linked immunosorbent assay* (ELISA) en *Radio-immuno assay* (RIA) zijn de bekendste (Hoek en Beljaars 1989); ook receptortesten, die couplant gebruikt worden voor antibiotica, horen erbij (Charm en Chi 1988). De gevoelige chemische analyses, zoals *High Performance Thin Layer Chromatography* (HPTLC), *High Performance Liquid Chromatography* (HPLC) en gaschromatografie (GC), al dan niet aangevuld met massaspectrometrie (MS), vormen de derde categorie. De eerste twee categorieën worden dikwijls beschouwd als screenings-testen; de laatste als bevestiging- en kwantificatiemethoden.

Meer dan eens wordt erop gewezen dat "absolute zekerheid" betreffende de aan- of afwezigheid van de gezochte substantie, het analyt, door geen enkele methode kan gegeven worden. M.a.w. geen enkele methode zou werkelijk volledig betrouwbaar zijn. Bij de derde categorie, de chemische methoden, is de kans op ver-

gissingen echter bijzonder klein en bijna verwaarloosbaar wanneer gebruik gemaakt wordt van massaspectrometrie om de identiteit van het analyt te bewijzen (Courtheyn *et al.* 1996). De grootste problemen stellen zich bij de biologische, de immunologische en de receptortesten. Dergelijke testen worden vaak als kit in de handel gebracht door commerciële firma's en soms, maar lang niet altijd, worden gegevens betreffende de betrouwbaarheid op de bijsluiter vermeld.

In het hiernavolgende overzicht wordt de betekenis van een aantal parameters, die de betrouwbaarheid van dergelijke screeningstesten weergeven, besproken. Ook wordt duidelijk gemaakt van welke factoren de eisen, die aan de verschillende categorieën analysemethoden gesteld worden, afhankelijk zijn. Tenslotte wordt besproken wanneer en waarom een screeningsmethode moet gevolgd worden door een chemische identificatiemethode.

BEOORDELING VAN SCREENINGSTESTEN

Definities van sensitiviteit en specificiteit

In Tabel 1 worden een aantal definities gegeven van parameters die de analyses praktijkgericht beoordelen. De parameters zijn van belang bij de interpretatie van de resultaten van commerciële testkits. Meestal zijn dit ELISA testen, doch de parameters gelden bijvoorbeeld ook voor inhibitietesten voor melk, die op de hoeve zelf kunnen uitgevoerd worden. Het ter plaatse testen van melk is een courante praktijk in de USA, zodat de veehouder zelf kan beslissen wanneer de melk opnieuw geschikt is voor consumptie (Cullor, 1993; Cullor *et al.*, 1994). In theorie zijn de besproken parameters ook van

Tabel 1. Definities van parameters nuttig bij de beoordeling van de prestaties van een analysemethode (volgens Cullor, 1993).

Vals-positief resultaat (false-positive result): identificatie van monster afkomstig van een niet-behandeld dier als test-positief.
Vals-negatief resultaat (false-negative result): identificatie van een monster dat residu bevat in hoeveelheden hoger dan de toegelaten limiet als test-negatief.
Epidemiologische specificiteit (epidemiologic specificity): de waarschijnlijkheid waarmee echte negatieve monsters juist geïdentificeerd worden.
Epidemiologische gevoeligheid (epidemiologic sensitivity): de waarschijnlijkheid waarmee echte positieve monsters juist geïdentificeerd worden.
Laboratorium-specificiteit (laboratory definition of specificity): het vermogen residuen van de opgespoorde substantie te onderscheiden van residuen van andere (meestal chemisch verwante) stoffen.
Laboratorium-gevoeligheid (laboratory definition of sensitivity): de detectielimiet van de op te sporen stof met de betreffende methode.
Betrouwbaarheid (reliability): de geloofwaardigheid van de test als middel om het residu op te sporen.
Prevalentie (prevalence): de proportie van de populatie (i.c. slachtdieren) waarbij het residu werkelijk aanwezig is.
Predictieve waarde/positieve test of PW/+ (predictive value/positive test; PV(+)): de waarschijnlijkheid dat een test-positief monster werkelijk het op te sporen residu bevat in hoeveelheden die hoger zijn dan toegelaten.
Predictieve waarde/negatieve test of PW/- (predictive value/negative test; PV(-)): de waarschijnlijkheid dat een test-negatief monster het op te sporen residu niet bevat in hoeveelheden die hoger zijn dan toegelaten.

toepassing op inhibitietesten voor vlees en nieren, maar in de praktijk kent men de gevoeligheid en de specificiteit (zie verder) van deze testen niet.

Tabel 1 maakt ook duidelijk dat sensitiviteit meerdere begrippen dekt, net zoals specificiteit verschillende betekenissen kan hebben. De definitie van sensitiviteit die in de chemische analyse de voorkeur geniet, is: de wijze waarop de analyse reageert op kleine veranderingen in concentratie van het analyt. De sensitiviteit is in deze gevallen evenredig met de helling van de ijkcurve van een chemische analyse. In andere gevallen wordt met sensitiviteit echter de *detectielimiet* bedoeld, dus de kleinste hoeveelheid van het analyt die nog kan aangetoond worden met een bepaalde zekerheid. In het hiernavolgende betoog zal deze laatste sensitiviteit altijd de *laboratoriumsensitiviteit of -gevoeligheid* genoemd worden. De detectielimiet is ook niet hetzelfde begrip als de beslissingslimiet. Een *beslissingslimiet* is, in het geval van residuanalyse, het gehalte aan residu boven hetwelke men het voedingsmiddel afkeurt. In het geval van toegelaten diergeneesmiddelen is dat bijvoorbeeld de *maximale residu-limiet* (MRL), opgelegd door de regelgeving van de Europese Unie. De beslissingslimiet voor niet toegelaten producten zal men zo laag mogelijk nemen.

Anderzijds betekent de specificiteit bij chemische analyses het vermogen om de gezochte stof te onderscheiden van andere verwante stoffen. Deze eigenschap wordt verder omschreven als laboratoriumspecificiteit.

Van sommige screeningstesten kunnen gegevens zoals sensitiviteit en specificiteit in de literatuur, of op de bijsluiters van commerciële kits, teruggevonden worden. Wanneer bij sensitiviteit de detectielimiet vermeld wordt, dan bedoelt men de laboratoriumsensitiviteit;

bij specificiteit worden soms een aantal verwante chemische producten vermeld met een percentage kruisreacties, en dit geeft de laboratoriumspecificiteit weer. Wanneer sensitiviteit en specificiteit echter weergegeven worden als percentages of als getallen tussen 0 en 1, zonder eenheden, dan wordt hiermee niet de laboratoriumsensitiviteit of -specificiteit bedoeld doch de epidemiologische sensitiviteit of specificiteit. De *epidemiologische sensitiviteit* geeft de waarschijnlijkheid weer waarmee echte positieve monsters juist geïdentificeerd worden, de *epidemiologische specificiteit* de waarschijnlijkheid waarmee echte negatieve monsters juist geïdentificeerd worden.

De epidemiologische gevoeligheid (of sensitiviteit) kan men ook mathematisch uitdrukken als de verhouding tussen de correct positief reagerende monsters en het totaal aantal positieve monsters (Tabel 2). Wanneer bijvoorbeeld de sensitiviteit 99 % bedraagt, dan betekent dit dat op 100 monsters die het residu bevatten, 99 zullen positief reageren en 1 negatief. Meteen betekent dit dat het aantal vals-negatieve reacties in dat geval 1 % zal bedragen.

Op dezelfde manier kan men de epidemiologische specificiteit uitdrukken als de verhouding van het aantal correct negatief reagerende monsters tot het totaal aantal negatieve. De relatie tussen specificiteit en aantal vals-positieve reacties, wanneer uitsluitend monsters worden onderzocht die het residu niet bevatten, is evident.

De epidemiologische sensitiviteit en specificiteit zijn voor de praktijk interessantere parameters dan de laboratorium-gevoeligheid en -specificiteit. De aanvragers van de analyse verlangen immers dat het laboratorium een betrouwbaar onderscheid kan ma-

Tabel 2. Mathematische voorstelling van de epidemiologische specificiteit en sensitiviteit, de voorspellende waarden en schijnbare en echte prevalenties (Martin *et al.*, 1987).

Testresultaat	Werkelijke toestand	
	Residu aanwezig (R+)	Residu afwezig (R-)
Aantal positief (T+)	a	b
Aantal negatief (T-)	c	d
Totaal (a+b+c+d=n)	a+c	b+d

Epidemiologische gevoeligheid : $a/(a+c)$

Epidemiologische specificiteit: $d/(b+d)$

Voorspellende waarde van een positief resultaat: $a/(a+b)$

Voorspellende waarde van een negatief resultaat: $c/(c+d)$

Schijnbare prevalentie: $(a+b)/n$ of $(T+)/n$

Werkelijke prevalentie: $(a+c)/n$

ken tussen monsters die aan een bepaalde voorwaarde voldoen, en monsters die daar niet aan voldoen. Epidemiologische specificiteit is ook een veel breder begrip dan laboratoriumspecificiteit: de chemische specificiteit van bacteriële inhibitietesten is bijvoorbeeld onbestaande, want de remming kan door om het even welk antibioticum veroorzaakt worden, doch de epidemiologische specificiteit om antibiotica in het algemeen op te sporen kan zeer hoog zijn, afhankelijk van de test en de wijze van aflezen.

Invloed van de wijze van aflezen op sensitiviteit en specificiteit

Voor de residu-analyse maakt men dikwijls gebruik van competitieve *Enzyme Linked Immuno-sorbent Assays* (ELISA). Dergelijke bepalingen worden colorimetrisch afgelezen; blanco's (die het residu niet bevatten) geven in de regel hogere optische densiteit (OD-) waarden dan positieve monsters. De gebruiker moet een bepaalde OD-waarde als *cutoff* waarde beschouwen: hogere OD-waarden worden afgelezen als negatief, lagere als positief. De keuze van de *cutoff* waarde is afhankelijk van de vereiste specificiteit en sensitiviteit (Golden, 1991).

Jackman (1996) en Everest *et al.* (1993, 1994) beschrijven een manier om zelf een *cutoff* waarde in te stellen, naargelang de doelstelling van de gebruiker. De door hen gevolgde methode laat toe vooraf te berekenen welke sensitiviteit en specificiteit bij de gekozen afleeswaarde mogen verwacht worden. Telkens worden een representatief aantal monsters onderzocht die het residu niet bevatten, en een aantal belaste monsters, d.w.z. monsters waaraan het analyt werd toegevoegd. De hoeveelheid analyt die toegevoegd wordt is afhankelijk van de vereiste laboratoriumgevoeligheid.

De resultaten van de colorimetrische testen worden verwerkt in waarschijnlijkheidscurven; daarbij wordt de B/B^0 waarde (optische densiteit monster/optische densiteit negatieve controle), of de procentuele B/B^0 waarde ($B/B^0 \times 100$) in ordinaat uitgezet ten opzichte van de cumulatieve waarschijnlijkheid, uitgedrukt in de normale probabiliteitsschaal, dat het monster deze B/B^0 waarde vertoont. Bij een normale verdeling van

de B/B^0 waarden wordt met deze schaal zowel voor de negatieve monsters als voor de belaste monsters een rechte lijn bekomen. Een horizontale lijn door een willekeurige B/B^0 waarde geeft het te verwachten aantal vals-positieve en vals-negatieve resultaten, wanneer deze waarde als *cutoff* waarde zou gebruikt worden. Op de kruispunten van de horizontale lijn met de twee rechten kan immers de waarschijnlijkheid afgelezen worden dat een negatief monster positief reageert, alsook de waarschijnlijkheid dat een positief monster negatief reageert.

Figuur 1 geeft dergelijke waarschijnlijkheidscurven van een commerciële ELISA test voor het opsporen van oxytetracycline residuen in vlees, die door ons werd gevalideerd. De rechten geven de cumulatieve waarschijnlijkheid weer dat blanco monsters en monsters belast met resp. 100 $\mu\text{g/kg}$ oxytetracycline en 100 $\mu\text{g/kg}$ doxycycline (gelijk aan de MRL voor vlees) de overeenstemmende B/B^0 waarden vertonen. De B^0 waarde is de OD waarde van het extractiemiddel (fosfaatbuffer). Met deze ELISA test is ook de B/B^0 waarde van blanco monsters meestal lager dan 1. Wanneer uitsluitend rekening gehouden wordt met oxytetracycline, kan een B/B^0 waarde van 0,63 of 63 % gebruikt worden als *cutoff* waarde, waarbij zowel de kans op vals-positieve als op vals-negatieve resultaten kleiner is dan 1 op 100. Bij deze *cutoff* waarde wordt echter slechts 20 % van de monsters die 100 $\mu\text{g/kg}$ doxycycline bevatten positief afgelezen. Bij een B/B^0 waarde van 0,80 reageert slechts 7 % van deze laatste groep negatief, doch dan bedraagt het aantal vals-positieve reacties 16 %.

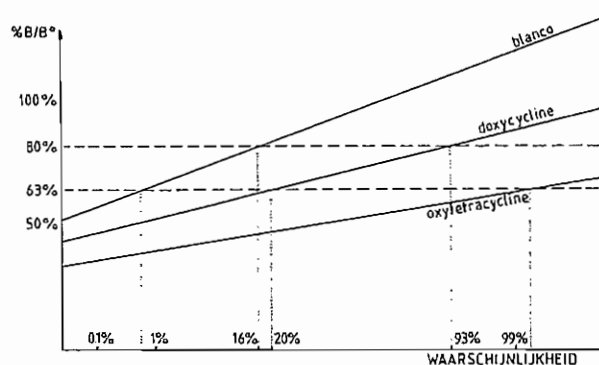


Fig. 1. Cumulatieve waarschijnlijkheidscurven van monsters belast met 100 $\mu\text{g/kg}$ oxytetracycline of met 100 $\mu\text{g/kg}$ doxycycline en van blanco monsters, bij gebruik van een commercieel verkrijgbare competitieve ELISA test. Aflezing bij een B/B^0 waarde van 0,80 geeft 7 % vals-negatieve resultaten voor monsters met doxycycline en 16 % vals-positieve resultaten. Aflezing bij een B/B^0 waarde van 0,63 geeft minder dan 1 % vals-positieve resultaten en detecteert meer dan 99 % van de monsters met oxytetracycline doch slechts 20 % van de monsters met doxycycline.

Op dergelijke waarschijnlijkheidscurven kan men ook zien dat bij om het even welke *cutoff* waarde vals-positieve zowel als vals-negatieve resultaten in theorie mogelijk zijn. Enkele fictieve voorbeelden zijn te vinden in figuren 2a en 2b. Wanneer de waarschijnlijk-

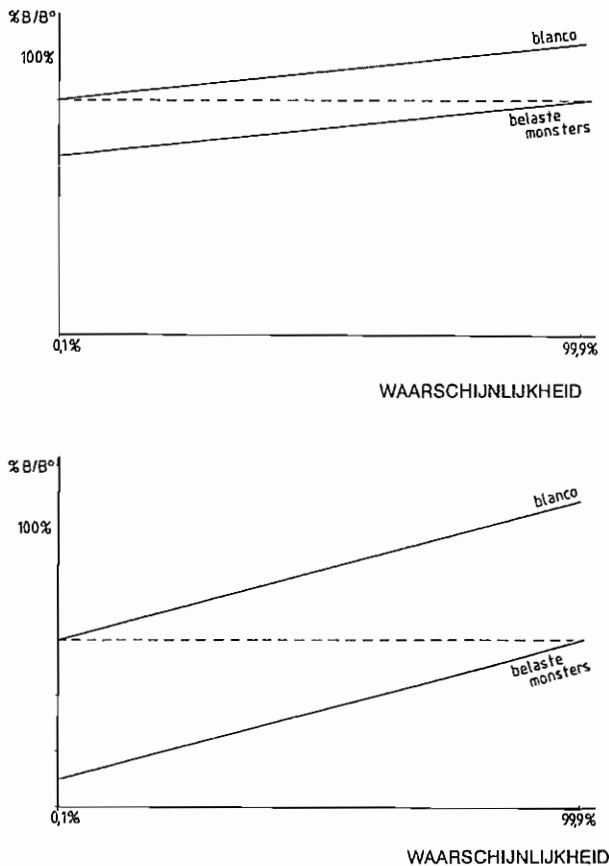


Fig. 2a en 2b. Fictieve voorbeelden van waarschijnlijkheidscurven van blanco en belaste monsters waarbij de variatie tussen de monsters onderling gering is (Fig. 2a) of waarbij de laboratoriumgevoeligheid veel lager is dan het gehalte dat men wil aantonen. In het eerste geval vertonen de curven een geringe helling, in het tweede geval liggen ze ver uit elkaar. In beide gevallen kan een *cutoff* waarde ingesteld worden waarbij de kans op vals-positieve en op vals-negatieve resultaten aanvaardbaar is.

heidscurven van zowel de negatieve als de positieve monsters een geringe helling vertonen (dus wanneer de variatie tussen gelijkwaardige monsters onderling gering is) (Fig. 2a) en/of ver uit elkaar liggen (dit is wanneer de laboratoriumgevoeligheid van de test veel lager is dan de beslissingslimiet) (Fig. 2b), dan kan de B/B° waarde zo gekozen worden dat de snijpunten van de horizontale lijn met de curven voor positieve en negatieve monsters voldoende ver van het middelpunt liggen, zodat de kans op vals-positieve en vals-negatieve resultaten bijvoorbeeld kleiner wordt dan één op een miljoen. Meestal is dit niet het geval bij dergelijke kits en moet men de keuze maken tussen meer vals-positieve of meer vals-negatieve resultaten.

In bovenstaand voorbeeld werd vergeleken hoe een testkit reageert op twee verschillende, chemisch verwante residuen. Op dezelfde manier kan men een verschillende *cutoff* waarde instellen naargelang de concentratie van het analyt die men wil aantonen. Jackman (1996) vergeleek de prestaties van een EIA test voor clenbuterol voor monsters waaraan respectievelijk 1ng/g en 0,25ng/g van het analyt was toegevoegd.

Deze test gaf minder dan 0,01 % vals-positieve en vals-negatieve resultaten voor de monsters met 1ng/g clenbuterol. Wanneer de vereiste laboratoriumsensitiviteit 0,25ng/g bedroeg werd de kit afgelezen bij een hogere B/B° waarde en gaf 1 % vals-positieve en 6 % vals-negatieve resultaten.

Een tweede voorbeeld uit de literatuur betreft een enzymatische analyse voor de determinatie van β -lactam antibiotica, die in de handel gebracht wordt als kit onder de naam Penzym (UCB) (Frère *et al.* 1980). Ook deze testmethode is competitief en geeft dus de hoogste optische densiteiten bij monsters die het residu niet bevatten. Everest *et al.* (1993, 1994) berekenden met waarschijnlijkheidscurven dat de analyse met het oorspronkelijke substraat van de β -lactam testkit, bij een optimale keuze van aflezing, aanleiding gaf tot 17 % vals-negatieve en 5 % vals-positieve resultaten. Vervanging van het oorspronkelijke substraat resulteerde in een sterke verbetering van de prestaties van de kit: het aantal vals-positieve en vals-negatieve resultaten bedroeg dan resp. 3 % en 1 %.

Deze drie voorbeelden illustreren dat de *cutoff* waarde kan aangepast worden naargelang de vereisten die aan de analyse gesteld worden. Indien het de bedoeling is om de aanwezigheid van het analyt verder chemisch aan te tonen, dan verlangt men zo weinig mogelijk vals-negatieve resultaten, doch een beperkt aantal vals-positieve resultaten worden getolereerd. Indien anderszits de resultaten van de analyse rechtstreeks aanleiding geven tot sancties tegen de producent dan moeten vals-positieve resultaten zoveel mogelijk uitgesloten zijn. Het zou handig zijn als producenten van commerciële kits dergelijke waarschijnlijkheidscurven op de bijsluiter zouden weergeven, zodat het laboratorium de *cutoff* waarde zelf kan kiezen in functie van de toepassing.

Het aanpassen van de *cutoff* waarde is één manier om het aantal vals-positieve resultaten tot een minimum te beperken. Een andere manier bestaat erin om twee screeningstesten, die niet gebaseerd zijn op hetzelfde principe, na elkaar toe te passen (Martin *et al.* 1987). Wanneer beide screeningstesten bijvoorbeeld 99 % specifiek zijn dan krijgt men $1\% \times 1\% = 0,01\%$ vals-positieve resultaten.

Invloed van prevalentie op betrouwbaarheid van positieve en negatieve resultaten: berekening van positieve en negatieve voorspellende waarden

Met prevalentie wordt bedoeld: de verhouding van het aantal monsters die het residu werkelijk bevatten ten opzichte van het totaal aantal monsters (Tabel 1). Dit gegeven is bij residu-analyse in de meeste gevallen uiteraard niet bekend. Wanneer exact geweten is hoeveel de epidemiologische sensitiviteit en specificiteit bedragen dan kan uit de resultaten van de analyse op een groot aantal monsters de werkelijke prevalentie berekend worden (zie verder).

In Tabel 1 worden ook nog de voorspellende of predictieve waarden (predictive values) voor positieve en negatieve testen, $PW(+)$ en $PW(-)$, vermeld. Deze kunnen worden uitgedrukt in % (Golden 1991) of als een waarde tussen 0 en 1 (Van Eenennaam *et al.* 1993). Optimaal bedraagt de predictieve waarde 1 of 100 %.

Uit de epidemiologische gevoeligheid en specificiteit enerzijds en de prevalentie anderzijds kan men de $PW(+)$ en $PW(-)$ van een test berekenen. Deze parameters geven immers het aandeel weer van de echte positieve (resp. negatieve) binnen de groep van de positief (resp. negatief) reagerende monsters. Twee voorbeelden worden hieronder uitgewerkt:

Voorbeeld 1

Eigenschappen van de test: gevoeligheid = 98 % (dus 2 % vals-negatieve); specificiteit = 96 % (dus 4 % vals-positieve).

Bij een prevalentie van 50 % bekomt men:

% positieve testen: 49 % + 2 % vals-positieve = 51 %.

% negatieve testen: 48 % + 1 % vals-negatieve = 49 %.

$PW(+)$ = 49 %/51 % = 0,96 of 96 %; slechts 4 % van de positieve resultaten is vals.

$PW(-)$ = 48 %/49 % = 0,98 of 98 %; slechts 2 % van de negatieve resultaten is vals.

Doch wanneer de prevalentie 10 % bedraagt

% positieve testen: 9,8 % + 3,6 % vals-positieve = 13,4 %.

% negatieve testen: 86,4 % + 0,2 % vals-negatieve = 86,6 %.

$PW(+)$ = 9,8 %/13,4 % = 0,73 of 73 %; 27 % van de positieve resultaten is vals.

$PW(-)$ = 86,4 %/86,6 % = >0,99 of >99 %; minder dan 1 % van de negatieve resultaten is vals.

Bij een prevalentie van 2 % bekomt men het volgende:

% positieve testen: 1,96 % + 3,92 % = 5,88 %.

% negatieve testen: 94,08 % + 0,04 % = 94,12 %.

$PW(+)$ = 1,96 %/5,88 % = 0,33 of 33 %; 67 % van de positieve resultaten is vals.

$PW(-)$ = 94,08 %/94,12 % = >0,99 of >99 %; bijna alle negatieve resultaten zijn juist.

Voorbeeld 2

Een testkit die 99 % specifiek en 99 % sensitief is, zal bij een prevalentie van 10 % de volgende resultaten geven:

% positieve testen: 9,9 % + 0,9 % vals-positieve = 10,8 %.

% negatieve testen: 89,1 % + 0,1 % vals-negatieve = 89,2 %.

$PW(+)$ = 9,9 %/10,8 % = 0,92 of 92 %; 8 % van de positieve resultaten is vals.

$PW(-)$ = 89,1 %/89,2 % = >0,99 of >99 %; bijna alle negatieve resultaten zijn juist.

Bij suboptimale specificiteit en een lage prevalentie zal een relatief groot aandeel van de positieve testen dus vals-positief zijn. Het totaal aantal uitgevoerde testen speelt geen rol. In het eerste voorbeeld zijn bij prevalenties lager dan 10 % reeds meer dan 1 op 4 van de positieve resultaten vals; bij prevalenties lager dan 2 % zijn meer dan 2 op 3 van de positieve resultaten vals. In het tweede voorbeeld, bij een specificiteit van 99 % en een prevalentie van 10 %, zijn nog altijd 8 % of bijna 1 op 10 van de positieve resultaten vals. In de praktijk komen lage prevalenties veel meer voor dan

hoge. Dit betekent dat positieve resultaten bijna altijd zullen moeten bevestigd worden, wil men er diagnostische waarde aan hechten.

Wanneer de specificiteit en de sensitiviteit van een analysemethode gekend zijn, dan kunnen de $PW(+)$ en $PW(-)$ waarden in curven worden uitgezet in functie van de prevalentie. De vorm van deze curven geeft de betrouwbaarheid van de methode duidelijk weer. Twee voorbeelden worden uitgewerkt in figuren 3 en 4. Nemen we een voorbeeld van een test die 5 % vals-positieve en 17 % vals-negatieve resultaten geeft, zoals de hoger besproken enzymatische methode voor het aantonen van β -lactam antibiotica. Op figuur 3 kunnen de $PW(+)$ en $PW(-)$ waarden bij iedere mogelijke prevalentie afgelezen worden. Het tweede voorbeeld (Fig. 4) is de verbeterde versie van dezelfde test, waardoor het aantal vals-positieve en vals-negatieve resultaten gereduceerd werden tot respectievelijk 3 % en 1 %. Bij deze tweede curve naderen zowel de $PW(+)$ en $PW(-)$ curven de ideale waarden van 100 % bij lage resp. hoge prevalenties.

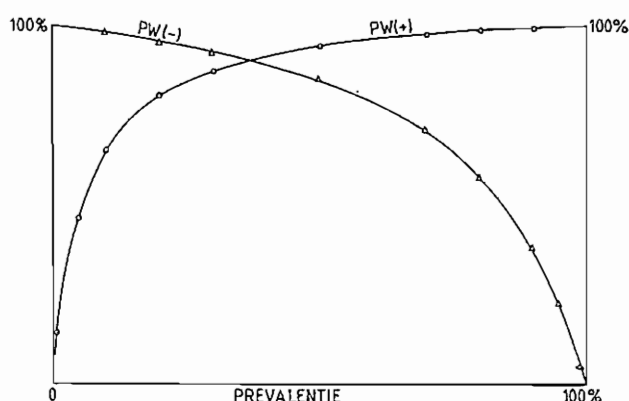


Fig. 3. $PW(+)$ en $PW(-)$ curven van een enzymatische test, geschikt voor het opsporen van β -lactam antibiotica, die 83 % sensitief is en 95 % specifiek bij toepassing op nierweefsel (Everest *et al.* 1994). De $PW(+)$ en $PW(-)$ waarden kunnen bij elke mogelijke prevalentie afgelezen worden.

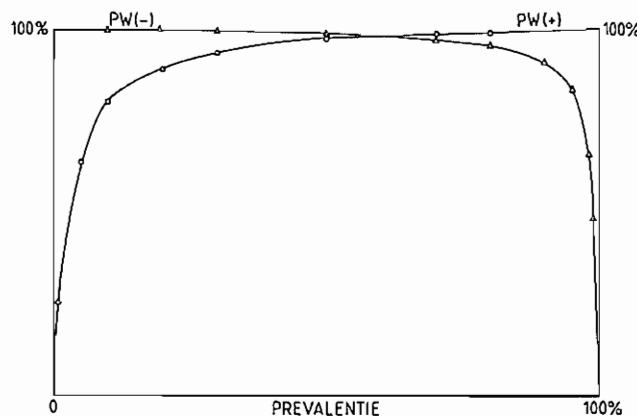


Fig. 4. $PW(+)$ en $PW(-)$ curven van dezelfde methode wanneer een ander substraat gebruikt wordt, waardoor de sensitiviteit en specificiteit resp. 99 % en 97 % bedragen. In vergelijking met figuur 2 benaderen de $PW(+)$ en $PW(-)$ waarden dicht bij de optimale waarden bij lage resp. hoge prevalenties.

Praktische toepassing: beoordeling van commerciële kits voor analyse van melk op aanwezigheid van antibiotica

Van Eenennaam *et al.* (1993) pasten deze theoretische gegevens toe op een aantal commerciële kits die geschikt zijn voor analyse van melk op aanwezigheid van antibiotica. Deze kits werden door hen getest op melk afkomstig van drie groepen van dieren met klinische mastitis. Twee van de drie groepen waren elk behandeld met een antibioticum van de β -lactam familie; de controlegroep was behandeld met oxytocine. In deze omstandigheden bleken sommige testkits zeer veel vals-positieve resultaten te geven, zodat ze in de praktijk niet bruikbaar beschouwd werden voor melk afkomstig van dieren met mastitis. De enige geteste kit die op ELISA gebaseerd was, nl. LacTek® (β -lactam) (Idetek) was ook de enige die in de gegeven omstandigheden relatief goed presteerde.

De betrouwbaarheid van de kits werd weergegeven in PW(+) en PW(-) curven (Van Eenennaam *et al.* 1993). De drie inhibitietesten gaven een bijna perfecte PW(-) curve en een veel minder goede (voor Delvotest P®, Gist-Brocades en Bacillus stearothermophilus disc assay) tot waardeloze (Charm Farm®, Charm Sciences lim.) PW(+) curve; met andere woorden ze gaven zeer veel vals-positieve resultaten. Voor de CITE® probe (β -lactam) test (Idexx) benaderden noch de PW(+) noch de PW(-) curve de optimale vorm.

Beide parameters, PW(+) en PW(-), zijn van groot praktisch belang. In de melkanalyse bijvoorbeeld betekent een lage PW(+) dat een groot aantal melkmonsters ten onrechte zal afgekeurd worden; bij een lage PW(-) komt dikwijls melk in de handel die niet voldoet aan de wettelijke limiet. Daarom raden Cullor *et al.* (1994) aan om bij de validatie van testkits voor melkanalyse, speciale aandacht te besteden aan vals-positieve en vals-negatieve uitslagen. Om vals-positieve resultaten uit te sluiten moet de kit o.a. ook getest worden met antibioticum-vrije melk afkomstig van dieren met mastitis, en met melk waaraan natuurlijke inhibitoren, zoals lactoferrine, werden toegevoegd.

Dergelijk vergelijkend onderzoek is tot nu toe nog niet uitgevoerd op testkits of testmethoden die gebruikt worden voor de analyse van vlees op aanwezigheid van residuen. De specificiteit en sensitiviteit zijn meestal niet bekend voor de testkits, wanneer ze gebruikt worden op een andere matrix dan deze waarvoor ze oorspronkelijk ontworpen werden.

Berekening van de echte prevalentie

De vorige redeneringen en berekeningen zijn theoretisch: de echte prevalentie is niet bekend in de meeste gevallen. Omgekeerd kan men de echte prevalentie berekenen uit de analyseresultaten, wanneer de sensitiviteit en de specificiteit van de test bekend zijn. Dit wordt toegepast bij de interpretatie van screeningstesten die subklinische ziekte-toestanden opsporen (Martin *et al.*, 1987), en is eveneens geldig voor het onderzoek op residuen. Bij deze berekening maakt men gebruik van de volgende formule:

$$p(R+) = \frac{p(T+) - p(T+/R-)}{1 - [p(T+/R-) + p(T-/R+)]}$$

waarbij: $p(R+)$ = werkelijke proportie residu-positieve dieren;

$p(T+)$ = proportie testpositieve dieren;

$p(T+/R-)$ = proportie vals positieve dieren of 1 - specificiteit;

$p(T-/R+)$ = proportie vals negatieve dieren of 1 - sensitiviteit.

Een voorbeeld wordt hierna uitgewerkt. Stel dat een bacteriële inhibitietest, die antibiotica-residuen opspoor, 96 % sensitief is en 96 % specifiek. Wanneer men 5 % positieve resultaten bekomt betekent dit dat de werkelijke prevalentie $(0,05-0,04)/[1-(0,04+0,04)] = 0,01$ of 1 % bedraagt. Wanneer men 15 % positieve resultaten vindt bekomt men voor de werkelijke prevalentie $(0,15-0,04)/[1-(0,04+0,04)] = 0,12$ of 12 %. Opnieuw is de afwijking het grootst bij lage schijnbare en echte prevalenties.

Deze cijfers maken wel duidelijk dat stijgingen of dalingen van de aantallen positieve resultaten van screeningstesten, die altijd op dezelfde manier uitgevoerd worden, niet te wijten kunnen zijn aan vals-positieve resultaten, doch werkelijk een weerspiegeling zijn van een gewijzigde toestand in het te onderzoeken materiaal. Zelfs indien een screeningstest minder dan 99 % specifiek en minder dan 99 % sensitief is, dan nog kan die test toegepast worden op grote aantallen monsters, bijvoorbeeld om statistieken op te stellen of om vergelijkingen te maken tussen verschillende groepen monsters.

LITERATUUR

- Charm S.E., Chi R. (1988). Microbial receptor assay for rapid detection and identification of seven families of antimicrobial drugs in milk; collaborative study. *Journal of the Association of Official Analytical Chemists* 71, 304-316.
- Courthey D., De Brabander H., Vercammen J., Batjoens P., Logghe M., De Wasch K. (1996). MS-MS: Applications and tentative quality criteria. *Proceedings of the Euroresidue III Conference, Veldhoven, Nederland, 6-8 mei 1996*, pp.75-85.
- Cullor J.S. (1993). Antibiotic residue tests for mammary gland secretions. *Veterinary Clinics of North America: Food Animal Practice* 9, 609-620.
- Cullor J.S., Van Eenennaam A., Gardner I., Perani L., Dellinger J., Smith W.L., Thompson T., Payne M.A., Jensen L., en Guterbock W.M. (1994). Performance of various tests used to screen antibiotic residues in milk samples from individual animals. *Journal of the Association of the Official Analytical Chemist* 77, 862-870.
- Everest D.J., Everest S.J., Jackman R. (1993). Spectrophotometric assay for β -lactam residues in kidney tissue. *Analytica Chimica Acta* 275, 249-252.
- Everest D.J., Jackman R., Thorne L., Everest S.J. (1994). Improved spectrophotometric assay for β -lactam residues in kidney tissue. *Analyst* 119, 2727-2729.
- Frère J-M., Klein D., Ghuysen J-M. (1980). Enzymatic method for rapid and sensitive determination of β -lactam antibiotics. *Antimicrobial Agents and Chemotherapy* 18, 506-510.

- Golden C. (1991). Overview of the state of the art of immunoassay screening tests. *Journal of the American Veterinary Medical Association* 198, 827-830.
- Hoek A.C., Beljaars P.R. (1989). ELISA's in levensmiddelen: een literatuuroverzicht. *De ware(n)chemicus* 19, 37-68.
- Jackman R. (1996). Antimicrobials: a strategy for linking screening and confirmatory methods. Proceedings of the Euroresidue III Conference, Veldhoven, Nederland, 6-8 mei 1996, pp. 99-112.
- Martin S.W., Meek A.H., Willeberg P. (1987). Measurement of Disease Frequency and Production. Chapter 3 in: *Veterinary Epidemiology - Principles and Methods*. Iowa State University Press, Ames, USA.
- Van Eenennaam A.L., Cullor J.S., Perani L., Gardner A., Smithe W.L., Dellinger J., Guterbock W.M. (1993). Evaluation of milk antibiotic residue screening tests in cattle with naturally occurring clinical mastitis. *Journal of Dairy Science* 76, 3041-3053.
-