



Sensitiviteit, specificiteit en COVID-19

Testen wordt genoemd als een van de middelen om het verspreiden van het SARS-CoV-2 virus te stuiten. Hoe effectief medische testen zijn hangt nauw samen met de sensitiviteit, specificiteit en prevalentie van het virus in de geteste populatie. Alleen die prevalentie in testgroepen is vanuit beleid en testadvies te beïnvloeden. Afhankelijk van de testuitslagen is veel patiënten in de testgroep opnemen wenselijk of juist niet.

ANTON MEIJBURG

Nederland is nu een jaar de ban van het SARS-CoV-2 virus, dat de ziekte COVID-19 veroorzaakt, en het stoppen ervan. Naast vaccins wordt ook testen veelvuldig genoemd als manier om de verspreiding te stuiten. Het percentage positieve uitslagen is niet gelijk aan het percentage COVID-19-patiënten in je testgroep. Vanuit individueel perspectief is de kans dat de testuitslag juist is ook niet constant. Prevalentie, sensitiviteit en specificiteit beïnvloeden deze kansen. Dit komt in de media nauwe-

lijks aan bod. Reden genoeg om het testen nader te bekijken.

Sensitiviteit, specificiteit en prevalentie

Een fout-positieve uitslag noemt de statistiek een type I fout, een fout-negatieve uitslag een type II fout. Bij medisch-diagnostische tests spreekt men vaak over de ver-

wante begrippen 'sensitiviteit' en 'specificiteit'. Wanneer een uitslag positief is, heeft de patiënt in deze casus COVID-19 :

Sensitiviteit: Het **aantal juist-positieve** testuitslagen **als percentage van alle COVID-19-patiënten**

Specificiteit: Het **aantal juist-negatieve** testuitslagen **als percentage van alle niet COVID-19-patiënten**.

In het document *Over de betrouwbaarheid van de PCR-test voor SARS-CoV-2* geeft het RIVM aan dat de sensitiviteit van de testen tussen de 67% en 98% ligt. Dat is nogal brede schatting die samenhangt met de hoeveelheid virusuitscheiding, moment van monsterafname en de kwaliteit van het monster. De specificiteit is preciezer bekend, die ligt tussen de 96,0% en 99,5%. De prevalentie is het percentage COVID-19-patiënten in de testgroep.

Rekenvoorbeeld week 2 2021

Op de website van het RIVM is te zien hoeveel testen wekelijks uitgevoerd worden en wat de uitkomsten zijn. In week 2 lieten 294.712 mensen zich testen en daaruit kwamen 32.535 positieve resultaten. Maar wat zegt dit nu precies? Uitgaande van de hoogste schatting in sensitiviteit en specificiteit (respectievelijk 98% en 99,5%) leidt dit tot de cijfers in tabel 1, waarbij EC staat voor 'Echt COVID-19' en EN voor 'Echt Niet COVID-19':

We zien 2 vergelijkingen met EC en EN als onbekenden, oplossen geeft de cijfers die tabel 2 laat zien.

Met de laagste schattingen sensitiviteit en specificiteit (respectievelijk 67% en 96,0%) zijn de cijfers zoals weergegeven in tabel 3.

Met deze testuitslagen loopt de schatting voor het totaal aantal werkelijke patiënten in de groep niet heel erg uiteen. Hoeveel patiënten juist geïdentificeerd worden wel.

	Uitslag positief	Uitslag negatief	Totaal
COVID-19	0,98*EC	0,02*EC	0,98*EC+0,02*EC
Geen COVID-19	0,005*EN	0,995*EN	0,005*EN+0,995*EN
Totaal	32.535	262.177	294.712

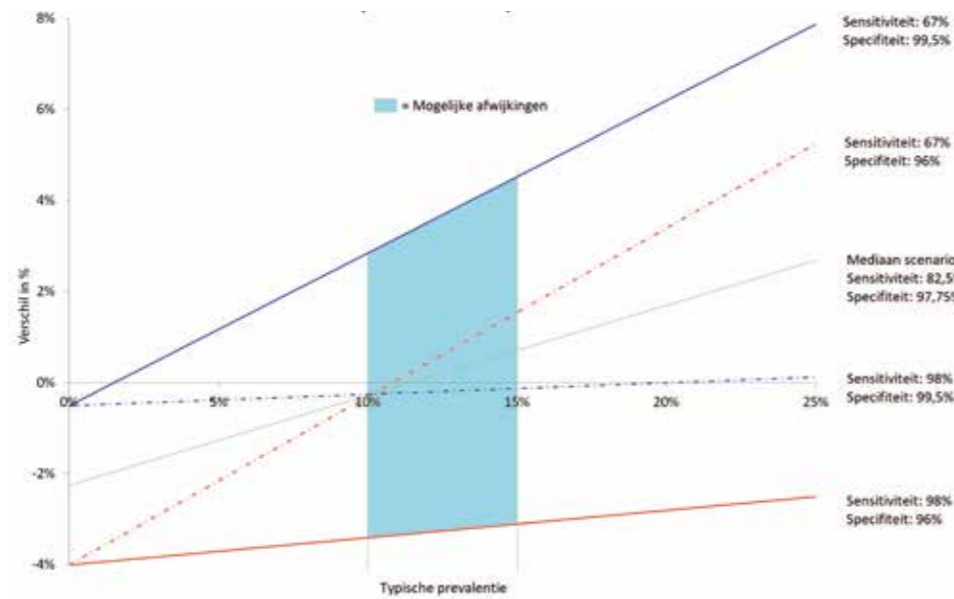
Tabel 1. Berekeningen week 2 met hoogste sensitiviteit en specificiteit

	Uitslag positief	Uitslag negatief	Totaal
COVID-19	31.221	637	31.858
Geen COVID-19	1.314	261.540	262.854
Totaal	32.535	262.177	294.712

Tabel 2. Uitkomsten week 2 met hoogste sensitiviteit en specificiteit

	Uitslag positief	Uitslag negatief	Totaal
COVID-19	22.064	10.867	32.931
Geen COVID-19	10.471	251.310	261.781
Totaal	32.535	262.177	294.712

Tabel 3. Uitkomsten week 2 met laagste sensitiviteit en specificiteit



Figuur 1. Verschil prevalentie en percentage positief testresultaat

De schattingsfout in prevalentie

Het percentage positieve uitslagen ligt doorgaans tussen de 10% en 15% in de testgroepen maar niet elke patiënt krijgt dus een positieve uitslag. Het verschil tussen het aantal positieve uitslagen als percentage van het totaal aantal uitslagen en de werkelijke prevalentie in de testgroep staat in figuur 1 voor verschillende scenario's voor sensitiviteit en specificiteit. Het mediane scenario zit qua sensitiviteit en specificiteit in het midden van de schattingen, op respectievelijk 82,5% en 97,75%.

Bij een prevalentie tussen 10% en 15% is het percentage positieve uitslagen maximaal 3,4% minder tot 4,5% meer dan die prevalentie. Op een totale prevalentie van maximaal 15% zijn dit grote marges. Bij een sensitiviteit van 98% en een specificiteit van 99,5% zijn de onzekerheden uiteraard het kleinst.

De grafiek laat zien dat een sensitiviteitsverschil van 31% een grote variatie geeft in de schattingsfout voor de werkelijk prevalentie. Het verschil in schattingsfout bij een gelijkblijvende sensitiviteit maar een 3,5%-verandering in specificiteit is al de helft van de inschattingsfout bij een sensitiviteitswijziging van 31%. Dus een kleine verandering in specificiteit heeft een veel grotere invloed op het mischatten van de prevalentie dan een *gelijke* verandering in sensitiviteit.

Dit komt omdat 4% van de groep met een negatieve uitslag (die fout-blijkt) in aantal uitslagen al de helft is van het aantal juist-positieve uitslagen bij typische testgroepen. Wanneer dit percentage fout-negatieven daalt naar 0,5% gaat het meteen om 8 keer zo weinig uitslagen. Het aantal fout-negatieve uitslagen verandert dan van de helft naar minder dan een tiende van het aantal

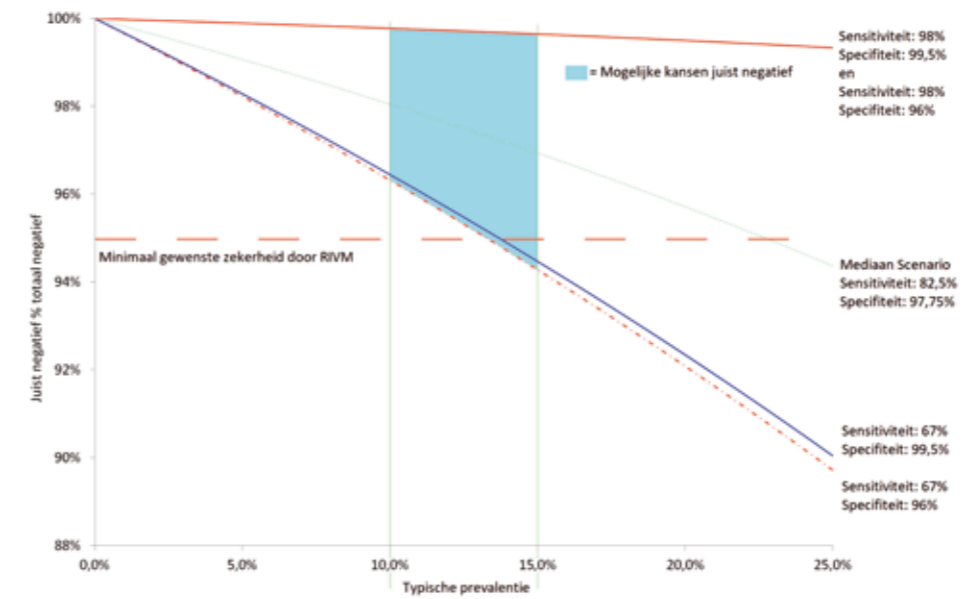
juist-positieve uitslagen waardoor een veel groter deel van de positieve uitslagen juist is.

De inschattingsfout is bij een gegeven sensitiviteit en specificiteit ook niet constant voor verschillende prevalenties. De prevalentie is het aantal juist-positieve uitslagen opgeteld bij het aantal fout-negatieve uitslagen. Het aantal positieve uitslagen is het aantal juist-positieven opgeteld bij het aantal fout-positieven. Bij de prevalenties in de grafiek zijn zelfs 3 van de 4 scenario's bij een bepaalde prevalentie helemaal juist. De grafiek geeft de indruk dat voor de testgroepen het scenario met de laagste sensitiviteit en specificiteit het op een na beste scenario is.

Op die punten waarin geen voorspellingsfout gemaakt wordt in het aantal patiënten is het aantal fout-positieven gelijk aan het aantal fout-negatieven. Zo is het totaal aantal positieve uitslagen gelijk aan het aantal werkelijke COVID-19-patiënten. Het zijn alleen niet allemaal de juiste mensen die de positieve uitslagen krijgen. Dus voor 3 van de 4 test scenario's is er wel een punt waarin het juiste aantal mensen geïdentificeerd wordt en in quarantaine zit, maar een flink deel daarvan is niet de juiste persoon. In tabellen 2 en 3 doet dit fenomeen zich ook voor.

Groepspectief versus individueel perspectief

De vraag 'Welk deel van de COVID-19 patiënten wordt gevonden?' heeft een 'groepspectief'. Mensen die zelf een uitslag krijgen willen het antwoord op de vraag: 'Wat is de kans dat mijn uitslag juist-positief of juist-negatief is?'. Deze vraag heeft een 'individueel perspectief'. Daarbij is de hoeveelheid juist-positieven en juist-negatieven *als percentage van alle of positieve of negatieve uitkomsten* van belang.



Figuur 2. Juist-negatieve uitslagen als percentage van alle negatieve uitslagen

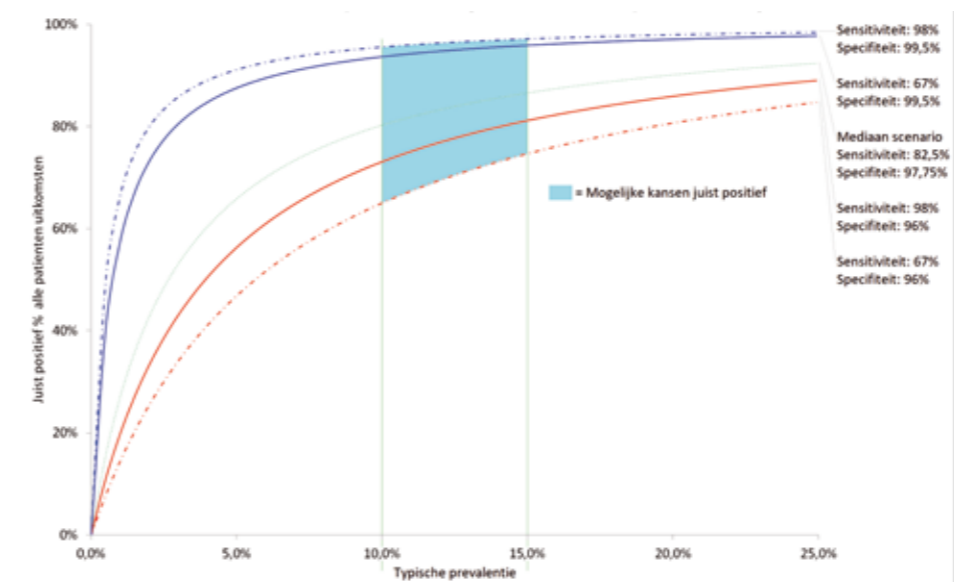
Figuur 2 toont de cijfers voor een negatieve uitkomst: De kans dat een negatieve uitslag juist is wanneer je zo'n uitslag krijgt ligt tussen de 94% en 100%. Het RIVM wenst dat de kans op een juist-negatieve uitslag minimaal 95% is, dit is een zekerheid die in de praktijk nagenoeg altijd gehaald zal worden. Deze keer heeft de sensitiviteit een grote invloed. Wanneer die hoog is zijn er relatief veel mensen met een juist-positieve uitslag. Dan zijn er nauwelijks patiënten over die een negatieve uitslag kunnen krijgen, het aantal fout-negatieven is noodzakelijkerwijs klein.

Ook is te zien dat hoe hoger de prevalentie is, hoe groter de kans is dat een negatieve uitslag onjuist is. Het aantal patiënten dat geen juiste uitslag krijgt wordt namelijk groter en de groep niet-patiënten kleiner. Als percentage van de niet-patiënten neemt het aantal patiënten met een fout-negatieve uitslag hard toe.

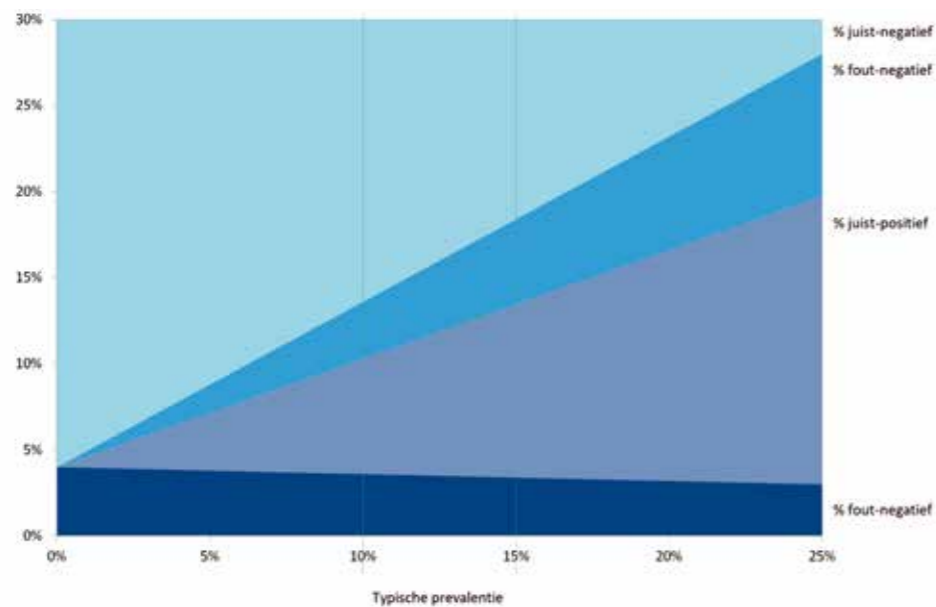
Belang van individu met negatieve uitslag tegen gesteld aan belang van individu met positieve uitslag

Voor het testen op COVID-19 ligt de kans dat een positieve uitslag juist is tussen de 65% en 97,2%. Voor de kans op een juist-positieve uitslag heeft de kleine verandering in specificiteit van 3,5% een veel groter effect dan enorme verandering in sensitiviteit van 31%. Dit is weer het gevolg van het feit dat het merendeel van de geteste mensen geen COVID-19 heeft. Een net wat groter deel van die mensen als correct negatief identificeren is relatief veel mensen ten opzichte van het aantal COVID-19-patiënten. Deze mensen kunnen dan niet meer een fout-positieve uitslag krijgen. Zie figuur 3.

Wanneer er gekozen wordt om ook te testen zonder



Figuur 3. Juist-positieve uitslagen als percentage alle positieve uitslagen



Figuur 4. Percentage alle uitslagen, mediaan scenario

klachten zal de prevalentie in de testgroep flink afnemen en de kans op een juist-positieve uitslag daalt dan ook snel. Hoe hoog de specificiteit dan ook is, bij heel weinig patiënten is een kleine hoeveelheid fout-positieven al snel een groot aantal vergeleken met het aantal patiënten. De groep die een positieve uitslag ontvangt bevat dat veel niet-patiënten.

Meer COVID-19-patiënten in de testgroep is in het belang van de mensen die een positieve uitslag ontvangen. De kans dat die daadwerkelijk klopt is groter. Het gaat juist tegen het belang in van mensen die een negatieve uitslag krijgen, de kans dat die klopt wordt juist kleiner. De sensitiviteit en specificiteit zijn met beleid niet goed te beïnvloeden. Maar hoeveel COVID-19 patiënten in de testgroep zitten kan wél worden beïnvloed door adviezen te geven over wanneer wel en niet te testen. Een beleidsmaker moet kiezen tussen meer mensen met COVID-19 rond laten lopen of meer mensen zonder COVID-19 isoleren.

Het totale plaatje

Inzicht in de aandelen juiste en onjuiste uitkomsten van negatieve en positieve testen voor de gehele testgroep is wenselijk. Voor het mediane testscenario is de opbouw van alle testuitslagen zoals figuur 4 toont

De absolute hoeveelheid fout-negatieven neemt sneller toe dan de hoeveelheid fout-positieven afneemt bij een hogere prevalentie. Dus het aantal onjuiste uitslagen is groter bij een toenemende prevalentie. Dat komt omdat voor deze testen de sensitiviteit doorgaans lager is

dan de specificiteit. Al met al blijft de hoeveelheid juiste uitslagen (negatief en positief) groot.

Testbeleid in de praktijk

Het testen om COVID-19-patiënten op te sporen is steeds effectiever naarmate het percentage patiënten in de testgroep toeneemt. Hoe hoger de prevalentie, hoe groter het aandeel patiënten dat je vindt en hoe groter de kans dat een positieve uitslag juist is. Het testbeleid is er ten tijde van dit schrijven dan ook op gericht om het aandeel COVID-19-patiënten in test groepen zo hoog mogelijk te maken. De kans dat een negatieve uitslag juist is neemt dan wat af maar blijft doorgaans boven de 95%. Een minimale wijziging in de specificiteit heeft een grote invloed op het aantal gevonden patiënten. Om mensen meer vrijheid te geven wordt gedacht over continue testen. De prevalentie van COVID-19 in de testgroep is dan laag en dan zal een positieve uitslag zal weinig toegevoegde waarde hebben. De kans op een fout-positieve uitslag is immers erg groot.

LITERATUUR

RIVM. (z.d.) *Over de betrouwbaarheid van de PCR-test voor SARS-CoV-2*. RIVM.

ANTON MEIJBURG studeerde Toegepaste Wiskunde aan de Stenden NHL Hogeschool Leeuwarden en is Statistisch Analist A. Hij werkt bij PGGM Vermogensbeheer als medewerker bestuurlijke informatie.
E-mail: antonmeijburg@gmail.com.



Onlangs bevatte *STATOR* een artikel waarin werd beschreven hoe 19^e-eeuwse statistici zich inzetten tegen drankmisbruik. Toen ik het las moest ik meteen denken aan een van de zeldzame keren in mijn leven dat ik behoorlijk aangeschoten ben geraakt, en dat nog wel in functie!

Het was aan het eind van de jaren 90 van de vorige eeuw, tijdens een van de tweejaarlijkse ISI-congressen. Een gebruikelijk onderdeel van die congressen is het diner dat de Directeur-Generaal voor de Statistiek van het gastland geeft voor al zijn collega's die het congres bezoeken en nog wat gasten van vergelijkbare status. Afhankelijk van budget, luim van de gastheer en ruimte in de zaal worden dan, naast de directeur van het Permanent Office van het ISI, ook enkele medewerkers van dat Office uitgenodigd. Ik heb gedurende de ruim 25 jaar dat ik als vrijwilliger bij dat Permanent Office werk vrijwel alle congressen bezocht en regelmatig aan zo'n diner deelgenomen.

Ook ditmaal was ik uitgenodigd en men had mij tegenover de DG van een Balkanland geplaatst, een land dat nauw betrokken was bij de vele vijandigheden die volgden op het uiteenvallen van het voormalige Joegoslavië. De goede man sprak amper Engels en werd daarom vergezeld van een assistente die als tolk fungeerde. Na het voorgerecht ging het mis, hij was tot de ontdekking gekomen dat ik uit Nederland kwam, een van de landen die participeerden in het bombarderen van Sarajevo. Hij hield mij zo ongeveer persoonlijk medeverantwoordelijk voor die bommen. Met een verontschuldigende glimlach vertaalde zijn assistente zijn woedende opmerkingen in mijn richting. Daarbij vermoedde ik dat ze de ál te scherpe kantjes weglief. Nu heb ik altijd weinig interesse en inzicht gehad in politieke ontwikkelingen, en de ingewikkelde verhoudingen in de Balkan waren al helemaal onbegrijpelijk voor me. Maar ik kon de heftige verbale aanval

len moeilijk negeren, ik zat daar tenslotte niet alleen en allengs begonnen meer mensen mee te luisteren.

Ik besloot me totaal te onthouden van enig politiek commentaar, dat was niet moeilijk want zoals gezegd wist ik daar niets van. Ik heb verteld dat ik me verontschuldigde voor mijn totale politieke ignorantie en zei dat dit een wetenschappelijke bijeenkomst was waarbij ik aanwezig was als wetenschapper. Ik zei dat hij dat ongetwijfeld ook was en stelde daarom voor om, als collega-wetenschappers, samen het glas te heffen op onze gezamenlijke interesse en de politiek er verder buiten te laten. Waarschijnlijk was de man op voornamelijk politieke gronden in zijn belangrijke functie benoemd want hij was zichtbaar gevlod dat ik hem als wetenschapper beschouwde. Tot mijn grote opluchting sloeg zijn eerst zo vijandige houding om als een blad aan een boom, hij omhelsde me, greep zijn glas, leegde het in een teug en schonk onze glazen nog eens in.

De politiek kwam die avond verder niet meer ter sprake, wel moest er vele malen op onze wetenschappelijke vriendschap worden geklonken. Ondanks de uiterst zuinige nipjes die ik nam was ik aan het eind van het diner toch meer aangeschoten geraakt dan me lief was. Maar gelukkig schoot ook daar die assistente weer te hulp, zij gaf me een steuntje op de smalle loopplank naar de boot die ons terug zou brengen van het eiland waar het diner plaatsvond. Eenmaal op de vaste wal, buiten gehoorbereik van haar baas, verontschuldigde ze zijn eerdere aanvallen en zei 'You handled that awful situation very well'.

Tsja, er zullen niet veel wetenschappers zijn die kunnen zeggen dat ze uit hoofde van hun functie aangeschoten zijn geraakt.

GERRIT STEMERDINK is eindredacteur van *STATOR*.
E-mail: gjstemerding@hotmail.com