

Kwantificeren van de accuratesse van sneltesten in verschillende contexten

Achtergrond

Het is algemeen bekend dat de accuratesse (sensitiviteit, specificiteit en voorspellende waarden) en dus effectiviteit van medische testen varieert per context. Zo ook van de corona-sneltesten.

Zo is de accuratesse van de corona-sneltest zeer waarschijnlijk anders tussen/in, bijvoorbeeld:

1. Mensen met en zonder COVID-19 symptomen/klachten; ouderen versus niet-ouderen;
2. Mensen met recent gestarte (bv. 1 tot 3 dagen) versus langer aanwezige (bv. 7 of meer dagen) COVID-19 symptomen/klachten;
3. Klachtenvrije mensen met of zonder een indicatie van langdurige (>15 min) in nabijheid van een besmet persoon;
4. Mensen die werken in risicoberoepen (zorgpersoneel, slachthuizen, leerkrachten, etc.) versus andere beroepen;
5. Situaties waarin de sneltest ad random wordt toegepast als screeningstest (d.w.z. testen van mensen zonder klachten/symptomen en zonder indicatie van langdurig nauw contact met besmet persoon) bij bijvoorbeeld profsporters, evenementen, werknemers, of studenten;
6. Situaties waarin de sneltest wordt afgenomen door een zorgprofessional of door de persoon zelf in een thuissituatie.
7. Situaties waarin de sneltest wordt afgenomen met materiaal afgenomen uit de neus, keel, nasopharyngeaal, of beiden.
8. Of in een thuissituatie van kwetsbare ouderen/kinderen waarin de test niet wordt uitgevoerd door de persoon zelf maar door (bijvoorbeeld) een zorgprofessional of mantelzorger.

Bovenstaande doelgroepen/situaties zijn slechts voorbeelden waarin de sneltesten gebruikt en ingezet (kunnen) gaan worden. Afhankelijk van het belang van de doelgroep en toepassing kan en moet wellicht geprioriteerd worden. Het is van belang zijn om zo snel mogelijk de accuratesse van de sneltesten gekwantificeerd te hebben in elk van die doelgroepen/situaties waarin deze (mogelijk) ingezet gaat worden.

Onderstaande is toegespitst op de zogenoemde antigeen sneltesten, maar in feite geldt onderstaande ook grotendeels voor bijvoorbeeld eNoses (blaastesten), de LAMP-sneltest, en andersoortige sneltesten.

Living healthcare systeem

Idealiter evalueren we, voordat we nieuwe testen in de reguliere zorg of als screeningstest toelaten, de accuratesse en waarde van de nieuwe allereerst via formeel opgezette en uitgevoerde wetenschappelijke studies. Bij de studiedeelnemers wordt dan zowel de nieuwe test als de huidige referentietest (ook wel gouden standaard genoemd) uitgevoerd. Vervolgens worden beide testuitslagen met elkaar vergeleken en de waarde van de nieuwe test t.o.v. de referentie uitgerekend (zie onder). Indien de nieuwe test in dergelijke wetenschappelijke (gecontroleerde) studies accuraat genoeg is bevonden, kan men besluiten

de test in de reguliere settings (waarin de test onderzocht ofwel gevalideerd was) toe te laten.

Deze luxe en tijd kunnen we in de huidige pandemie waarin het Cov-19 virus zich razendsnel verspreidt, helaas niet veroorloven. We zullen deze evaluaties al doende, direct in de beoogde contexten, moeten uitvoeren: learning while doing in een zogenoemd living healthcare systeem. Dit is voor de reeds CE gemarkeerde testen ook legitiem, maar kan ook gedaan – en moet zeker overwogen worden – voor nieuwe testen (sneltesten of niet) die nog geen CE hebben.

Het evalueren van de nieuwe (snel)testen direct in de reguliere contexten, heeft ook heel veel voordelen. De test wordt namelijk direct op haar waarde geëvalueerd in de context waarin het toegepast gaat worden, zoals bijvoorbeeld in een formele GGD-teststraat, in setting risicoberoepen, bij ouderen, bij asymptomatische nauwe contacten uit BCO of CoronaMelder, of in thuissituatie als zelftest, uitgevoerd door persoon zelf of door een zorgverlener of mantelzorgver. Per context kunnen alle bevorderende en belemmerende factoren (inclusief logistieke factoren) die van invloed zijn op de uitvoer, accuratesse en dus op de effectiviteit van de sneltest, direct worden meegenomen en beoordeeld.

Te kwantificeren parameters

In de vergelijking tussen de uitslagen van de sneltest en die van de huidige referentietest (PCR), gaat het om het per doelgroep/situatie kwantificeren van zowel de sensitiviteit (ook vaak de gevoeligheid genoemd) en specificiteit als ook van de voorspellende waarden van de positieve en negatieve sneltest uitslagen. Want per doelgroep/situatie kunnen deze parameters nu eenmaal erg verschillen.

- De sensitiviteit en specificiteit zijn zogenoemde *groepsparameters*
 - o *De sensitiviteit* geeft aan welk percentage van de daadwerkelijk (door de PCR geclassificeerde) COVID-19 besmette personen ook terecht door de sneltest worden opgespoord doordat beide testen positief resultaat geven (terecht-positieven); en welke percentage wordt gemist doordat de sneltest negatief aangeeft en de PCR-referentietest discordant, negatief, is (vals-negatieven). Let op: het gaat hier over besmette personen zoals door de PCR bepaald, en niet over besmettelijke personen.
 - o *De specificiteit* geeft aan welk percentage van de daadwerkelijk (door de PCR geclassificeerde) niet-besmette personen ook terecht door de sneltest worden aangeduid als niet besmet (terecht-negatieven) en welk percentage onterecht door de sneltest wordt aangeduid als besmet – dus weer een discordant testresultaat, in dit geval vals-positief van de sneltest.
- De voorspellende waarden zijn de zogenoemde individuele kansen, die gelden voor de persoon die de sneltest ondergaat.
 - o De voorspellende waarde van de positieve sneltest uitslag geeft de kans aan dat de geteste persoon ook daadwerkelijk COVID-19 heeft. 1 minus die kans is dus de kans dat ondanks dat de test positief is de persoon toch geen COVID-19 heeft.

- o De voorspellende waarde van de negatieve sneltest uitslag geeft de kans aan dat de geteste persoon ook daadwerkelijk geen COVID-19 heeft. 1 minus die kans is dus de kans dat ondanks dat de test negatief is de persoon toch COVID-19 heeft.
- Voor alle 4 de parameters geldt dat er altijd onzekerheid (ofwel imprecisie ofwel toeval) in de gekwantificeerde waarden zitten. Hoe kleiner de groep deelnemers waarop het onderzoek is uitgevoerd, hoe groter die onzekerheid (toevalsbevinding). Deze onzekerheid wordt per parameter uitgedrukt in de zogenoemde standaard fout, en die volgt direct uit de verzamelde data. Met deze standaard fout wordt per parameter het zogenoemde betrouwbaarheidsinterval berekend.

Nota Bene.

- Vaak spreekt men in geval van sneltesten over de gevoeligheid van de test. Dit verwijst echter alleen naar de sensitiviteit – en betreft dus slechts een van de vier accuratesseparameters die bij de beoordeling van de testen in acht genomen moeten worden. En het is afhankelijk van de situatie of de sensitiviteit idealiter zo hoog mogelijk is (dus weinig FN – minder schijnveiligheid) of de specificiteit (weinig FP – minder mensen onterecht als ziek gelabeld en dus in quarantaine)
- Sensitiviteit en specificiteit zijn dus parameters die de waarde van de test op populatieniveau aangeven – en van belang zijn voor beleidsmakers. Ze geven aan: hoeveel van alle werkelijke besmette mensen (volgens PCR) we ook zullen opsporen met de sneltest, en hoeveel van de werkelijke niet-besmette mensen we inderdaad als niet-besmet bestempelen met de test. De voorspellende waarden zijn echter de absolute kansen op werkelijk besmet of niet-besmet zijn volgens de sneltest, die direct voor de geteste persoon gelden. Men moet bij besluitvorming over de inzet van de sneltesten altijd alle vier de parameters – inclusief hun precisie/betrouwbaarheidsintervallen - meenemen. En helaas is het vaak kiezen tussen twee kwaden: afhankelijk welke discordantie we het meeste van belang vinden en willen minimaliseren: FN of FP, en in welke situatie.

Generieke onderzoeksopzet

Onderstaande opzet voor het kwantificeren van de accuratesse van corona-sneltesten geldt in feite voor elke sneltest, en voor elke context waarin de sneltest uitgevoerd zou kunnen gaan worden, inclusief als zelftest. Uiteraard gelden voor de laatste situatie ook andere aspecten zoals het inrichten van de meldplicht van positieve sneltestuitslagen en de invoer ervan in bestaande databases, maar dat valt buiten de scope van dit document (zie andere documenten).

In feite is, ongeacht de context, de onderzoeksopzet grofweg zoals hieronder beschreven. Uiteraard wordt per context (bv. sneltesten in teststraten uitgevoerd door ervaren zorgpersoneel of als zelftesten) de opzet, uitvoer en analyse op maat afgestemd.

- Definieer helder en objectief de context (doelgroep, wijze van afname, etc.) waarin de sneltest mogelijk ingezet gaat worden.

Bij iedere deelnemer wordt de sneltest uitgevoerd alsmede de PCR test als referentie. Aangezien er meerdere PCR methoden in omloop zijn, moet goed vooraf goed bepaald worden welke PCR methode in welke evaluatie gebruikt wordt, en ook voor alle deelnemers in die setting dezelfde PCR methode gebruikt wordt.

- Zowel de sneltest als de PCR analyse gebeurt idealiter op hetzelfde afgenomen keel/neus materiaal. Zo voorkom je dat, bijvoorbeeld, het materiaal voor de sneltest iets anders was verkregen dan voor de PCR test (bv. wattenstaafje in ene geval iets minder diep is ingebracht). Indien dit niet mogelijk is dan beide materiaalafname door zelfde persoon laten doen.
- Zelftest-situatie. Indien het een zelftest-situatie betreft, kan gekozen worden om de uitvoer van de sneltest en het afnemen van het materiaal voor de PCR test bewust *niet* door dezelfde persoon uit te laten voeren, om zoveel mogelijk de beoogde zelftest-context inclusief methode van materiaalafname na te bootsen. Ook is het denkbaar dat in kwetsbare groepen (bv. ouderen, kinderen op scholen) de zelftest niet door de persoon zelf, maar door een getraind persoon (bv. verpleegkundige of mantelzorg) op locatie wordt uitgevoerd. Ook dit moet dan direct in die hoedanigheid worden meegenomen – zodat we direct de waarde van de zelftest in de beoogde context kunnen meten. De instructie moet op dezelfde manier gegeven worden als tijdens het beoogde gebruik na het onderzoek, conform een living healthcare opzet.
- Om na te gaan of accuratesse van de sneltest mogelijk varieert tussen bepaalde situaties/condities, leggen we tevens een aantal andere kenmerken van de deelnemers vast zoals: klachten ja/nee, aantal dagen klachten, ernst van klachten, door wie is materiaal afgenomen (bv. zorgprofessional, zelf, mantelzorg), hoe is het sneltest materiaal bewerkt en geanalyseerd (indien van toepassing), etc.
- Groepsgrootte: Definieer vooraf de mate van standaardfout (en dus breedte van het betrouwbaarheidsinterval) dat men wil accepteren. Hoe kleiner, hoe preciser men wil zijn, hoe groter de benodigde groepsgrootte. Dit kan vooraf door methodologen en inhoudelijke experts worden berekend. De benodigde groepsgrootte varieert dus per context.
- Data analyse: bereken per context de 4 bovengenoemde accuratesse parameters met betrouwbaarheidsintervallen. Bereken ook of deze accuratesseparameters variëren onder verschillende omstandigheden/subgroepen (bv. mensen met kortdurende versus langdurende klachten; met of zonder nauw contact waarschuwing; etc.). Een dergelijke subgroep (gestratificeerde) of zelfs multivariabele analyse kan eenvoudig worden uitgevoerd, afhankelijk van het deelnemers. Dit kan vooraf berekend worden.

Nieuwe testen die mogelijk beter zijn dan de huidige PCR testen

Bovenstaande onderzoeksopzetten en accuratesse parameters gaan uit van het evalueren van de uitslagen van de sneltesten t.o.v. uitslagen van de huidige PCR referentie (gouden)

standaard. Het gaat hier dus niet om het aantonen of de nieuwe (snel)testen wellicht accurater ofwel beter zijn dan de huidige PCR referentietest: daarvoor zijn andere onderzoeksopzetten nodig dan hierboven beschreven. Hierboven zijn de sneltesten vergeleken *ten opzichte van* PCR. In deze opzet – waarbij de PCR uitslag de waarheid is – kan de nieuwe sneltest nooit beter uitkomen dan de PCR – hooguit even goed. Elke discordantie in testresultaten (d.w.z. de FP en FN situaties) worden toebedeeld aan de sneltest – niet aan de PCR. Als er een test (sneltest of niet) komt die wellicht beter is dan de huidige referentie (PCR), dan moet deze nieuwe test en de PCR niet ten opzichte van elkaar maar *naast elkaar* worden geëvalueerd, waarbij er een derde referentietestmethode nodig is. Als een dergelijke derde referentiemethode niet aanwezig is, wat vaak het geval is in de geneeskunde, wordt veelal het beloop van de ziekte als referentie gebruikt (verergering van symptomen of klachten). Dit wordt vaak bij chronische aandoeningen gedaan of aandoeningen met specifieke klachten. Covid-19 is echter een acute aandoening, kent veelal een a-specifieke symptomatologie, en is in het gros van de geïnfecteerde personen (gelukkig) zelf-limiterend. Beloop van de aandoening als referentiemethode is hier dus lastig.

Oplossingen in geval van COVID-19 bestaan derhalve uit: 1. een derde referentietest (bv. in speciale ziekenhuispopulaties kan men denken aan een aanvullende beeldvormende test); 2. beide testen (nieuwe test en PCR) direct na de eerste test nog eens uitvoeren (in acht nemen dat de tijd dan al geleid kan hebben tot andere virale load/spontaan herstel); 3. Geavanceerde statistische analyses – zogenoemde latente klasse modellen – die omgaan met situaties waarin twee testen worden vergeleken waarvan geen van beide als referentietest kan worden aangeduid.