

MØRKETALLET OG DE SVÆRE BETINGEDE SANDSYNLIGHEDER

Selv med en god test, der viser om man har antistoffer mod COVID-19-virus, er der ofte kun cirka 50 % chance for, at man reelt har haft sygdommen, selvom man tester positiv. Det får vi den matematiske forklaring på her.

I gennem corona-krisen har vi ofte hørt om mørketal. Det henviser til alle de mennesker, der har været smittet med COVID-19, men som ikke optræder i statistikkerne, fordi de aldrig er testet positive - blandt andet på grund af manglen på testkapacitet i starten, eller fordi de ikke har haft symptomer. Vi ved derfor ikke, hvor stor en andel af befolkningen, der reelt har haft COVID-19, og derfor kan antages at være immune - det er et mørketal.

Hvorfor er det så vigtigt at kende mørketallet?

Fra en samfundsmæssig synsvinkel vil vi gerne kunne forudsige, hvor mange syge vi kan forvente i den kommende tid, og da er mørketallet vigtigt, fordi jo større andel af befolkningen, der er immune og ikke kan smitte videre, jo mindre spredning vil der være af sygdommen. Men også fra et personligt synspunkt vil vi gerne vide, om vi er immune og derfor ikke risikerer at blive syge eller at smitte andre.

Det viser sig imidlertid, at selvom vi kan lave pålidelige befolkningsundersøgelser, der kan estimere mørketallet, er det meget sværere at estimere nøjagtigt, hvem de immune personer er. Det skyldes

en kontrainuitiv effekt, når små sandsynligheder er i spil, der kan være svær at forstå. Denne artikel handler om disse sandsynligheder, og hvordan de spiller sammen.

Krav til en god antistoftest

Det er ikke kun mørketallet, vi ikke kender. Der er stadig mange ting, vi ikke ved om COVID-19. Vi ved for eksempel ikke, om man bliver immun, og i så fald, hvor længe man er immun. Der er rapporteret få tilfælde af personer, der er testet positiv for COVID-19 to gange. Men det skal det ikke handle om her.

Lad os for nu antage, at man bliver immun, i hvert fald for en tid, og vi derfor kan bruge mørketallet til at forudsige udviklingen i smitten i befolkningen. Statens Serum Institut har for nyligt igangsat et større studie af befolkningen i Danmark for at afdække mørketallet, *Vi tester Danmark*, hvor en halv million personer har fået tilbudt en antistoftest.

En COVID-19-antistoftest er en blodprøve, der kan påvise, om en person har antistoffer mod SARS-CoV-2, den virus, der giver COVID-19-infektion. Mange antistof-test er i øjeblikket under udvikling, og der er allerede mange gode test tilgængelige, hvor "god" betyder, at testen har en høj *sensitivitet* og en høj *specificitet*. Sensitivitet er sandsynligheden for, at testen korrekt identificerer en person med antistoffer, også kendt som *den sande positivrate*.

En meget følsom test vil identificere de fleste mennesker, der har antistoffer, og kun en lille del af dem med antistoffer vil ikke blive fundet af testen, også kaldet *falske negative*. Specificitet er sandsynligheden for, at testen korrekt identificerer en person uden antistoffer. Dette er kendt som *den sande negativrate*. En test med høj specificitet vil identificere de fleste mennesker uden antistoffer, og kun en lille del af dem uden antistoffer vil testen identificere som havende antistoffer, også kaldet *falske positive*.

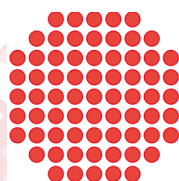
Bemærk, at vi vil gerne have en test, hvor begge disse sandsynligheder er høje. Vi kan altid lave en test med 100% sensitivitet, hvis vi



Foto: Lars Sørensen

Forfatteren

Susanne Ditlevsen er professor ved Institut for Matematiske fag, Københavns Universitet. Hun forsker i statistisk analyse, stokastiske processer og matematiske modeller i biologi. susanne@math.ku.dk



Logoet for Vi tester Danmark.

**VI TESTER
DANMARK**

Epidemiologiske begreber

Figuren opsummerer mulighederne, når befolkningen testes for antistoffer mod SARS-CoV-2, den virus der giver COVID-19-infektion. Antag, at befolkningen består af N personer. Man kan da inddele dem i fire grupper givet ved, om deres test vil være positiv eller negativ, og givet ved, om de har antistoffer eller ej. Hvis personen har antistoffer, og testen er positiv, kaldes det en sand positiv, og på samme vis for de andre felter.

De epidemiologiske nøglebegreber, kan aflæses af figuren.

Prævalens (mørketal): andelen af befolkningen med en bestemt tilstand (Her er det andelen af befolkningen, der har antistoffer mod SARS-CoV-2). Mørketallet eller prævalensen kan findes ved n_{ja}/N i diagrammet.

Sensitivitet: Sandsynligheden for at en test er positiv givet, at personen, der testes, har det, der testes for (her antistoffer). Sensitiviteten kan findes ved n_{11}/n_{ja} i diagrammet.

Specificitet: Sandsynligheden for, at en test er negativ givet, at personen, der testes, ikke har det, der testes for, her antistoffer. Specificiteten kan findes ved n_{22}/n_{nej} i diagrammet.

Positiv prædiktiv værdi: sandsynligheden for, at en person har det, der testes for givet, at personen er testet positiv. I vores tilfælde er det sandsynligheden for at have antistoffer givet, at man er testet positiv. Den positive prædiktive værdi kan findes ved n_{11}/n_{pos} i diagrammet.

Sand positiv: Et positivt testresultat for en person, der har det, der testes for, her antistoffer.

Sand negativ: Et negativt testresultat for en person, der ikke har det, der testes for.

Falsk positiv: Et positivt testresultat selvom personen ikke har det, der testes for.

Falsk negativ: Et negativt testresultat selvom personen har det, der testes for.

Anti-stoffer	Ja	Nej	
Test			
Positiv	n_{11} Sand positiv	n_{12} Falsk positiv	n_{pos}
Negativ	n_{21} Falsk negativ	n_{22} Sand negativ	n_{neg}
	n_{ja}	n_{nej}	N

Rækkesummerne angiver antallet af positive henholdsvis negative tests, $n_{pos} = n_{11} + n_{12}$, og $n_{neg} = n_{21} + n_{22}$.

Søjlesummerne angiver antallet med og uden antistoffer

$$n_{ja} = n_{11} + n_{21} \text{ og } n_{nej} = n_{12} + n_{22}$$

Vi får således, at $n_{pos} + n_{neg} = n_{ja} + n_{nej} = N$.

er ligeglade med specificiteten – vi kan jo bare sige, at alle er positive! Men det er åbenlyst ikke brugbart. Der er allerede udviklet flere test med en sensitivitet på næsten 100% og også en høj specificitet på omkring 95-99%. Med så store sandsynligheder må vi da kunne stole på resultatet af testen! Men, desværre, sådan er det ikke. Lad os se, hvorfor det ikke er tilfældet.

Når mørketallet spiller ind

Sandsynligheden for, at personer med et positivt testresultat, virkelig har antistoffer kalder vi den *positive prædiktive værdi*. Det er altså den, vi er interesseret i, når vi bliver testet og vil vide, hvad sandsynligheden er for, at vi er immune, forudsat at testen er positiv. Men den positive prædiktive værdi er

ikke det samme som sensitiviteten, og det er derfor, det er så vanskeligt. Mens sensitiviteten kun afhænger af egenskaber ved testen, afhænger den positive prædiktive værdi nemlig også af mørketallet – det vil sige andelen af befolkningen, der har antistoffer i blodet på tidspunktet for testen. Den positive prædiktive værdi vil typisk være meget mindre end sensitiviteten, hvis der ikke er så mange, der har antistoffer.

Sensitiviteten og den positive prædiktive værdi er kun de samme, hvis sandsynligheden for, at en person har antistoffer, er den samme som sandsynligheden for, at testen er positiv. Det kræver, at sensitiviteten og specificiteten er 100%, men det sker kun meget

sjældent (hvis nogensinde!) for hvilken som helst test. Faktisk er de meget forskellige, hvis mørketallet er lavt, som det nok er de fleste steder på dette tidspunkt i pandemien. Det er nemlig sådan, at jo lavere mørketallet er, desto lavere er den positive prædiktive værdi. Det betyder, at en COVID-19-antistof-test, selv med høj sensitivitet og specificitet, der anvendes i områder, hvor mørketallet er lavt, vil have en lavere positiv prædiktiv værdi end i et område, hvor mørketallet er højere. Mørketallet varierer formentlig meget fra område til område, men de fleste steder i Danmark er et skøn på 2-5% nok ikke langt fra sandheden i skrivende stund.

Lad os antage en test med sen-

Betingede sandsynligheder og Bayes' formel

En *betinget sandsynlighed* er sandsynligheden for, at en bestemt hændelse indtræffer givet, at en anden hændelse er indtruffet. Lad os definere to hændelser A og B. Da betegnes den betingede sandsynlighed for, at A indtræffer givet, at B er indtruffet, for $P(A|B)$, der læses "sandsynligheden for A, givet B".

Eksempel: lad hændelserne A og B være

A: en person har antistoffer

B: personen har et positivt testresultat

Sandsynligheden $P(A)$ er mørketallet, og $P(B)$ er sandsynligheden for, at testen er positiv for en tilfældigt udvalgt person. Da kan sensitiviteten skrives som den betingede sandsynlighed

$$P(B|A)$$

og den positive prædiktive værdi kan skrives som

$$P(A|B).$$

De er ikke det samme! Bayes' formel fortæller os, hvad forholdet er mellem de to betingede sandsynligheder. Bayes' formel siger, at

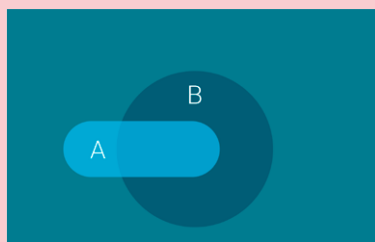
$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

hvor vi antager, at $P(B) > 0$. Vi ser således, at de to betingede sandsynligheder kun er det samme, hvis $P(A) = P(B)$. Her er $P(A)$ og $P(B)$ *marginale* sandsynligheder, dvs. sandsynligheder for enkelthændelser uden at inddrage eventuelle sammenhænge med andre hændelser. Sandsynligheden $P(B)$ kan findes ved loven om total sandsynlighed, hvor man opdeler efter, om personen har antistoffer eller ej,

$$\begin{aligned} P(B) &= P(B|A)P(A) + P(B|\text{ikke } A)P(\text{ikke } A) \\ &= \text{sensitivitet} \times \text{prævalens} + \\ &\quad (1 - \text{specificitet}) \times (1 - \text{prævalens}). \end{aligned}$$

Her har vi brugt, at $P(\text{ikke } A) = 1 - P(A)$ og $P(B|\text{ikke } A) = 1 - P(\text{ikke } B|\text{ikke } A)$, hvor $P(\text{ikke } B|\text{ikke } A)$ er lig med specificiteten.

Illustration af de betingede sandsynligheder. Hele firkanten symboliserer hele befolkningen, og arealet af mængderne A og B er proportionale med sandsynligheden for hændelserne. Fællesmængden for A og B er repræsenteret ved overlappet mellem de to mængder, og arealet er proportional med sandsynligheden for at begge hændelser sker.



$$\begin{aligned} P(A) &= \frac{\text{area af A}}{\text{area af firkant}} & P(A|B) &= \frac{\text{area af A}}{\text{area af B}} \\ P(B|A) &= \frac{\text{area af B}}{\text{area af A}} \end{aligned}$$

sitivitet på 100% og specificitet på 98%, der bruges i et område, hvor mørketallet er 2%. Så er den positive prædiktive værdi lig med 0,5 (se Faktaboks 3). Selv med så god en test er sandsynligheden for at have haft COVID-19 altså kun fifty-fifty, selvom du tester positiv for antistoffer!

Usandsynlige hændelser i spil

Årsagen er, at der er to usandsynlige hændelser i spil: sandsynligheden for, at du er immun (en lille sandsynlighed givet ved mørketallet), og derfor testes positiv (med stor sandsynlighed, i eksemplet ovenfor med sandsynlighed en) – eller sandsynligheden for, at du vil

blive testet positiv, selvom du ikke har antistoffer (en lille sandsynlighed), men mange mennesker uden antistoffer vil blive testet (fordi mørketallet er lille), og dermed vil antallet af falsk positive være stort.

Hvorfor dette er så kontraintuitivt forklares smukt i den stærkt anbefalelsesværdige bog af nobelprisvinder Daniel Kahneman, *Thinking, Fast and Slow*, der viser, hvordan vi psykologisk har tendens til at glemme eller ignorere forekomsten af det, vi undersøger (mørketallet) i lyset af yderligere information (resultatet af testen).

En lav positiv prædiktiv værdi betyder, at flere individer får et falsk positivt resultat, hvilket er farligt, da et positivt testresultat sandsynligvis vil gøre os mere afslappede og måske ændre adfærd, fordi vi ikke er bange for at blive smittet eller at smitte andre. I dette specielle eksempel kan vi være sikre på, at vi ikke har antistoffer, hvis vi får et negativt resultat, da sensitiviteten er 100%, så der er ingen falske negative.

Et aktuelt eksempel

Lad os lave udregningerne for den antistoftest, der er blevet tilbudt et stort udsnit af den danske befolkning i kampagnen *Vi tester Danmark*. Her er sensitiviteten omkring 93% og specificiteten 98%. Testen er altså ikke lige så god som de bedste test, man kan få. Hvis mørketallet er 2%, da bliver den positive prædiktive værdi lig med 0,49. Det gør altså ikke den store forskel for den positive prædiktive værdi, at sensitiviteten er så meget lavere. Det er fordi, den største fejlkilde er falsk positive (og ikke falsk negative).

Formentlig er mørketallet højere visse steder, for eksempel i København, og det gør faktisk en stor forskel. Lad os antage, at den er 5%, da får vi, at den positive prædiktive værdi er lig med 0,71. Den positive prædiktive værdi stiger altså fra omkring 50% til omkring

Udregning af den positive prædiktive værdi

Den positive prædiktive værdi er givet ved udtrykket:

$$\frac{\text{sensitivitet} \times \text{prævalens}}{\text{sensitivitet} \times \text{prævalens} + (1 - \text{specificitet}) \times (1 - \text{prævalens})}$$

Lad os antage en test med sensitivitet på 100% og specificitet på 98%, der bruges i et område med prævalens på 2%. Så er den positive prædiktive værdi:

$$\frac{1 \cdot 0,02}{1 \cdot 0,02 + (1 - 0,98)(1 - 0,02)} = 0,5.$$

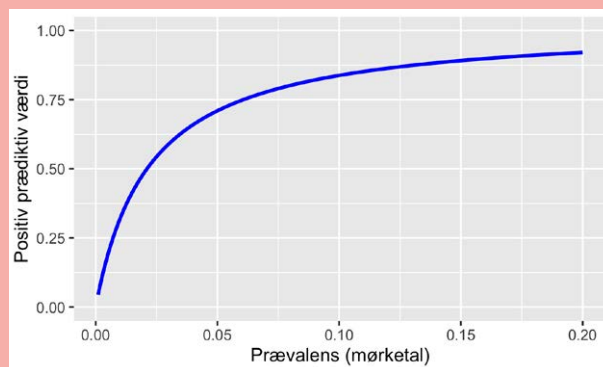
Mange af de tests, der bruges nu, har så høj sensitivitet og specificitet (dog nok ikke dem, man kan købe på apoteket).

Lad os regne nogle eksempler med de tests, der benyttes i det store danske studie *Vi tester Danmark*. Her er sensitiviteten omkring 93% og specificiteten 98%. Hvis prævalensen er 2%, da bliver den positive prædiktive værdi lig med:

$$\frac{0,93 \cdot 0,02}{0,93 \cdot 0,02 + (1 - 0,98)(1 - 0,02)} = 0,49.$$

Hvis prævalensen er 5%, da får vi:

$$\frac{0,93 \cdot 0,05}{0,93 \cdot 0,05 + (1 - 0,98)(1 - 0,05)} = 0,71.$$



I figuren er den positive prædiktive værdi afbildet som funktion af prævalensen for en test med sensitivitet på 93% og specificitet på 98%.

70%, blot fordi mørketallet stiger fra 2 til 5%.

Selvom det er svært at identificere de specifikke personer, der har antistoffer, kan vi stadig bruge testene til at opnå gode estimater af

mørketallet. Hvis vi tester mange, kan vi korrigere for det forventede antal falske positive og falske negative. Vi kan derfor få rimelige befolkningsestimater, selvom de individuelle estimater ikke er gode.

Litteratur

Statens Serum Institut, Vi tester Danmark. www.vitesterdanmark.dk.

Statens Serum Institut, Fokus på mørketallet for COVID-19.

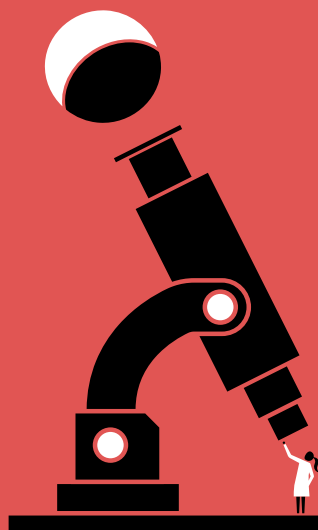
www.ssi.dk/aktuelt/nyheder/2020/fokus-pa-morketallet-for-covid-19.

FDA U.S. Food & Drug Administration. EUA Authorized Serology Test Performance. www.fda.gov/medical-devices/coronavirus-disease-2019-covid-19-emergency-use-authorizations-medical-devices/eua-authorized-serology-test-performance. Retrieved on September 9, 2020.

D. Kahneman. (2020). Thinking, Fast and Slow. Penguin Books. 2011

Tag på camp på SDU

Kom tættere på dit studievalg



- **Fysik & Astronomi Camp 6.-7. marts 2021**
sdu.dk/fysikcamp
- **Kemi Camp 6.-7. marts 2021**
sdu.dk/kemicamp
- **Datalogi Camp 6. marts 2021**
sdu.dk/datalogicamp
- **Matematik Camp 7. marts 2021**
sdu.dk/matematikcamp
- **Biotek Camp 6.-7. marts 2021**
sdu.dk/biotekcamp
- **IT Camp for Piger 20.-21. marts 2021**
sdu.dk/itcampforpiger