
Feilbruk av diagnostiske laboratorietester – et beslutningsanalytisk perspektiv

AKTUELT PROBLEM

ELLING ULVESTAD

Email: elling.ulvestad@haukeland.no
Avdeling for mikrobiologi og immunologi
Gades Institutt
Haukeland Sykehus
5021 Bergen

Økende bruk av medisinske tester skyldes neppe bare økt forekomst av sykdom i befolkningen, og det har vært spekulert på om økningen kan skyldes feilbruk av tester. Feilbruk av tester får medisinske og økonomiske konsekvenser, og det er derfor av betydning å redusere denne form for feilbruk.

En viktig egenskap ved tester er at de som verktøy skal skaffe til veie informasjon for medisinske beslutninger. Gjennom beregning av prediktive verdier for positive og negative prøvesvar vil man kunne måle testers nytteverdi i en enhetlig "valuta". Ved dette vil testbruk i ulike beslutningssituasjoner kunne sammenliknes, og feilbruk kan defineres som rekvisisjon av tester som ikke gir beslutningsbærende informasjon.

I artikkelen argumenteres det for at eksterne standarder for riktig bruk av diagnostiske tester vanskelig kan la seg realisere. For å redusere feilbruk blir det derfor viktig at leger erkjenner og etablerer interne beslutningsanalytisk funderte standarder. Slik vil praktiserende leger kunne evaluere testbruken i egen praksis.

Tester rekvireres hyppigst i forbindelse med diagnostikk og overvåking, men også som følge av nysgjerrighet, forskningsinteresse, vane, ønske om juridisk ryggdekning, som resultat av usikkerhet, påtrykk fra pasienten eller som ledd i behandling (1, 2). Resultatet fra testen inngår i en brukssammenheng eller, mer generelt, i en betydningssammenheng med øvrige tegn og symptomer. Ved dette dannes grunnlag for diagnose, behandling og prognose. Mens laboratoriedata bidrar til diagnostisering av mindre enn 10 % av

sykdomstilfellene i allmennpraksis (3), hevdes det at laboratoriedata utgjør grunnlaget for 60 – 70 % av alle beslutninger som fattes vedrørende innleggelse, behandling og utskrivning fra sykehus (4).

Undersøkelser viser stor variasjon mellom legers bruk av tester, noe som tyder på at kliniske indikasjoner for testbruk sjelden er absolutte. Variasjonen er systematisk koblet til ulike karakteristika ved legen eller legens praksis. Faktorer som delvis kan relateres til ulik bruk av tester er type utdanning, antall år i virksomhet som lege, spesialitet, type arbeidsplass, kjennskap til testkostnader og forventninger fra pasientene (5, 6). I tillegg vil ulikheter i kognitiv funksjon (7), som har mer å gjøre med hvordan man tenker enn hva man vet, kunne få konsekvenser for estimering av pretestsannsynlighet for sykdom og derved for hvordan en gitt medisinsk problemstilling løses ved hjelp av tester. En variert bruk av tester er dermed ikke automatisk uttrykk for at enkelte leger bruker testene feil.

I løpet av det siste tiår har kravene til medisinsk praksis gjennomgått store forandringer. Det settes i økende grad søkelys på kunnskapsbasert medisin, praksisretningslinjer og ekstern kontroll. Leger må i større grad enn tidligere redegjøre for den medisinske kvalitet på behandlingen de gir og for utgifter de genererer ved sin praksis. Begrensede økonomiske og personellmessige ressurser gjør det nødvendig å identifisere aktivitet som ikke bidrar til økt helse, slik at korrigerende tiltak kan iverksettes. Når det kan hevdes at variasjon mellom legers bruk av laboratorietester i liten grad får følger for behandlingsresultatet (8), er det rimelig at helsemyndigheter og sykehusledelse vurderer de senere års økende bruk av laboratorietester som et område for mulig feilbruk.

Målsettingen med denne artikkelen er å søke en klargjøring og operasjonalisering av begrepet feilbruk. Diagnostiske tester anvendes for å øke informasjonsgrunnlaget i situasjoner preget av usikkerhet. Innledningsvis gjennomgås derfor ulike modeller for beslutning under tvil. Deretter blir en del karakteristika ved tester og legers diagnostiske strategi omtalt, før en metode for evaluering av testbruk blir vurdert. Det konkluderes med at den foreslåtte metode kan være nyttig for evaluering av testbruk i egen praksis, men at det er tvilsomt om metoden vil kunne tilfredsstille myndighetenes styringsbehov.

Beslutning under tvil

Den klassiske beslutningsteori har sitt utspring fra økonomisk og samfunnsvitenskapelig teori. Filosofisk er teorien basert på arbeidet av Descartes, som mente å ha utarbeidet regler for rettledning av forstanden, på Leibniz' idé om at all tenkning kan beskrives ved hjelp av matematiske formler, samt på utilitaristiske prinsipper, der nytte er en sentral utfallsvariabel. En grunnpilar i modellen er at legen antas å være en målrettet aktør som velger de midler som best tjener til å oppnå målet – korrekt diagnose og behandling. I den bayesianske versjon av modellen forutsettes det at legen som rasjonell aktør kan estimere og rangere sannsynligheter for alle mulige diagnoser som er i overensstemmelse med pasientenes symptomer, og dernest at han kan liste

opp alle alternative diagnostiske strategier for å komme frem til korrekt diagnose. Ved å velge den beste strategien vil pasientens nytte kunne optimaliseres gjennom adekvat behandling. I sin idealtypiske form kan modellen bidra til å klargjøre hvordan man bør handle basert på tilgjengelig empirisk kunnskap, og modellen gir også mulighet for å klassifisere en testprosedyre som korrekt eller feil, sett i relasjon til en gitt standard.

Motsatt denne modellens pretensjoner er medisinsk aktivitet i den virkelige verden preget av usikkerhet på alle nivåer, noe som fører til at medisinske beslutninger sjelden fattes som følge av rasjonelle valg mellom veldefinerte alternativer (9). Tvert imot, beslutninger fattes i situasjoner preget av tidspress, ufullstendig og til tider motsigelsesfull informasjon, og i et miljø der beslutninger må avveies mot administrative og økonomiske begrensninger. I tillegg synes rekkefølgen på informasjonen som innhentes å ha en avgjørende innflytelse på hvordan sannsynligheten for sykdom vurderes (10). Sett fra et slikt perspektiv har legen ingen objektiv basis å velge ut fra, de valg han foretar, gjøres ut fra en allerede foreliggende kompleks situasjon som han så å si er kastet inn i (11). Legen møter pasienten med et sett av diagnostiske fordommer, basert blant annet på pasientens alder, kjønn, yrke og bosted. Disse fordommene, som får betydning for hvilke diagnostiske overveielser han gjør, søker legen så å modifisere og presisere gjennom anamneseopptak og undersøkelse. I hermeneutisk filosofi betegnes slike fordommer som forforståelse (11), innen bayesiansk sannsynlighetsteori tilsvarer fordommene pretestsannsynlighet for sykdom (12).

I beslutningssituasjoner vil handling kunne baseres på fordommer eller på ervervet kunnskap (13). I hvor stor grad handling baseres på instinkt eller på læring vil være en avveining mot miljømessige krav og vil variere fra situasjon til situasjon. En nyere beslutningsteoretisk retning, i hovedsak basert på arbeider fra forskere ved Max Planck-instituttet i München og Berlin, forfekter en teori som nettopp søker å ivareta læringens og de evolusjonsfrembrakte instinkters betydning ved beslutninger (14). Teorien hevder at mennesket gjennom sin mentale evolusjonshistorie har utviklet en evne til å fatte raske og ofte riktige beslutninger ved at enkle mentale regler over tid er blitt tilpasset nøkkeldata fra miljøet. Det poengteres at ulike aktører vektlegger data ulikt, og at bare en del av den informasjon som er tilgjengelig, anvendes når det skal tas en beslutning. Sett i et slikt perspektiv blir rasjonalitet å forstå som økologisk tilpasningsevne. Evolusjonsteoretisk kan dette gi mening hvis en begrenset rasjonalitet, som samtidig både er impulsiv, emosjonell, fleksibel og inkonsistent, bidrar til økt reproduksjon og overlevelse. Når målene er uklare, alternativene vanskelige å fremstille og beregningene tidkrevende, er det, for å uttrykke det paradoksalt, rasjonelt å avstå fra å handle som en rasjonell aktør. I den virkelige verden dreier det seg mer om å treffe en beslutning som virker enn å finne det beste alternativet. Raske og dataknappe beslutninger kan gi nøyaktigere prediksjoner enn langsomme og datarike, blant annet fordi beslutninger basert på kombinasjoner av store mengder informasjon lett kan bli for situasjonsspesifikke og dermed i mindre grad vil være generaliserbare.

Enkelhet som rasjonalitetsfaktor

Enkelthet ved innsamling og fortolkning av data kan være en viktig rasjonalitetsfaktor for beslutninger (15). Enkelhet, som gjennom århundrer har vært fremmet som et ideal for vitenskapelige modeller, er kanskje mest kjent fra William av Ockhams filosofi. "Ockhams barberhøvel", som proklamerer at antall forklarende faktorer ikke skal økes utover det nødvendige, er fortsatt et viktig vitenskapsteoretisk begrep. Sett fra et pragmatisk synspunkt vil enkelhet være viktig fordi enkle modeller gir bedre prediksjoner enn kompliserte modeller, men ikke nødvendigvis fordi de enkle modellene er nærmere "sannheten" enn mer kompliserte modeller (16).

Selv om det intuitivt kan virke riktig at evnen til å fatte raske og riktige beslutninger er å anse som et ledd i artens økologiske tilpasning, vil dette evolusjonsteoretiske begrepets empiriske innhold vanskelig kunne etterprøves eller anvendes praktisk hvis ikke grad av "tilpasning" lar seg måle. I et forsøk på å etterkomme et slikt krav har Forster nylig foreslått å operasjonalisere "økologisk tilpasning" som prediktiv nøyaktighet, det vil si å måle i hvor stor grad ulike beslutningsregler gir korrekt prediksjon av nye data (17). Et slikt pragmatisk syn har mye for seg, men det kan innvendes at vi ved å godta en slik operasjonalisering, samtidig også må godta en viss grad av antirealisme ved diagnostikk, det vil si at medisinske hypoteser og modeller som ikke er fullt forenlige med virkeligheten, likevel kan levere prediktiv informasjon som er av høyere kvalitet enn hypoteser og modeller som er mer virkelighetsnære. Slik sett rører vi her ved et viktig skille mellom målsettingen ved den diagnostiske og den vitenskapelige oppdagelse. Blant annet fordi diagnostiske markører ikke nødvendigvis trenger å være kausalitetsmarkører, vil testdata kunne gi diagnostikeren viktig beslutningsbærende kunnskap, mens vitenskapsmannen ved å anvende de samme data risikerer å fjerne seg fra kunnskap om den patofysiologi han skal klargjøre. I så måte er antirealisme i første omgang et større problem for vitenskapsmannen enn for diagnostikeren.

Testing som ledd i en beslutningsprosess

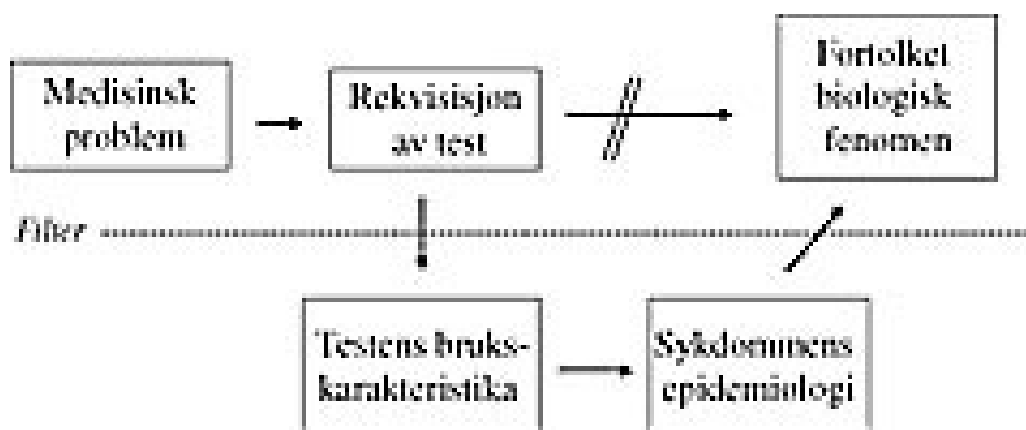
I uoversiktlige situasjoner rekvireres tester for å øke informasjonsgrunnlaget før en beslutning fattes. Ifølge Gigerenzer & Todds (14) teori vil slike beslutninger fungere best når de tar hensyn til kontekstuelle parametere – selv om noe fungerer godt som grunnlag for raske og riktige beslutninger i én kontekst, kan dette "noe" være ubrukelig i en annen kontekst. Applisert på diagnostiske problemstillinger vil et slikt syn være kompatibelt med og velkjent fra bayesiansk statistikk – pretestsannsynlighet for sykdom vil være ulik i ulike diagnostiske kontekster, det vil f.eks. være forskjell i allmennpraksis og i spesialistpraksis. I brukersituasjonen blir testbruk å forstå nettopp som et kontekstavhengig hjelpemiddel – leger har ulik erfaringsbakgrunn, ulike typer

sykdom krever ulik tilnærming, ulike fremtredelsesformer av samme sykdom kan kreve ulik diagnostisk tilnærming, og ulike tester har ulik teknisk og medisinsk verdi.

Diagnostikk er ikke et mål, det er et middel. Ved diagnostikk skal ikke legen nødvendigvis tilstrebe absolutt sikkerhet, hans oppgave er tvert imot å redusere usikkerheten til et akseptabelt nivå. Hvilken informasjon som vurderes å være av betydning, vil i stor grad være preget av hvilke patofysiologiske modeller som legges til grunn for sykdom og av hvilke idealiserte eksempler på riktig resonnering legen møter i sin utvikling som student og praktiserende lege (18). At leger vektlegger beslutningsbærende informasjon ulikt, er tydeliggjort i undersøkelser som viser at enkelte leger bedre enn andre er i stand til å ta vare på den informasjon de får fra negative resultater (19, 20). Når det i tillegg viser seg at enkelte leger bruker en uformell og direkte metode for å avgjøre en tests diagnostiske verdi i klinisk praksis, mens andre bruker en mer formalisert tilnærming (21), burde det ikke komme som noen overraskelse at legers evne til korrekt å vurdere sannsynlighet for sykdom viser stor variasjon, selv ved enkle og avgrensede problemer (22).

Formidlet objektivitet

En fare ved bruk av tester er at de, som annet verktøy, i daglig bruk etter hvert mister sin fremskutte plass i vår bevissthet. Som hammeren i snekkerens hender glir testen inn i bakgrunnen, det som trer frem og blir vektlagt er det ferdige resultat (fig 1). Men tar vi testene for gitt og lar resultatene fra dem få et skinn av uformidlet objektivitet, risikerer vi samtidig å fjerne oss fra den biologi og den patofysiologi vi ønsker å klargjøre. Når testresultatene på falske premisser overtar styringen av våre resonnementer, blir vi ufrie i vår forstående tilnærming til sykdom. Når testresultatet får selvstendig liv og griper makten over rekvirenten, fremstår resultatet som et faktum som i verste fall styrer rekvirentens videre resonnementer. Dette oppleves når diagnostisk falskt positive resultater blir tolket som uttrykk for sykdom. Meningen med en laboratorietest er, i likhet med meningen med en hammer, konstituert av den situasjon de brukes i. Slik hammeren brukes for å besørge et byggverk, skal testen brukes for å besørge en diagnose. For å skape et effektivt byggverk må snekkerens og legens aktivitet være formålsrettet. Helheten og delene i byggverket må forstås gjennom hverandre, og denne forståelsen må relateres til målsettingen med den aktivitet som utføres. Ved å løsrive testresultatet fra en kontekst kan resultatet bli en ørkesløs utredning uten mål og mening. Kun ved å tilegne seg grundig praktisk og teoretisk kjennskap til testene og deres begrensninger kan legen gjenvinne herredømmet over testresultatet og dermed gjenopprette kontakt med det medisinske mangfold.



Figur 1 Det er feil å anta at man ved rekvisisjon av en test får tilbake et uformidlet svar vedrørende et biologisk fenomen. Informasjon som hentes fra testen vil være "filtrert" og må derfor alltid tolkes "gjennom" testens brukskarakteristika (sensitivitet og spesifisitet) og epidemiologisk kunnskap om angjeldende sykdom

Som en erkjennelse av at testresultater er å betrakte som formidlet medisinsk objektivitet, er det utarbeidet en rekke forslag til formalisert tolking av resultater fra tester og diagnostiske prosedyrer (23, 24). Disse metodene, basert på bayesiansk sannsynlighetsteori, undervises studenter regelmessig i (25). En grundig forståelse av testenenes begrensede informasjonsverdi er nødvendig, blant annet fordi diagnostiske tester sjelden er av en slik kvalitet at prøveresultatene gir entydige svar på om pasienten har eller ikke har en bestemt sykdom. Avhengig av epidemiologiske forhold ved en sykdom kan det i enkelte tilfeller endog være slik at negative resultater kan ha stor differensialdiagnostisk verdi, mens positive resultater nærmest er diagnostisk verdiløse, selv om det er de positive resultatene som er sterkest assosiert med sykdom (26). Blant annet på grunn av vanskelig tilgjengelig og lite relevant informasjon om testenenes brukskarakteristika har anvendelse av beslutningsteoretisk verktøy ikke fått allmenn utbredelse. I stedet har mange praktiserende leger utviklet en mer intuitiv tolking av tester, der egen erfaring med den enkelte test tillegges stor betydning (12). Det å etablere egne mål på testens brukskarakteristika vil imidlertid heller øke enn minke tolkingsproblemene, dette fordi man i en typisk brukssituasjon aldri kan få vurdert en tests validitet. I brukssituasjonen tester man diagnostiske hypoteser mot hverandre, testens sensitivitet og spesifisitet kan ikke vurderes i en slik situasjon og må derfor ansees for gitt.

Problemrettet versus dataorientert diagnostikk

Som et ledd i sin teoretiske utdanning lærer medisinerstudenter en logisk og sekvensiell tilnærming til diagnostikk – problembasert læring forfektes som et ideal (27, 28). Samtidig er studentenes praktiske opplæring preget av maler for skriving av pasientjournaler, noe som krever en systematisk og bred, men ikke problemrettet innsamling av data. I møtet med pasienten trenes studentene først opp til å innhente generelle anamnesticke opplysninger, deretter skal de foreta en grundig klinisk undersøkelse, og til sist, når alle data er for hånden, er det tid for å formulere og teste diagnostiske hypoteser. Misforholdet mellom teori og praksis blir tydeliggjort når studenter erfarer at drevne klinikere slett

ikke foretar en formålsløs datainnsamling, men nettopp praktiserer den problemrettede diagnostikk de proklamerer i sin undervisning. Av denne grunn har enkelte kritiske røster hevdet at man bør anvende Ockhams barberhøvel også på klinisk datainnsamling, og at man snarest bør kvitte seg med den dataorienterte tilnærming til diagnostikk (29).

Ved diagnostikk etter hypotetisk-deduktivt mønster søker legen å samle inn data som har relevans for pasientens problemstilling. Avhengig av hvilken informasjon dette gir, estimeres pretestsannsynligheter for sykdom, differensialdiagnoser overveies, og adekvate diagnostiske tester rekvireres. Etter hvert som laboratoriene har klart å levere raske og rimelige prøvesvar, har et alternativ til denne sekvensielle diagnostiske strategi utviklet seg. Ved første kontakt med pasienten rekvireres et bredt spekter av tester, for om mulig å innhente betydningsfulle resultater raskere, noe som igjen kan tenkes å spare pasienten for lidelse til en lavere total kostnad. Denne typen dataorientert strategi er delvis en videreføring av gammel klinisk-medisinsk tradisjon, men baserer seg også på en erkjennelse av at pasienters sykdomsmanifestasjoner kan være et resultat av ulike patofysiologiske prosesser samt at pasienter kan ha flere sykdommer samtidig.

Slike resonnementer til tross, det har vært vanskelig å vise at en dataorientert fremgangsmåte fører til økonomiske besparelser eller økt effekt av pasientbehandlingen målt som kortere liggetid, kortere tid som går med til å finne korrekt diagnose, oppdagelse av uventet sykdom eller redusert sykkelighet og dødelighet (30). Delvis kan dette skyldes at et flertall av de tester vi i dag anvender innen mikrobiologisk og immunologisk diagnostikk ikke er egnet for en dataorientert strategi, men også at leger som anvender denne strategien, ikke har kunnskap nok til å atskille relevante fra ikke-relevante resultater. Et tredje argument for å erstatte dataorienterte tilfældigheter med en problemrettet strategi er at legen dermed får mulighet til å evaluere sin diagnostiske praksis og slik kan lære av sine feiltakelser. Når det i tillegg kan hevdes at diagnostiske feiltakelser ofte skyldes manglende evne til å kombinere data på en adekvat måte (31), er det naturlig å sette et stort spørsmålsteget ved nytteverdien av en dataorientert diagnostisk strategi.

At tester i økende grad brukes som et ledd i en dataorientert diagnostisk prosess, har ført til en generell økning i forbruket av diagnostiske tester. Siden det ser ut til å være en årsakssammenheng mellom det å ta mange prøver og det å produsere ubegrunnede henvisninger og unødvendige innleggelser i sykehus (32), ville det vært av interesse å vite om det å ha en dataorientert tilnærming til diagnostikk, fører til "diagnoseglidning" og tilsynelatende økning av produksjonen i sykehus ved at pasienter får diagnoser som medfører høyere ressursmessig uttelling for sykehuset. Økningen betegnes som tilsynelatende, fordi mange av prøvene kan representere falsk positivitet og dermed ikke vil være uttrykk for sykdom.

Den økende bruk av diagnostiske tester er også muligens et utslag av et ønske om diagnostisk sikkerhet før medisinsk intervensjon (33). Dette ønsket kan neppe realiseres ved å øke testaktiviteten. Tvert imot, ved beslutningsanalyse kan man vise at den diagnostiske usikkerhet sjelden kan reduseres til null, det å øke antall tester vil paradoksalt nok kunne bidra til å øke risikoen for feil. I et

slikt perspektiv blir såkalt defensiv testing, det vil si testing som utføres for at legen skal sikre seg mot eventuelt rettslige etterspill, en klar målforskyvning. For leger kan defensiv testing være et middel for å sikre seg mot eventuelle emosjonelle, finansielle eller profesjonelle konsekvenser av beskyldninger om malpraksis. Et argument om at defensiv testing fører til økte omkostninger for helsevesenet, vil i en slik situasjon ha liten gjennomslagskraft. Når man derimot ved beslutningsanalyse kan vise at defensiv testing i enkelte situasjoner vil være til stor nytte for legen og legens forsikringsselskap, men av liten eller negativ nytte for pasientene, burde argumentet for å redusere testing på svake indikasjoner styrkes betraktelig (34). Dette eksemplet viser også at når legen som rasjonell aktør søker å oppnå størst mulig nytte, er det samtidig nødvendig å vurdere hvem han skal være til nytte for.

Feilbruk versus feilbarlig bruk av laboratorietester

Den økende bruk av mikrobiologiske og immunologiske tester skyldes neppe bare økt forekomst av epidemisk sykdom eller redusert helse i befolkningen, og det har vært spekulert på om økningen delvis kan skyldes økt forekomst av feilbruk. Feilbruk av tester har økonomiske og administrative konsekvenser ved at ressursene ikke utnyttes optimalt, men kan også få alvorlige medisinske konsekvenser ved at falskt positive resultater fører til ytterligere invasiv testing eller ved at feildiagnostisering fører til feilbehandling. Av denne grunn er det viktig å evaluere omfanget av og finne eventuelle årsaker til feilbruk av tester.

Feil i medisinsk forstand er et meget komplekst begrep, med både normative, empiriske, kognitive, emosjonelle og kontekstuelle aspekter. For å få en mer differensiert diskusjon rundt begrepet feil har jeg tidligere argumentert for at det er nyttig å skille mellom medisinsk feil og medisinsk feilbarlighet (35). Med medisinsk feil forstår jeg en hendelse som ikke stemmer overens med gitte spesifikasjoner og som kunne vært unngått, for eksempel feil forårsaket av en aktørs uoppmerksomhet eller manglende kunnskaper. Med medisinsk feilbarlighet forstår jeg derimot feil som av ulike grunner er uunngåelige, enten som følge av vitenskapelig utilstrekkelighet eller på grunn av medisinsens probabilistiske natur. Et tilsvarende skille mellom feilbruk av tester og testers feilbarlighet kan være nyttig. Feilbruk av en test er knyttet til brukeraspektet ved testen, til legens beslutning om å rekvirere en test i en gitt situasjon. Siden feilbruk skyldes beslutninger fattet av den enkelte lege, kan feilbruk justeres lokalt. Feilbarlighet av testene, derimot, er knyttet til resultatet som produseres, og skyldes tekniske, biologiske, epidemiologiske eller vitenskapelige forhold som ikke kan justeres lokalt.

Når vi skal diskutere muligheter for å redusere feilbruk av tester, er det altså brukeraspektet og ikke testresultatene og deres fortolkning vi diskuterer. Dette til tross, det er en nær relasjon mellom de to begrepene. Fra et epidemiologisk synspunkt vil for eksempel testbruk på feilaktige indikasjoner kunne resultere i lav testeffektivitet hvis prøver fra mange pasienter gir falskt positive eller falskt negative resultater. Ved dette mister testen sin troverdighet som undersøkelsesverktøy, det vil si at feilbarligheten øker.

Prediktiv sannsynlighetsberegning

For å begrense testbruk og hindre feilbruk krever Medicare dokumentasjon på at testen har relasjon til pasientens ICD-9-CM-kode (International Classification of Diseases, Ninth Edition, Clinical Modification) før utgifter til testing blir refundert. Medicare har ikke operasjonalisert begrepet feilbruk, men henviser til at bruken av testene skal være forstandig og nødvendig (36). Å sjekke relasjonen mellom testbruk og diagnose blir imidlertid en post festum-kontroll som vanskelig kan fange opp legens usikkerhet forut for diagnosen. Det er nettopp denne prediagnostiske usikkerhet som avgjør om en test skal rekvireres eller ikke.

Det å sjekke relasjonen mellom testbruk og diagnose blir en lite valid metode til måling av feilbruk. For å avgjøre om en test brukes feil, må man etter min oppfatning ta utgangspunkt i legens brukende omgang med testen. Dette gir mening fordi den enkelte test konstitueres som diagnostisk verktøy nettopp i brukssituasjonen. Brukssammenhengen vil dermed være med på å definere hvilken betydning vi kan tillegge den enkelte test og testresultatet, og dermed også for å vurdere om testen brukes riktig eller feil.

Evaluerings av testbruk gjennom denne form for prosesskontroll vil være kvalitetssikrende og vil kunne tilfredsstille krav til læring i egen praksis ved at den enkelte lege selv kan evaluere og definere sin diagnostisering som god eller dårlig ut fra gitte kriterier (37). Metoden kan imidlertid ikke brukes av eksterne kontrollinstanser, for eksempel trygdekontor eller tilsynsmyndigheter, for å vurdere om en lege bruker tester feil. Denne vurderingen er det kun rekvirerende lege selv som kan gjøre. Dette fundamentale skille mellom deltaker og tilskuer kan, slik jeg ser det, vanskelig overskrides.

En lege kan forut for rekvisisjon av en test anvende prediktive verdier for å besvare de to enkle spørsmålene "Hvordan bør jeg diagnostisere?" og "Hvordan er jeg i ferd med å diagnostisere?". Analyse av avvik mellom hvordan man bør teste og hvordan man intenderer å teste kan gi klarhet i om tester er i ferd med å anvendes feil. Et midlertidig problem ved denne form for evaluering er at verdier for de to viktigste karakteristika ved testen, sensitivitet og spesifisitet, i liten grad er tilgjengelige utenfor de store laboratoriene. På sikt kan imidlertid dette problemet løses ved hjelp av informasjonsteknologi. Vellykkede forsøk har vært gjort på å redusere unødvendig testing ved å supplere rekvirenter med prediktiv verdi gitt positivt eller negativt prøvesvar forut for rekvisisjon av tester (38). Etter hvert som IT-teknologien utvikles, vil de fleste legekontorer kunne kobles opp mot slik PC-basert informasjon og dermed få brukervennlig kunnskap om tester og deres egenskaper.

Et viktig fellestrekk ved bruk av tester er at de som verktøy skal skaffe til veie informasjon for medisinske beslutninger. For hver enkelt diagnostisk situasjon vil man kunne relatere testbruk til testformål og deretter beregne prediktive verdier for positive og negative prøvesvar. Informasjonsverdien fra tester vil således kunne måles i samme probabilistiske "valuta" og dermed sammenliknes over ulike beslutningssituasjoner. Feilbruk kan da defineres som rekvisisjon av

tester der den prediktive verdien for testen i brukssituasjonen er så lav at den ikke gir beslutningsbærende informasjon. For eksempel vil det være uakseptabelt å bruke HLA-B27-testen som diagnostisk markør for Bekhterevs sykdom hos en pasient med ryggsmarter. Slike pasienter har en pretestsannsynlighet for Bekhterevs sykdom på 4,4 %. Ved positivitet for HLA-B27 øker posttestsannsynlighet for sykdom til 30,4 %. En positiv prediktiv verdi på 30,4 % er ubrukelig i diagnostisk sammenheng, og testbruken må derfor karakteriseres som feil. Hvis derimot legen i samme kliniske situasjon ønsker å sannsynliggjøre at pasienten ikke har Bekhterevs sykdom og derfor rekvirerer HLA-B27-testen, vil han finne en negativ prediktiv verdi på 99,7 %. I dette tilfellet ville det være korrekt å rekvirere testen (26). Eksakt hvor grensen for feilbruk bør settes, eller om det i det hele tatt bør settes en grense som skal gjelde uavhengig av kontekst, kan diskuteres. For å rettferdiggjøre rekvisisjon av en test bør resultatet fra testen bidra til å øke rekvirentens beslutningsevne utover den informasjon han alt sitter inne med, og det kan derfor argumenteres for at prediktiv verdi i det minste bør

være høyere enn 50 % for å rekvirere enkelttester. Hvis testen rekvireres som et første ledd i en rekke med tester, vil imidlertid prediktiv verdi kunne være lavere enn 50 % og likevel gi meningsfull informasjon (26).

En feilbarlig metode

Ved hjelp av relevant informasjon og adekvate metoder kan sannsynlighet for fremtidige hendelser estimeres. I medisinsk praksis vil både informasjon og metode bygge på et vitenskapelig fundament. Medisinsk praksis skiller seg likevel fra vitenskapelig praksis ved at legen søker en partikulær sannhet der vitenskapsmannen søker en universell sannhet. Spissformulert kan vi med Wittgenstein hevde at vitenskapen gjennom sin ”streben mot det generelle” samtidig uttrykker en ”foraktende holdning overfor det partikulære kasus” (39). Det partikulære kasus betraktes i vitenskapelig sammenheng kun som et middel til å føre vitenskapsmannen frem til erkjennelse av den reale essens. Heri ligger også legens dilemma – som diagnostiker må han ved hjelp av Bayes’ teorem invertere vitenskapsidealet og på bakgrunn av det generelle uttale seg om det partikulære. Dette er tidvis en usikker streben, hyllet inn i probabilismens slør, og hvor risikoen for feilvurderinger er stor.

Når vi ved hjelp av prediktive verdier beregner en pasients sannsynlighet for sykdom, vil sannsynlighetsverdiene variere mellom 0 % og 100 %, kun unntaksvis vil ytterpunktene på skalaen tangeres. Dette kan virke kontraintuitivt for legen som søker å utrede sin pasient – enten har pasienten sykdommen eller så har hun den ikke, tertium non datur. Men så lenge det ikke er en deterministisk sammenheng mellom testresultat og sykdom, så lenge vi mangler relevant informasjon for beregning av pretestverdier og så lenge metodologiske mangler ved den underliggende forskning forekommer (40), vil beregning av prediktive verdier nødvendigvis bli usikre. I tillegg er det en innebygd svakhet ved metoden at den fordrer en direkte kobling mellom testresultat og diagnose. Ved å sette søkelyset på sluttresultatet og ikke på de patofysiologiske resonnementer som fører frem til sluttresultatet, øker risikoen

for feilslutninger. Selv om vitenskapen skulle vise at alle med en sykdom har en spesiell markør, vil det være en feilslutning å anta at alle med markøren har sykdommen. Dette fordi mange uten sykdommen også kan ha markøren, noe som vil bidra til falsk positivitet når det undersøkes for markøren i en uselektert befolkning. Denne usikkerheten kan i liten grad elimineres og vil derfor være en konstant trussel mot Descartes' rasjonalistiske visjon om absolutt erkjennelse.

Monitorering av sykdom

I motsetning til ved diagnostikk er det ved monitorering av sykdom mindre komplisert å klargjøre regler for hva som er å betrakte som feilbruk av tester. Dette kan gjøres ved å sammenholde faktisk testbruk med kunnskap om angjeldende sykdoms etiologi og patogenese. For eksempel gir det mening å snakke om feilbruk når en test rekvireres fra samme pasient flere ganger og så nær i tid at den underliggende patofysiologiske prosess vanskelig har kunnet endre seg. Den gjentatte testen gir da ingen ny beslutningsbærende informasjon. Ut fra en slik antakelse fant Bates og medarbeidere (41) at 71 % av gjentatte konsentrasjonsmålinger av gentamicin og 26 % av gjentatte dyrkinger for sputum, urin og avføring var overflødige. Tilsvarende erfaring fra eget laboratorium tyder på at et ikke ubetydelig antall undersøkelser er overflødige og derfor representerer feilbruk.

Avslutning

Det er i dag liten avstand mellom medisinsk forskning på den ene side og medisinsk diagnostikk og behandling på den annen – klinisk forskning og diagnostikk er ofte to sider av samme prosess. På grunn av denne nære relasjonen kan vi nå tilby pasienter mer målrettet diagnostikk enn vi kunne tidligere, men prisen for fremskrittene er økende risiko for feilbarlighet. Vi er her ved en grunnleggende etisk problemstilling – skal vi unnlate å undersøke pasienter fordi falskt positive eller falskt negative resultater kan oppstå, eller skal vi akseptere en viss usikkerhet hvis sannsynligheten for korrekt prøvesvar og dermed helbredelse er stor? Og helt fundamentalt, er det akseptabelt å teste for å lære? Det siste spørsmålet anskueliggjør det nære slektskapet mellom etikk og vitenskap ved medisinske beslutninger.

Et gjennomgangstema i denne artikkelen har vært usikkerheten som er knyttet til avgjørelsen om man skal teste eller ikke, og til usikkerheten som ligger i testenes resultater. Videre har jeg søkt å vise at allmenngyldige regler for rettledning av forstanden vanskelig kan la seg realisere i legers diagnostiske praksis. I den diagnostiske prosess skal vi, ut fra en refleksjon over kjente observasjonsdata, utlede noe vi ikke vet. Det er vanskelig å ha den klarhet som Descartes fordrer i forløpet av en slik prosess – etter som noen spørsmål blir besvart, dukker nye opp. Det å løse og skape problemer blir dermed ikke to atskilte deler av den diagnostiske prosess, men én. Den klassiske tro på en "oppgagelsens logikk" har fortsatt ikke latt seg realisere i moderne

vitenskapsteori – det finnes ingen maksimer som uunngåelig fører til oppdagelser. Ved alle oppdagelser er det et element av intuisjon og kreativitet som ikke kan forklares rasjonelt. Dette elementets kompleksitet kommer særlig tydelig frem ved de vanskeligheter man har støtt på i forsøk på diagnostisering ved hjelp av kunstig intelligens (42), og er en faktor som den kunnskapsbaserte medisinen vanskelig kan overskride. Intellektets sprang fra observasjon til hypotese betegnes som slutning til beste forklaring, og er en viktig, men upredikerbar del av både forskning og diagnostikk. Det er kanskje her vi sporer ”kunsten” i medisinen, dette udefinerbare som går utenpå det rent vitenskapelige og som av Descartes ble betegnet som vårt lumen naturale. Med dette får ordet diagnose, som er utledet fra det greske ”dia” (gjennom) og ”gnosis” (kunnskap), et videre betydningsinnhold enn hva den kunnskapsbaserte medisin kan tilby.

LITTERATUR

1. Allison JJ, Kiefe CI, Cook EF, Gerrity MS, Orav EJ, Centor R. The association of physician attitudes about uncertainty and risk taking with resource use in a Medicare HMO. *Med Decis Making* 1998; 18: 320 – 9.
2. Sox HC jr., Margulies I, Sox CH, Psychologically mediated effects of diagnostic tests. *Ann Intern Med* 1981; 95: 680 – 5.
3. Hampton JR, Harrison MJG, Mitchell JRA, Prichard JS, Seymour C. Relative contributions of history-taking, physical examination, and laboratory investigation to diagnosis and management of medical outpatients. *BMJ* 1975; 2: 486 – 9.
4. Forsman RW. Why is the laboratory an afterthought for managed care organizations? *Clin Chem* 1996; 42: 813 – 6.
5. Axt-Adam P, van der Wouden JC, van der Does E. Influencing behavior of physicians ordering laboratory tests: a literature study. *Med Care* 1993; 31: 784 – 94.
6. Epstein AM, McNeil BJ. Physician characteristics and organizational factors influencing use of amulatory tests. *Med Decis Making* 1985; 5: 401 – 15.
7. Dawson NV. Physician judgement in clinical settings: methodological influences and cognitive performance. *Clin Chem* 1993; 39: 1468 – 78.
8. Schroeder SA, Schliftman A, Piemme TE. Variation among physicians in use of laboratory tests: relation to quality of care. *Med Care* 1974; 12: 709 – 13.
9. Bradley GW. *Disease, diagnosis and decisions*. Chichester: Wiley, 1993.
10. Bergus GR, Chapman GB, Levy BT, Ely JW, Opplinger RA. Clinical diagnosis and the order of information. *Med Decis Making* 1998; 18: 412 – 7.

11. Heidegger M. Being and time. Oxford: Blackwell, 1962.
12. Upshur REG. Priors and prejudice. *Theor Med* 1998; 19: 59 – 72.
13. Sober E. From a biological point of view. Cambridge: Cambridge University Press, 1994: 50 – 70.
14. Gigerenzer G, Todd P. Fast and frugal heuristics. The adaptive toolbox. I: Gigerenzer G, Todd PM, and the ABC research group. Simple heuristics that make us smart. New York: Oxford University Press, 1999: 3 – 34.
15. Kassirer JP. Diagnostic reasoning. *Ann Intern Med* 1989; 110: 893 – 900.
16. Forster M, Sober E. How to tell when simpler, more unified, or less ad hoc theories will provide more accurate predictions. *Br J Phil Sci* 1994; 45: 1 – 35.
17. Forster MR. How do simple rules "fit to reality" in a complex world? *Minds Machines* 1999; 9: 543 – 64.
18. Schaffner KF. Exemplar reasoning about biological models in disease: a relation between the philosophy of medicine and philosophy of science. *J Med Philos* 1986; 11: 63 – 80.
19. Gorrry GA, Pauker SG, Schwartz WB. The diagnostic importance of the normal finding. *N Engl J Med* 1978; 298: 486 – 9.
20. Christensen-Szalanski JJJ, Busyhead JB. Physicians' misunderstanding of normal findings. *Med Decis Making* 1983; 3: 169 – 75.
21. Reid MC, Lande DA, Feinstein AR. Academic calculations versus clinical judgments: practicing physicians' use of quantitative measures of test accuracy. *Am J Med* 1998; 104: 374 – 80.
22. Poses RM, Randall DC, Collins M, Fager SS. The accuracy of experienced physicians' probability estimates for patients with sore throats. *JAMA* 1985; 254: 925 – 9.
23. Griner PR, Mayewski RJ, Mushlin AI, Greenland P. Selection and interpretation of diagnostic tests and procedures. Principles and applications. *Ann Intern Med* 1981; 94: 557 – 92.
24. Åsberg A. Diagnostiske tester – rekvirering og tolking. *Tidsskr Nor Lægeforen* 1993; 113: 604 – 8.
25. Johannessen T. Klinisk beslutningslære ved medisinstudiet i Trondheim. *Tidsskr Nor Lægeforen* 1998; 118: 921 – 4.
26. Ulvestad E. HLA-B27 ved Bekhterevs sykdom *Tidsskr Nor Lægeforen* 2000; 120: 1317 – 22.
27. Lycke KH. Problembasert læring – dokumenterte effekter og teoretisk forankring. *Tidsskr Nor Lægeforen* 1995; 115: 718 – 20.
28. Pettersen HB, Bovim G, Brodal P, Øgreid D, Fønnebø V. Medisinstudiet i Norge. Én felles utdanning, fire ulike studiemodeller. *Tidsskr Nor Lægeforen*

1997; 117: 2798 – 803.

29. Hoffbrand BI. Away with the system review: a plea for parsimony. *BMJ* 1989; 298: 817 – 9.
 30. Benson ES. Initiatives toward effective decision making and laboratory use. *Hum Pathol* 1980; 11: 440 – 8.
 31. Førde R. Competing conceptions of diagnostic reasoning – is there a way out? *Theor Med Bioeth* 1998; 19: 59 – 72.
 32. Malde K, Kvamme O, Ebbing H. Aksjon riktig takstbruk – storm i vannglass? *Tidsskr Nor Lægeforen* 1999; 119: 3804 – 7.
 33. Kassirer JP. Our stubborn quest for diagnostic certainty. A cause of excessive testing. *N Engl J Med* 1989; 320: 1489 – 91.
 34. DeKay ML, Asch DA. Is the defensive use of diagnostic tests good for patients, or bad? *Med Decis Making* 1998; 18: 19 – 28.
 35. Ulvestad E. Medisinsk feilbarlighet. *Bergens Tidende* 19.8.1998.
 36. Young DS, Bekeris LG. The inordinate complexity of delivering laboratory services in the USA. *Ann Clin Biochem* 1999; 36: 123 – 32.
 37. Carlsen T, Bratland SZ, Claudi T, Cooper J, Telje J, Waaler HM et al. Effektiv læring med data fra egen praksis – erfaringer fra SATS-prosjektet. *Tidsskr Nor Lægeforen* 1999; 119: 4306 – 9.
 38. Solomon DH, Shmerling RH, Schur PH, Lew R, Fiskio R, Bates DW. A computer based intervention to reduce unnecessary serologic testing. *J Rheumatol* 1999; 26: 2578 – 84.
 39. Wittgenstein L. The blue and brown books. 2. utg. New York: Harper & Row, 1960: 17 – 8.
 40. Reid MC, Lachs MS, Feinstein AR. Use of methodological standards in diagnostic test research. Getting better but still not good. *JAMA* 1995; 274: 645 – 51.
 41. Bates DW, Boyle DL, Rittenberg E, Kuperman GJ, Ma’Luf N, Menkin V et al. What proportion of common diagnostic tests appear redundant? *Am J Med* 1998; 104: 361 – 8.
 42. Josephson JR, Josephson SG, red. Abductive inference. *Computation, philosophy, technology*. Cambridge: Cambridge University Press, 1994.
-

Publisert: 20. august 2000. *Tidsskr Nor Lægeforen*.

© Tidsskrift for Den norske legeforening 2025. Lastet ned fra tidsskriftet.no 20. desember 2025.