# Kan vi stole på dopingtestingen?

I jakten på dopede idrettsutøvere tas det for lite hensyn til faren for å dømme en uskyldig. Dopingjegerne synes å være mer opptatt av at alle skal tas enn å vurdere risikoen for at noen av testresultatene kan være falskt positive.

En norsk idrettsutøver (Erik Tysse) testet i 2010 ifølge Laboratorio Antidoping FMSI i Roma positivt på det forbudte preparatet Cera (Continuous Erythropoietin Receptor Activator), som de mente å ha påvist i utøverens urin. Metoden som ble brukt, var en såkalt isoelektrisk fokusering (IEF). Både A-prøven og B-prøven var rapportert positive, begge først etter gjentatte analyser. Saken er behandlet av påtalenemnda i Norges idrettsforbund, som vedtok å anmelde den til domsutvalget. Der ble Tysse dømt, men han anket. Vi vil her komme med noen refleksjoner om dopingtesting generelt og om testingen i den spesifikke saken spesielt.

## Generelt om biologiske tester

Biologiske tester kan karakteriseres ved to atskilte egenskaper: repeterbarhet og validitet. Repeterbarhet uttrykker testmetodens evne til å frembringe samme resultat ved gjentatte målinger. Målinger er utsatt for to forskjellige feilkilder: de tilfeldige og de systematiske. Det er viktig at disse feilene, som ofte er uunngåelige, er så små som mulig.

Systematiske feil kan være uten betydning. Hvorvidt mansjettmålinger av det intraarterielle blodtrykket utført lege artis er beheftet med systematiske feil, har begrenset betydning, siden alle kliniske forsøk med risikoanalyser og behandlingseffektivitet er utført med mansjettmåling.

Ved dopingtesting forsøker man å ta hensyn til tilfeldige feil ved å analysere både en A-prøve og, dersom denne er positiv, en B-prøve. A-prøven og B-prøven er identiske, den opprinnelige prøven er bare fordelt på to glass. De to prøvene analyseres med samme testmetode, men enkelte ganger utføres det en tilleggsanalyse av B-prøven. Dette gir selvsagt et mangelfullt grunnlag for å estimere repeterbarheten. For et slikt formål er to prøver for lite. Videre undersøkes B-prøven med den bakgrunn at A-prøven er positiv. Dette gir en forventning om hva man skal finne, noe som kan gi en skjevhet. Som eksempel på dette kan nevnes at sjefen for laboratoriet i Roma angivelig skal ha uttalt at det ville bli en skandale om A-prøven ikke ble bekreftet med en positiv B-prøve. Bekreftet ikke B-prøven den positive A-prøven, ville 25 personer ved laboratoriet i Roma miste jobben (1).

Et annet aspekt er at gjentakelsen med en B-prøve bare er en kontroll for visse typer feil, nemlig de som går på laboratoriets evne til å utføre testen korrekt. Hvis årsaken til utslaget på testen ligger i at utøveren har spesielle verdier på grunn av forhold ved vedkommendes fysiologi eller næringsinntak, vil ikke en A-prøve og en B-prøve gi noen bedret presisjon. En eventuell forurensning eller forbytning av prøven vil heller ikke nødvendigvis bli plukket opp.

Det antas av dopingjegerne at dersom både A-prøven og B-prøven er positiv, er konklusjonen gitt – atleten er skyldig. Dette er prinsipielt utilfredsstillende. Det hjelper ikke med perfekt repeterbarhet dersom testen mangler validitet.

## Testens validitet

*Validiteten* dreier seg om hvorvidt testresultat sier noe om det man er ute etter. Sier et mammogram noe om forekomsten av brystkreft? Sier en tuberkulintest noe om en person er infisert med Mycobacterium tuberculosis?

Det som uttrykker validiteten, er sensitiviteten og spesifisiteten. *Sensitiviteten* uttrykker testens evne til å gi et positivt svar dersom den aktuelle sykdom eller doping er til stede eller har funnet sted. En sensitivitet på 0,9 vil bety at én av ti med sykdommen ikke blir observert/oppdaget. Disse kalles som kjent «falskt negative». *Spesifisiteten* uttrykker testens evne til å gi negativt svar for dem som ikke har sykdommen eller ikke er dopet. En spesifisitet på 0,9 vil bety at én av ti prøver vil konkludere med at sykdom/doping er til stede for individer som ikke er syke eller dopet. Disse kalles «falskt positive».

Det er av betydning at både sensitivitet og spesifisitet er så nær 100 % som mulig. Men det er viktig å erkjenne at forsøk på å øke sensitiviteten til en test vil redusere spesifisiteten og vice versa. Det finnes ingen a priori-optimalitet i balansen mellom disse. Den optimale balansen er avhengig av sykdommens eller tilstandens art og av det relative forhold av kostnadene av dem. En meget høy sensitivitet er viktig dersom det finnes en effektiv behandling av en alvorlig sykdom. I dopingarbeidet er det spesifisiteten som er viktig – om noen dopede atleter slipper unna er mindre farlig enn det traumet det er for en ikke-dopet å bli dømt. Det er derfor avgjørende for rettssikkerheten å ha en meget høy spesifisitet og at man kan dokumentere dette med statistisk tilfredsstillende empiri. Uten kunnskap om spesifisiteten er det grunnleggende umulig å konkludere om sykdom (in casu doping) er til stede.

Dersom f.eks. 90 % av de syke har en positiv test, kan man da slutte at 90 % av dem med positiv test er syke? Det er ikke sjelden å høre en slik slutning. Men den er feil, og vi demonstrerer sammenhengen i tabell 1. La oss anta følgende verdier: sensitivitet 0,9, spesifisitet 0,9, pretestsannsynlighet (prevalens) 0,01.

Med pretestsannsynlighet for sykdom menes en antatt prevalens av sykdommen i den befolkning man tester, en sannsynlighet som er basert på tidligere erfaringer og kunnskap. Med disse tallene vil, i en befolkning på 10 000 personer, 100 være syke. Av de syke er det 90 som har en positiv test. Av de friske er det 990 som har en positiv test. Av de 1 080 som har en positiv test, er det 90 som er syke – dvs. det er bare 8,3 % av dem med positiv test som er syke.

Hva skjer dersom vi antar en pretestsannsynlighet som er lavere, f.eks. 0,005? Tabell 2 viser at da reduseres testens prediksjonskraft fra 8,3 % til 4,3 %.

Sannsynligheten for at en positiv test indikerer sykdom, er altså avhengig av prevalensen av sykdommen i den befolkningen man tester. Det betyr f.eks. at sannsynligheten for at en gitt mammogramskygge indikerer brystkreft er forskjellig i en tilfeldig valgt gruppe av kvinner i 50-årsalderen og i en gruppe av kvinner som har en familiær opphopning av brystkreft.

## Aktuell dopingsak

I den aktuelle saken er det største problemet at vi ikke kjenner testens validitet. Cera-testens validitet er nylig estimert (2). Den viser klart samspill mellom sensitivitet og spesifisitet og demonstrerer en klar mulighet for falskt positive svar, avhengig av den valgte grenseverdi for definisjon av positivitet. Denne analysen gjelder imidlertid bruk av en ELISA-metode, en annen analysetype enn den som er brukt i det aktuelle tilfellet (IEF).

Den siktedes advokat har gjentatte ganger forsøkt å få validitetsverdier for den anvendte Cera-testmetoden, men dette har vært forgjeves (advokat Kjenner, personlig meddelelse). I et brev av 29.10. 2010 til advokat Kjenner fra professor Hemmersbach, direktør for det norske dopinglaboratoriet, står det: «WADA is not aware of any false positive case reported for CERA using the IEF method.» Det er et interessant svar for så vidt som at det er en erkjennelse av at selv ikke WADA kjenner Cera-testens validitet.

Ingen av referansene Hammersbach angir, behandler spørsmålet om falskt posi-

tive. Vi må derfor begrense oss til å simulere verdier. Verdiene for sensitivitet og spesifisitet er gitt i tabell 3. Pretestsannsynligheten er 0,001, det vil si at man forventer at én av 1 000 utøvere er dopet. Estimatet er basert på tall fra protokollen fra domsutvalget (sak 25/10) der det står: «Romalaboratoriet opplyser å ha foretatt ca 5000 analyser hvorav 6 med positive CERA prøver.» Verdiene i tabell 3 viser at det er under 50 % sannsynlighet for at en positiv test indikerer faktisk doping.

Det fremgår dessuten fra saksdokumentene at den siktede kunne dokumentere et såkalt normalt blodpass (3). Det er kjent at dersom en utøvers blodpass er suspekt, så er det alene en sterk indikator på faktisk doping. Det betyr at dersom testpopulasjonen består av utøvere med normalt blodpass, så vil pretestsannsynligheten for positivitet være betydelig lavere enn i en uselektert befolkning. Da vil prediksjonskraften være enda lavere. I tabell 3 eksempel 2 har vi simulert en antatt prevalens for en slik gruppe med atleter til 0,0005. Da vil prediksjonskraften være 32,2 %, som er langt fra et grunnlag for fellende dom.

La oss på den annen side forutsette en pretestsannsynlighet på 25 % (f.eks. sykkelryttere med suspekt blodpass). Tabell 3, eksempel 3: Her vil prediksjonskraften på 99,69 være tilstrekkelig for en fellende dom.

**Tabell 1** Sammenhengen mellom sensitivitet (0,9), spesifisitet (0,9), prevalens (0,01) og prediksjonskraft i en befolkning på 10 000

| | Test + | Test – | Totalt |
|---|---|---|---|
| Tilstand + | 90 | 10 | 100 |
| Tilstand – | 990 | 8 910 | 9 900 |
| Totalt | 1 080 | 8 920 | 10 000 |
| Prediksjonskraft | 0,083 | | |

**Tabell 2** Sammenhengen mellom sensitivitet (0,9), spesifisitet (0,9), prevalens (0,005) og prediksjonskraft i en befolkning på 10 000

| | Test + | Test – | Totalt |
|---|---|---|---|
| Tilstand + | 45 | 5 | 50 |
| Tilstand – | 995 | 8 955 | 9 950 |
| Totalt | 1 040 | 8 960 | 10 000 |
| Prediksjonskraft | 0,043 | | |

**Tabell 3** Sammenhengen mellom sensitivitet, spesifisitet, prevalens og prediksjonskraft. Aktuell dopingsak

| Eksempel | Sensitivitet | Spesifisitet | Prevalens | Prediksjonskraft |
|---|---|---|---|---|
| 1 Generell testpopulasjon | 0,95 | 0,999 | 0,001 | 0,487 |
| 2 Normalt blodpass | 0,95 | 0,999 | 0,0005 | 0,322 |
| 3 Suspekt blodpass | 0,95 | 0,999 | 0,25 | 0,9969 |

## Konklusjon

Når det gjelder den aktuelle saken må vi tilkjennegi uro: Det fremgår av domsutvalgets kjennelse at de ikke er sikre på sakens fakta og at deres avgjørelse er basert på en oppfatning av at laboratoriet har fulgt WADAs prosedyrer. Leder av utvalget, Lars Erik Frisvold, sa rett ut under de muntlige forhandlingene at mye av diskusjonen foregikk over hodet på ham. Er det forsvarlig for rettssikkerheten at domsutvalget, uten egen kompetanse, blindt godtar WADAs prosedyrer?

For dopingtesting som for alle andre biologiske tester er kunnskap om testens validitet grunnleggende. Validiteten uttrykkes ved sensitivitet og spesifisitet. Om det er sensitiviteten eller spesifisiteten som er viktigst, avhenger av hvilken tilstand man ved hjelp av testen forsøker å avdekke. For dopingtesting er begge viktige. Det er viktig å fange alle juksemakere, men det er også viktig å unngå å dømme uskyldige. Det er kanskje bedre å la ti dopede gå fri enn å dømme én uskyldig.

Vi har i denne artikkelen konsentrert oss om doping med Cera og har måttet konkludere med at kontrollapparatet har trukket sine slutninger uten å kunne dokumentere testens validitet. Det virker som det a priori antas at spesifisiteten er 100 % for en kombinasjon av A- og B-prøvene, noe som ville utelukke mulighetene for falskt positive svar. Vi har vist at manglende dokumentasjon av testens validitet gjør en fellende dom urimelig og kan medføre justismord. Nødvendigheten av dokumentasjon av testers validitet gjelder for alle dopingtester – for å sikre atletenes rettssikkerhet.

**Hans Th. Waaler**
*hanswaaler@gmail.com*
**Harald Siem**
**Odd O. Aalen**

Hans Th. Waaler (f. 1926) er dr.philos. og professor emeritus.
**Ingen oppgitte interessekonflikter.**

Harald Siem (f. 1941) er lege, master of public health og ansatt ved Avdeling for global helse, Helsedirektoratet. Han har vært distriktslege på Aukra, har arbeidet ved Institutt for allmennmedisin ved Universitetet I Oslo, i Oslo helseråd, Arbeidsgiverforeningen og i ti år med internasjonalt helsearbeid i Genève.
**Ingen oppgitte interessekonflikter.**

Odd O. Aalen (f. 1947) er professor ved avdeling for biostatistikk ved Universitetet i Oslo.
**Ingen oppgitte interessekonflikter.**

**Litteratur**
1. Protokoll fra domsutvalget i Norges idrettsforbund og Norges olympiske og paralympiske komité. Sak 25/10. Oslo: Norges idrettsforbund, 2010.
2. Lamon S, Giraud S, Egli L et al. A high-throughput test to detect C.E.R.A. doping in blood. J Pharm Biomed Anal 2009; 50: 954–8.
3. ADN (Anti-Doping-Norge). Sak 25/10: ADN mot Erik Tysse. Oslo: Anti-Doping-Norge, 2010.

*Mottatt 18.5. 2011, første revisjon innsendt 16.6. 2011, godkjent 18.8. 2011. Medisinsk redaktør Anne Kveim Lie.*

*Engelsk oversettelse av hele artikkelen på www.tidsskriftet.no*