

**S.C. Høymork svarer:**

Torgeir Brun Wyller har nok ikke helt oppfattet hovedbudskapet i min lederartikkel (1): Det finnes behandling som ser ut til å ha god effekt hos enkeltpasienter, men som viser seg ikke å holde mål eller sågar påføre mer skade enn gagn når man går den nærmere etter i sømmene. Jeg trakk fram to eksempler som illustrerer dette, nemlig aktivert protein C til sepsispasienter og laserbehandling ved angina pectoris.

Selv har jeg nærmere tjue års bakgrunn fra klinisk arbeid som anesthesiolog og intensivlege. Jeg var selv blant entusiastene da aktivert protein C ble tilgjengelig. Entusiasmen skyldtes at jeg selv hadde behandlet én pasient, der et dystert klinisk bilde snudde etter at vi ga medikamentet. Jeg undres hvordan Bruun Wyller mener vi som fagmiljø skulle forholdt oss på dette tidspunktet i medikamentets historie. Mon tro han mener at vi skulle fortsatt ufortrødent å gi slik behandling på bakgrunn av enkeltobservasjoner og ikke samtidig etterspurt eller bidratt til å framskaffe den kunnskapen som vi seinere fikk, og som viste at behandlingen dessverre ikke hadde den effekten vi håpet på?

Mine to litt karikerte legehelter i henholdsvis avsnitt 1 og 2 har begge sine svakheter, slik også lederartikkelen illustrerer. Faglige retningslinjer kan aldri bli så omfattende og finmasket at de dekker enhver klinisk situasjon, og for noen fagområder er de også lite utviklet. Den store innsatsen som legges ned i flere fagmedisinske foreninger og andre fagmiljøer i å utarbeide ulike faglige retningslinjer, tyder imidlertid på at mange klinikere ser nytten av disse som støtte for klinisk arbeid.

**Siv Cathrine Høymork**  
sch@kunnskapssenteret.no

Siv Cathrine Høymork (f. 1965) er avdelingsdirektør i Nasjonalt kunnskapssenter for helsetjenesten.  
Ingen oppgitte interessekonflikter.

**Litteratur**

1. Høymork SC. Legehelter – og antihelter. Tidsskr Nor Legeforen 2015; 135: 1336.

**Re: Vi bør slutte å bruke ordet dement**

«Innen fagfeltet demens vil man nesten være tilbøyelig til å se på helsearbeidere som halvstuderte røvere dersom de omtaler en person som «dement»,» påstås det i språkspalten i Tidsskriftet (1). Er helsearbeidere som omtaler personer/pasienter som at han/hun er bipolar, schizofren, deprimert eller liknende, også «halvstuderte røvere»?

**Siri Lill Thowsen**  
siri.lill@yahoo.no

Siri Lill Thowsen (f. 1940) er pensjonist.  
Ingen oppgitte interessekonflikter.

**Litteratur**

1. Aga O. Vi bør slutte å bruke ordet dement. Tidsskr Nor Legeforen 2015; 135: 1572.

**O. Aga svarer:**

Takk for kommentar. La meg først slå fast at jeg ikke har noen problemer med å gi slipp på den noe provoserende «røver»-metaforen, verken når det gjelder de sykdommer Thowsen nevner eller for andre diagnoser. Mitt poeng er at som en hovedregel er det ingen pasienter som «er» sin sykdom, men derimot «har man», «lider man av», «er man rammet av» eventuelt «har man pådratt seg» og i noen tilfeller «er man blitt smittet av» en sykdom.

Samtidig benytter jeg denne anledningen til å gjøre oppmerksom på en uheldig glipp i mitt opprinnelige innleggs (1) siste avsnitt, der det mangler anførselstegn i forbindelse med ordet «epileptikere». Grunnen til at denne pinligheten er oppstått, er at for å spare meg selv for tid og spare meg for noen ekstra tastetrykk, kopierte jeg «epileptikere» fra det foregående avsnitt der ordet står i anførselstegn: «Temaet er omtalt i Tidsskriftet tidligere, bl.a. av nevrologen Karl O. Nakken som påpekte at man ikke bør snakke om «epileptikere», men om mennesker med epilepsi...». Når jeg så kopierte ordet, fikk jeg uheldigvis ikke med meg anførselstegnene, og dermed var det pinlige skjedd at jeg selv kom i skade for å benytte betegnelsen epileptikere som jeg noen linjer ovenfor har sitert Karl O. Nakken på som et ord som bør unngås.

Hadde jeg vært våken nok og ikke så opphengt i å «gjøre ting raskt og å spare tastetrykk», burde jeg i siste avsnitt naturligvis passet på at anførselstegnene kom med eller kanskje aller helst skrevet: «De aller fleste epilepsipasienter fungerer så bra kognitivt at de vil kunne sette ord på dette,...».

**Olav Aga**  
olavaga@broadpark.no

Olav Aga (f. 1938) er tidligere sykehjemsoverlege.  
Ingen oppgitte interessekonflikter.

**Litteratur**

1. Aga O. Vi bør slutte å bruke ordet dement. Tidsskr Nor Legeforen 2015; 135: 1572.

**Re: Samhandlingsreformen – hva nå?**

«Evaluering av samhandlingsreformen viser at det fortsatt er et stykke igjen før målene kan sies å være nådd,» skriver A. Grimsmo i en leder i Tidsskriftet (1). Samhandlingsreformen fører til store endringer i hvordan pasienter behandles. Helsedirektoratet har presentert data for reinnleggelser for pasienter fra sykehus som er tatt imot av kommunene (2), og jeg lurer på om noen har målt om det er endringer i mortaliteten 30 dager etter akutt innleggelse i sykehus for den samme gruppen?

**Ole Martin Steihaug**  
osteihaug@gmail.com

Ole Martin Steihaug (f. 1978) er lege i spesialisering i geriatri og forsker ved Haraldsplass sykehus.  
Ingen oppgitte interessekonflikter.

**Litteratur**

1. Grimsmo A. Samhandlingsreformen – hva nå? Tidsskr Nor Legeforen 2015; 135: 1528.
2. Helsedirektoratet. Samhandlingsstatistikk 2016–2014. <https://helsedirektoratet.no/Lists/Publikasjoner/Attachments/798/Samhandlingsstatistikk-2013-2014-IS-2245.pdf> [25.9.2015].

**A.Grimsmo svarer:**

Ole Martin Steihaug tar opp et viktig spørsmål. Vi har gode eksempler i Norge på at god planlegging av utskrivning kan redusere reinnleggelser (1). Det samme er tilfelle med et strukturert mottak av pasienter i primærhelsetjenesten (2). Reinnleggelser blir derfor ofte brukt som et kvalitetsmål. Årsakene til reinnleggelser er imidlertid komplekse (3). Man kan derfor ikke se på reinnleggelser alene, uten samtidig å ha oversikt over dødelighet i samme periode. Norge har den høyeste reinnleggelsesraten i Norden, men har også lav dødelighet etter utskrivning. Om det her er en sammenheng vet vi ikke. Når det gjelder økningen i reinnleggelser som Helsedirektoratet har registrert, så vil en forskergruppe som er ledet av professor Terje P. Hagen ved Avdeling for helseledelse og helseøkonomi på Universitetet i Oslo (4) kunne sette ord på dette,...

&gt;&gt;&gt;

tetet i Oslo, utgi statistikk senere i høst som også tar for seg dødeligheten etter utskrivning i samband med samhandlingsreformen.

**Anders Grimsmo**  
anders.grimsmo@ntnu.no

Anders Grimsmo (f. 1950) er professor ved Institutt for samfunnsmedisin, NTNU.

Ingen oppgitte interessekonflikter.

#### Litteratur

1. Prestmo A, Hagen G, Sletvold O et al. Comprehensive geriatric care for patients with hip fractures: a prospective, randomised, controlled trial. *Lancet* 2015; 385: 1623–33.
2. Garåsen H, Windspoll R, Johnsen R. Long-term patients' outcomes after intermediate care at a community hospital for elderly patients: 12-month follow-up of a randomized controlled trial. *Scand J Public Health* 2008; 36: 197–204.
3. Krumholz HM. Post-hospital syndrome—an acquired, transient condition of generalized risk. *N Engl J Med* 2013; 368: 100–2.

## Re: Et uunnværlig verktøy

I denne lederartikkelen om statistikk (1) gjentar Eva Skovlund synpunkter fra en leder hun skrev i 2013 (2). To spørsmål jeg stilte henne i en artikkelkommentar (2) til lederen fra 2013 er fremdeles ubesvart: 1) Hvorfor gir 100 tester med ett spørsmål i hver test riktigere resultater enn 100 spørsmål i en test? 2) Hva er den prinsipielle forskjellen i den informasjon p-verdier og konfidensintervaller (CI) gir? I en kronikk om p-verdier (3) refererer Are Hugo Pripp til en diskusjon om dette siste (4, 5). Argumentet synes å være at bredden på CI, i motsetning til avstanden mellom dem (det er avstanden eller overlappingen som tilsvarende p-verdier), skal gi ekstra informasjon som p-verdien ikke har.

Problemet er imidlertid at bredden på CI (og følgelig også avstanden mellom CI) varierer med antall observasjoner. Jo flere observasjoner desto smalere CI, større avstand mellom CI og lavere p-verdier. Ønsker man å fremme et gitt budskap, kan dette påvirkes ved å justere antall observasjoner. På forhånd å beregne antall observasjoner man trenger for å vise, for eksempel om en behandling har effekt (å oppnå en p-verdi  $< 0,05$ ), er en akseptert fremgangsmåte, men det er egentlig å fiske etter et gitt resultat.

P-verdier og avstanden mellom CI forteller oss ikke noe om to forhold som vi trenger for å trekke praktiske konklusjoner, nemlig hvor sterk en effekt er og hvordan variasjonen i behandlingseffekten er fra person til person (spredning). P-verdier sier oss bare at det sannsynligvis er en effekt, men ikke hvor sterk den er. Videre er den, slik vi nå gjør våre analyser, kun knyttet til gjennomsnittet, ikke til variasjonen. Variasjonen i behandlingseffekt, også om den har en unormal fordeling, kan være like viktig å kjenne til som den gjennomsnittlige effekten.

Man kan hevde at bredden på CI gir ekstra informasjon ved å si noe om presisjonen på det estimerte gjennomsnittet, en opplysning som kan være av betydning. Men, denne parameter må ikke forveksles med informasjon om behandlingseffektens variasjon, en feiltolkning som jeg tror er vanlig (og som kanskje kynisk utnyttes av noen kunnskapsrike forfattere).

Disse statistiske parametere gir oss altså begrenset informasjon om det virkelige livet, det vil si om forhold som vi trenger å kjenne til for å fatte beslutninger og å gi pasientene anbefalinger. Jeg har merket meg at man heldigvis begynner å gi disse bearbejdede, teoretiske, statistiske estimater mindre betydning ved at plots med CI fortreges av «box.plots», «bee-swarm plots», eller en kombinasjon av «box plots» og, for eksempel, søylediagrammer. Disse viser oss de målte resultater, det vil si at vi får «the whole complete information, nothing is hidden, you see the sample size, the distribution, possible problems/outliers...everything» (6).

Jeg vil hevde at vårt manglende krav til presentasjon av «det virkelige livet» gir mulighet til å selge (nær) verdiløse helsekostprodukter og også til å skremme med bagatellmessige risikofaktorer: I en reklame for bruk av vitamin K for å bevare benhelsen vises til et

arbeide hvor det er vist signifikant mindre tap av beinmasse (BMD) ved tilskudd av vitamin K (7). Den faktiske forskjellen mellom gruppene med og uten dette tilskuddet var imidlertid minimal, sannsynligvis helt ubetydelig, men statistisk signifikant takket være et relativt stort antall observasjoner. Slik sett er påstanden i reklamen korrekt, men misvisende.

**Arne Høiseith**  
arnhois@online.no

Arne Høiseith (f. 1944) er konsulent ved Curato røntgeninstitutt. Ingen oppgitte interessekonflikter.

#### Litteratur

1. Skovlund E. Et uunnværlig verktøy. *Tidsskr Nor Legeforen* 2015; 135: 1424.
2. Høiseith A. Spør først, regn siden. *Kommentar. Tidsskr Nor Legeforen* 2013; 133: 10. <http://tidsskriftet.no/article/2949352> [12.10.2015].
3. Pripp AH. Hvorfor p-verdien er signifikant. *Tidsskr Nor Legeforen* 2015; 135: 1462–4.
4. Mitchell MS, Yu MC, Whiteside TL. The tyranny of statistics in medicine: a critique of unthinking adherence to an arbitrary p value. *Cancer Immunol Immunother* 2010; 59: 1137–40.
5. VanderWeele TJ. Re: The ongoing tyranny of statistical significance testing in biomedical research. *Eur J Epidemiol* 2010; 25: 843–5, author reply 844–5.
6. Wilhelm Jochen. How to handle Narrow Confidence Intervals? *Research Gate* 22.4.15. [www.researchgate.net/post/How\\_to\\_handle\\_Narrow\\_Confidence\\_Intervals](http://www.researchgate.net/post/How_to_handle_Narrow_Confidence_Intervals) [22.9.2015].
7. Knapen MH, Drummen NE, Smit E et al. Three-year low-dose menaquinone-7 supplementation helps decrease bone loss in healthy postmenopausal women. *Osteoporos Int* 2013; 24: 2499–507.

## E. Skovlund svarer:

Det viktigste poenget med å utføre signifikanstester er etter mitt skjønn at vi ikke skal overtolke våre observasjoner. Vi stiller spørsmålet «hva er sannsynligheten for å observere det resultatet vi ser, eller en enda større effekt, gitt at nullhypotesen (for eksempel at det ikke er en forskjell i effekt av to behandlinger) er sann?». Hvis denne sannsynligheten (p-verdien) er stor, er det grunn til å mistenke at en observert forskjell ikke er uttrykk for sann effekt. Dersom p-verdien er liten, peker det i retning av at vi har observert en reell effekt, gitt at behandlingsgruppene er sammenlignbare.

Høiseith spør nokså upresist hvorfor 100 tester med ett spørsmål i hver test gir riktigere resultater enn 100 spørsmål i en test. Hver enkelt statistisk test man utfører forsøker å gi svar på ett spørsmål. Utvalget i en studie kan være skjevt og lite representativt for populasjonen man ønsker å studere. Dersom 100 forskningsspørsmål blir forsøkt besvart basert på det samme skjeve utvalget, vil denne svakheten kunne ramme mange av konklusjonene man trekker. Uavhengige forsøk er derfor av stor verdi. Innlegget «Data torturing» (1) presenterer for øvrig både problemer med multiple signifikanstester og andre fallgruver knyttet til presentasjon av forskningsresultater på en utmerket og forståelig måte.

Konfidensintervaller hjelper oss å kvantifisere usikkerhet og inneholder det vi kan kalle plausible verdier av sann effekt. De er nært beslektet med p-verdier, men gir viktig tilleggsmessig informasjon fordi vi estimerer størrelsen av en eventuell effekt. Dermed kan vi avgjøre om effekten er stor nok til at den har klinisk betydning. Bruker vi grensene i intervallet til å trekke slutninger om statistisk signifikans, har de selvfølgelig samme svakheter som p-verdier.

Både antall observasjoner og variabilitet (spredning) er viktige for bredden av et konfidensintervall. Jo flere observasjoner vi har, desto smalere blir intervallet. Økt presisjon betyr ikke juks – en våken leser vil klare å avdekke at en gjennomsnittlig endring i blodtrykk på 0,5 mmHg med et 95 % konfidensintervall som strekker seg fra 0,3 til 0,7 neppe har klinisk relevans selv om endringen er statistisk signifikant ( $p < 0,05$ ). Det er her konfidensintervallet viser sin verdi. Vi ser med en gang at gjennomsnittseffekten er svært liten, men den er i dette eksemplet presist estimert, og vi kan selv vurdere hvorvidt den er stor nok til å ha klinisk betydning.

Vi er for øvrig åpenbart ikke uenige om at det forekommer et tankeløst overforbruk av p-verdier i medisinsk forskning. Men det er etter min oppfatning misbruket som fortjener kritikk, ikke

>>>