

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 [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

>>>

metoden som sådan. Riktig brukt gir p-verdier og konfidensintervaller nyttige bidrag til innsikt.

Eva Skovlund
eva.skovlund@ntnu.no

Eva Skovlund (f. 1959) er professor ved NTNU. Ingen oppgitte interessekonflikter.

Litteratur

1. Mills JL. Data torturing. *N Engl J Med* 1993; 329: 1196–9.

Re: En kvinne i 60-årene med kløe på armene

I denne kasuistikken om en pasient med brakioradial kløe (1), refererer forfatterne til enkeltrapper som har vist effekt av akupunktur (2). Jeg vil i den anledning også gjøre oppmerksom på en systematisk gjennomgang som ser på akupunktur og kløe (3) og finner holdpunkter for at akupunktur kan ha en plass i behandling av kløe.

Merete Lindén Dahle
merete@klinikkfrogner.no

Merete Lindén Dahle (f. 1966) er akupunktør ved Klinikk Frogner. Ingen oppgitte interessekonflikter.

Litteratur

1. Halvorsen JA, Ørstavik K. En kvinne i 60-årene med kløe på armene. *Tidsskr Nor Legeforen* 2015; 135: 1459–61.
2. Marziniak M, Ständer S. Brachioradial Pruritus. I: Misery L, Ständer S, red. *Pruritus*. Berlin: Springer Verlag, 2010.
3. Yu C, Zhang P, Lv ZT et al. Efficacy of acupuncture in itch: a systematic review and meta-analysis of clinical randomized controlled trials. *Evid Based Complement Alternat Med*. 2015; 2015: 208690.

J.A. Halvorsen & K. Ørstavik svarer:

Behandling av kløe bør ta utgangspunkt i kløens årsak. Kløen kan være ledd i en dermatologisk eller systemisk sykdom og ha ulike patogenetiske mekanismer. De tre studiene som er inkludert i den omtalte metaanalysen (1), hadde totalt 35 pasienter med uremisk kløe eller atopisk eksem i tillegg til kontrollene. Brakioradial kløe har en annen patogenese, muligens av nevrologisk art. Det er derfor ikke dokumentert at akupunktur har effekt ved brakioradial kløe.

Jon Anders Halvorsen
j.a.halvorsen@medisin.uio.no
Kristin Ørstavik

Jon Anders Halvorsen er ph.d. og spesialist i hud- og veneriske sykdommer, overlege og seksjonsleder ved Seksjon for hudsykdommer, Oslo universitetssykehus. Ingen oppgitte interessekonflikter.

Kristin Ørstavik er dr.med. og spesialist i nevrologi og i klinisk nevrofysiologi, overlege og seksjonsleder ved Nevrologisk avdeling, Oslo universitetssykehus. Ingen oppgitte interessekonflikter.

Litteratur

1. Yu C, Zhang P, Lv ZT et al. Efficacy of acupuncture in itch: a systematic review and meta-analysis of clinical randomized controlled trials. *Evid Based Complement Alternat Med*. 2015; 2015: 208690.

Re: Alle selvmord skyldes ikke behandlingssvikt

I et «Brev til redaktøren» i Tidsskriftet, har Espen Øvre en kommentar knyttet til Norsk pasientskadeerstatnings bruk av sakkyndige ved vurderingen av selvmordsaker i psykisk helsevern (1).

Øvre og hans familie har opplevd en stor tragedie, og det er trist og gjør naturligvis sterkt inntrykk å lese om deres opplevelser. Jeg kan ikke gå inn på den saken her, men jeg ønsker å nyansere Øvres bilde av hvordan Norsk pasientskadeerstatning (NPE) jobber.

Saker som omhandler selvmord er blant de mest alvorlige vi behandler. Vi legger derfor stor vekt på en grundig håndtering av sakene og saksbehandlere med bred erfaring og kompetanse. De sakkyndige som vurderer de medisinskfaglige sidene av sakene er spesialister på fagområdet. Vurderingene de gjør er basert på faglig kunnskap, ikke egne følelser og meninger, slik Øvre antyder. Øvre har rett i at de sakkyndige vurderingene er sentrale premisser for avgjørelsene. Men det er NPE som vurderer de erstatningsrettslige spørsmålene og fatter vedtak. Det skjer etter at vi også har vurdert annen dokumentasjon, som for eksempel kommentarer og innspill fra erstatningssøkeren og behandlingsstedet.

Øvre er kritisk til at ikke flere sakkyndige vurderer hver sak. NPE mottar nærmere 6000 saker i året, og de aller fleste av disse blir vurdert av minst én sakkyndig. Ressurshensyn gjør at ordningen er ikke rigget slik at hver enkelt sak kan bli vurdert av flere sakkyndige. Det skjer imidlertid dersom saken omhandler flere medisinske spesialiteter eller de medisinske spørsmålene fremstår som spesielt kompliserte.

Øvre skriver at flere som har søkt om erstatning forteller at de opplever å ikke bli hørt. Slik skal det ikke være. Vi er opptatt av en god dialog med erstatningssøker underveis i prosessen. Dersom vi ikke tar kommentarer og innspill fra erstatningssøkeren til følge, betyr ikke det at vi ikke har vurdert innholdet i dem.

Det kan være vanskelig å forstå at tilsynsmyndighetene kan fatte en avgjørelse om svikt i behandlingen, mens NPE kommer til at det ikke er grunnlag for erstatning. Det kan imidlertid henge sammen med at tilsynsmyndighetene og NPE forvalter ulike lovverk og har forskjellige oppgaver. NPE vurderer erstatningskrav etter pasientskadeloven. Et vilkår for å tilkjenne erstatning er ikke bare at det foreligger en svikt i behandlingen, men at denne svikten med overvekt av sannsynlighet har ført til en pasientskade. Tilsynsmyndighetene foretar normalt ikke slike årsaksvurderinger. Derfor er det ikke slik at vi tolker de samme lovene ulikt, slik Øvre skriver. Påpekning av feil og mangler fra tilsynsmyndighetenes side vil altså ikke uten videre lede til at NPE tilkjenner erstatning.

Det Øvre tar opp i innlegget sitt, er vi interessert i å høre mer om. Jeg har derfor invitert han til et møte med oss der vi kan snakke mer om dette.

Rolf Gunnar Jørstad
rgj@npe.no

Rolf Gunnar Jørstad (f. 1958) er direktør Norsk pasientskadeerstatning. Ingen oppgitte interessekonflikter.

Litteratur

1. Øvre E. Re: Alle selvmord skyldes ikke behandlingssvikt. *Tidsskr Nor Legeforen* 2015; 135: 1531–2.

RETTELSE

Fet, feit eller bare overvektig?

Magnus Strømmen, Inger Johanne Bakken, Ellen Andenæs, Christian A. Klöckner, Ronald Mårvik, Bård Kulseng, Are Holen

Tidsskr Nor Legeforen 2015; 135: 1732–6

I Tidsskriftet nr. 19/2015 skal det på s. 1734 i siste setning i første avsnitt under Diskusjon stå: Som sagt var det kvinnene som var mest uttrykks-sensitive, og de som ble overvektige tidlig i livet, de med høyere utdanning og de som var utilfreds med vekten.

Vi beklager feilen, den er rettet på nett.