



Tidsskriftet  
DEN NORSKE LEGEFORENING

# Skal vi skrinlegge begrepet statistisk signifikans?

---

## MEDISIN OG TALL

MATS JULIUS STENSRUD

E-post: [m.j.stensrud@medisin.uio.no](mailto:m.j.stensrud@medisin.uio.no)

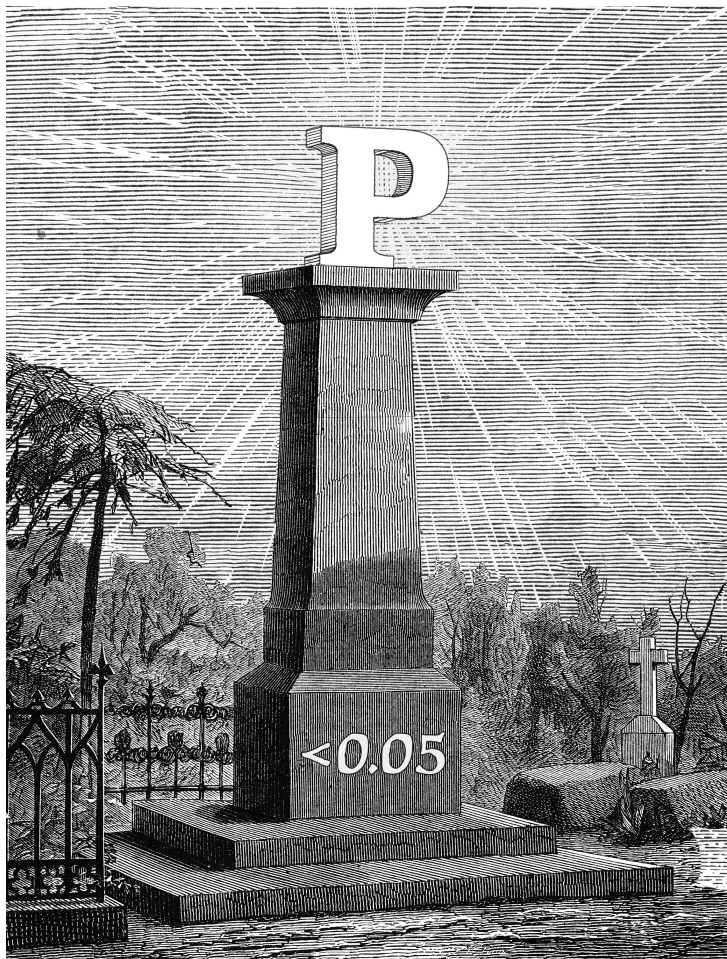
Mats Julius Stensrud er lege, dr.philos. og statistiker. Han er Fulbright-stipendiat ved Harvard School of Public Health og jobber ellers som postdoktor ved Avdeling for biostatistikk, Universitetet i Oslo. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

ODD O. AALEN

Odd O. Aalen er professor emeritus i statistikk ved Avdeling for biostatistikk, Universitetet i Oslo. Forfatteren har fylt ut ICMJE-skjemaet og oppgir ingen interessekonflikter.

---

Leger gjør beslutninger på usikkert grunnlag. Statistikk er det formelle verktøyet for å belyse usikkerheten, og begrepet *statistisk signifikans* brukes mye i den medisinske faglitteraturen. Men nå går hundrevis av forskere til opprop for å få begrepet fjernet.



Illustrasjon: clu/iStock, tilpasset av Tidsskriftet

Formålet med begrepet statistisk signifikans er å beskytte mot falske resultater og å sikre reproduserbarhet. I praksis brukes statistisk signifikans gjerne for å beskrive en hypotesetest som ga en p-verdi mindre enn 0,05. Dette er for øvrig en vilkårlig grense som like gjerne kunne vært lavere, for eksempel 0,01.

## Kobrafenomenet

Det er få som er uenig i at falske resultater og manglende reproduserbarhet er et problem. Likevel er det flere som mener at den omfattende bruken av begrepet statistisk signifikans er en dårlig løsning (1, 2). Noen sammenlikner det med et *kobrafenomen*, og refererer til det britiske imperiet: Britene var bekymret over de mange giftige kobraslangene i Dehli, India, og vedtok derfor å gi pengepremier i bytte mot døde kobraer for å redusere bestanden. I stedet for å ta livet av de ville kobraene var det mange som tjente store penger på å avle frem kobraer. «Løsningen» gjorde altså situasjonen verre, fordi insentivet ble misbrukt.

## Dikotomani og p-fisking

Et nøkkelproblem med statistisk signifikans er den såkalte *dikotomanien*: p-verdier brukes ofte til å lage kunstige todelinger, slik at en p-verdi mindre enn 0,05 tolkes som en reell effekt og en p-verdi større enn 0,05 tolkes som ingen effekt. Dette er en klar misforståelse av p-verdien. I det hele tatt skyldes nok mange av problemene med begrepet statistisk signifikans at p-verdier feiltolkes (3), også blant redaktører og fagfellevurderere i ledende medisinske tidsskrift.

Forhåpentligvis fører opprøret til at redaktører i mindre grad bruker statistisk signifikans som kriterium for å publisere artikler

Et relatert problem er den såkalte *p-fiskingen* (p-hacking): Noen forskere analyserer dataene sine på flere forskjellige måter, men oppgir bare analyser som gir *signifikante* p-verdier.

Dette fører selvsagt til overestimering av effektstørrelser og inflasjon i usikkerhetsmål. Vi kan redusere dette problemet ved å registrere alle analyser før studier blir utført. Likevel er det et tilleggsproblem at mange tidsskrifter bare er interessert i å publisere *statistisk signifikante* funn.

## Hva er alternativet?

Det er ikke opplagt at enten-eller-svar basert på statistisk signifikans er nødvendig i de fleste enkeltstudier. Riktignok må vi ta enten-eller-beslutninger i klinisk praksis, men et statistisk signifikant resultat fra en enkeltstudie er uansett ikke tilstrekkelig for å ta kliniske beslutninger. For eksempel legges det mye vekt på klinisk signifikans i de fleste introduksjonskurs i statistikk – altså hvorvidt en effekt har betydning i klinisk praksis. Den kliniske signifikansen er umulig å vurdere fra en p-verdi alene. Dessuten gir p-verdien ingen informasjon om kvaliteten på studiedesignet og datamaterialet. Selvsagt kan heller ikke p-verdien belyse kostnader eller praktiske konsekvenser av en klinisk beslutning.

I stedet for å fokusere på statistisk signifikans bør en medisinsk studie belyse et forskningsspørsmål på så mange måter som mulig, og effektestimater med konfidensintervaller gir nesten alltid mer informasjon enn p-verdier.

## Problemet er ikke p-verdien i seg selv

Når det er sagt, vil vi understreke at p-verdien i seg selv ikke er skyld i problemene med begrepet statistisk signifikans. P-verdien er et veldefinert matematisk uttrykk som beskriver usikkerhet. Problemet er misbruk og mistolkninger av p-verdier. De svarer ikke direkte på spørsmålet om en hypotese er korrekt, og må derfor alltid tolkes i kontekst: En liten p-verdi kan gi god dokumentasjon for en velfundert hypotese, men en liten p-verdi gir også svak dokumentasjon for en lite plausibel hypotese (4).

Noen statistikere mener at bayesianske metoder bør erstatte p-verdiene og konfidensintervallene. Bayesianske beslutningsregler kan imidlertid rammes av det samme kobrafenomenet som statistisk signifikans. Dessuten krever bayesianske metoder at man definerer antagelser om effekter a priori, altså før studien gjennomføres, og disse antakelsene er ofte subjektive.

## Konklusjon

Det er klare problemer med bruken av begrepet statistisk signifikans i medisinsk forskning, og ukritisk bruk av p-verdier har antakelig bidratt til mange ikke-reproduserbare resultater. Derfor belyser årets opprør mot bruken av statistisk signifikans et viktig problem. Forhåpentligvis fører opprøret til at redaktører i mindre grad bruker statistisk signifikans som kriterium for å publisere artikler. Medisinske forskere bør også fokusere mer på effektstørrelser og konfidensintervaller enn p-verdier alene.

---

### LITTERATUR:

1. McShane BB, Gal D, Gelman A et al. Abandon statistical significance. *Am Stat* 2019; 73: 235–45. [CrossRef]
2. Amrhein V, Greenland S, McShane B. Scientists rise up against statistical significance. *Nature* 2019; 567: 305–7. [PubMed][CrossRef]
3. Greenland S, Senn SJ, Rothman KJ et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. *Eur J Epidemiol* 2016; 31: 337–50. [PubMed][CrossRef]
4. Nuzzo R. Statistical errors: P values, the 'gold standard' of statistical validity, are not as reliable as many scientists assume. *Nature* 2014; 513: 150–3. [PubMed][CrossRef]

Publisert: 2. august 2019. Tidsskr Nor Legeforen. DOI: 10.4045/tidsskr.19.0311  
© Tidsskrift for Den norske legeforening 2020. Lastet ned fra tidsskriftet.no