

Antall som må behandles (NNT) – misvisende, misforstått, misbrukt?

Sammendrag

Antall pasienter som må behandles (number needed to treat, NNT) er ett av flere effektmål som ikke fanger opp tidsdimensjonen. Det er fordi det baseres på måling på et gitt, tilfeldig tidspunkt. Denne egenskapen gjør NNT til et problematisk effektmål for intervensjoner mot kroniske sykdomsprosesser, der tid til hendelse er en sentral problemstilling. NNT har i tillegg ugunstige statistiske egenskaper. Begrepet kan gi inntrykk av at uønskede sykdomshendelser unngås helt med behandling, mens effekten i realiteten er at de blir utsatt i tid. NNT bør derfor anvendes og tolkes med varsomhet.

Engelsk sammendrag finnes i artikkelen på www.tidsskriftet.no

Oppgitte interessekonflikter: Ingen

Torbjørn Wisløff

torbjorn.wisloff@kunnskapssenteret.no
Nasjonalt kunnskapssenter for helsetjenesten
Postboks 7004 St. Olavs plass
0130 Oslo

Peder A. Halvorsen

Institutt for samfunnsmedisin
Universitetet i Tromsø

Ivar Sønbo Kristiansen

Nasjonalt kunnskapssenter for helsetjenesten

Pasienter, leger, administratorene og politikere har behov for å bedømme effektstørrelsen av medisinsk behandling når de skal treffe beslutninger – hva enten det er på klinisk eller mer aggregert nivå. I 1988 foreslo Laupacis number needed to treat (NNT) som effektmål fordi det skulle være lett å forstå for klinikere (1). Begrepet NNT er i de senere år blitt fremhevet som særlig velegnet til å måle intervensjonseffekter (2-11). Mange synes også å mene at det er et mer korrekt effektmål enn f.eks. relativ risikoreduksjon (2, 12).

En studie av sirkulasjonsstans ved ventrikelflimmer kan vise hvordan man beregner NNT og andre aktuelle effektmål (13). I en undergruppe i denne studien mislyktes

gjenoppretting av spontan sirkulasjon hos 42 % av pasientene når de fikk kardiopulmonal resuscitering før elektroshokk ble gitt («intervensjonsgruppen»), mens det var mislykket hos 62 % når man gav elektroshokk uten forutgående resuscitering. I det følgende defineres andel med uteblitt spontansirkulasjon som D_i i intervensjonsgruppen (42 %) og D_k i kontrollgruppen (62 %). Basert på disse to andeler kan man uttrykke behandlingseffekten på ulike måter:

(Likning 1)

$$(D_k - D_i) / D_k = \text{relativ risikoreduksjon (RRR)}$$

beskriver den relative reduksjon i risiko for det uønskede utfall (uteblitt spontansirkulasjon).

(Likning 2)

$$D_k - D_i = \text{absolutt risikoreduksjon (ARR)}$$

beskriver den absolutte reduksjon i risiko for det uønskede utfall. NNT er den resiproke (inverse) av ARR.

(Likning 3)

$$1 / (D_k - D_i) = \text{number needed to treat (NNT)}$$

I gjenoppplivningseksempelet er $ARR = 62\% - 42\% = 20\%$ eller 0,20. RRR er 32% ($20\%/62\%$), mens NNT er 5 ($1/0,2$). Man må i gjennomsnitt behandle fem pasienter (med resuscitering før elektroshokk) for å redde én pasient, hvilket innebærer at i gjennomsnitt én av fem pasienter har nytte av behandlingen.

Ut fra andelene D_k og D_i kan man også uttrykke behandlingseffekten som oddsratio (OR) eller log-OR, hvilket er beskrevet i lærebøker om biostatistikk og epidemiologi. Det er verdt å legge merke til at alle fem effektmål er ulike matematiske formuleringer av de samme to andeler. De gir uttrykk for ulike aspekter av behandlingseffekten, men det ene er ikke mer korrekt enn noen av de andre. Flere studier har imidlertid vist at både pasienter, leger og helseadministratorene kan være mer velvillig innstilt til en behandling når effekten uttrykkes som relativ enn som absolutt risikoreduksjon (6, 14-20). Førstnevnte oppfattes derfor gjerne som misvisende fordi det «overdriver» behandlingseffekten (2, 16). Siden det ikke finnes noen gullstandard for hva som er korrekte beslutninger, kan man like gjerne si at det ene effektmålet undervurderer effekten som at det andre overvurderer den.

Behandling som

utsetter ønsket hendelse

I eksemplet om resuscitering var intervensjonen kortvarig og effekten kunne observeres nærmest umiddelbart. Ved langvarige intervensjoner rettet mot kroniske sykdomsprosesser som aterosklerose, diabetes eller osteoporose kan man ikke alltid unngå uønskede konsekvenser (f.eks. hjerteinfarkt eller nyresvikt), men de kan utsettes. Helsetjenesten kan som kjent ikke forebygge døden – bare utsette den i tid. Beregningen av NNT skjer på samme måte som vist i likningene ovenfor, der andelene (D_k og D_i) er målt på et gitt tidspunkt. *Tolkningen* er imidlertid mer komplisert ved kroniske tilstander, fordi det spiller en sentral rolle *når* i forløpet man velger å observere effekten. Figur 1 illustrerer dette ved hjelp av en hypotetisk intervensjonsstudie. For enkelthets skyld begrenses antallet pasienter her til 40. Av disse får 20 pasienter behandling (intervensjon), de andre 20 utgjør en kontrollgruppe. Hver pasient er illustrert med en horisontal linje, og linjens lengde illustrerer hvor lenge hver enkelt pasient lever fra behandlingsstart. I figur 1a er pasientene tilfeldig fordelt, mens de i figur 1b er ordnet etter levetid. Fra figur 1 ser man at behandlingen fremskynder døden for dem som dør tidlig i studieforløpet (dette kan f.eks. skyldes bivirkninger), mens den senere i forløpet utsetter døden. I gjennomsnitt lever de behandlede pasientene noe lenger enn de ubehandlede.

Figur 2 er en revidert utgave av figur 1b og gir et visuelt inntrykk av levetid og behandlingseffekt for den enkelte pasient. Y-aksen viser nå andelen som er i live, og denne andelen må nødvendigvis falle over tid. Ved hjelp av slike *overlevelseskurver* kan vi uttrykke effekt av intervensjoner mot kroniske tilstander på tre ulike måter: vertikale

Hovedbudskap

- Antall pasienter som må behandles (number needed to treat; NNT) fanger ikke opp tidsdimensjonen i kliniske studier
- NNT kan gi inntrykk av at uønskede sykdomshendelser unngås helt med behandling, mens de i realiteten blir utsatt i tid
- Begrepet bør brukes og tolkes med varsomhet

effekt mål, horisontale effekt mål og kombinerte effekt mål.

Vertikale effekt mål

Vertikale effekt mål uttrykker forskjellen mellom de to gruppene på ett bestemt tidspunkt. NNT, RRR og ARR er eksempler på vertikale effekt mål. Disse er mye brukt i vitenskapelige arbeider der man typisk presenterer forskjeller mellom intervensjons- og kontrollgruppe ved avslutning av studien, uten å ta hensyn til tidsaspekter. Vertikale effekt mål har også en fremtredende plass i retningslinjer for forebygging av hjerte- og karsykdom (21), og effekt av behandlingsoptimaliteter sammenliknes gjerne ved hjelp av NNT (22).

Problemen oppstår i fortolkningen av NNT. Eksempelvis fortolkes $NNT = 17$ som at «én av 17 som behandles i ti år, unngår en sykdomshendelse pga. behandlingen» (21) eller at ti ikke har effekt av behandlingen når NNT er 11 (1, 23). Så enkelt er det imidlertid ikke. I figur 1b angir den stiplede linjen et tidspunkt der det er ett dødsfall mindre i intervensjonsgruppen enn i kontrollgruppen. ARR er således $1/20$, NNT er 20. Imidlertid ser også de øvrige 19 ut til å ha blitt påvirket av intervensjonen. Seks pasienter dør tidligere enn uten intervensjon (negativ effekt) og 14 dør senere (positiv effekt). Det virker rimelig å anta at intervensjoner mot kroniske, langsomt progredierende sykdomsprosesser har en viss effekt på tidspunktet for den uønskede hendelse (død, hjerteinfarkt etc.) hos en stor andel av de behandlede. Dersom utsettelsen av hendelsen er liten, kan den absolutte risikoreduksjon bli liten og NNT stor, uten at andelen av pasientene som har nytte av intervensjonen nødvendigvis er liten. Andre vertikale effekt mål kan muligens gi inntrykk av at alle som behandles, har nytte av intervensjonen. Dette vil imidlertid oftest ikke være korrekt. Heller ikke ARR eller RRR kan si noe om sannsynligheten for effekt hos den enkelte. Alle vertikale effekt mål som måles på ett tidspunkt, angir bare gjennomsnittlig effekt ved dette tidspunktet.

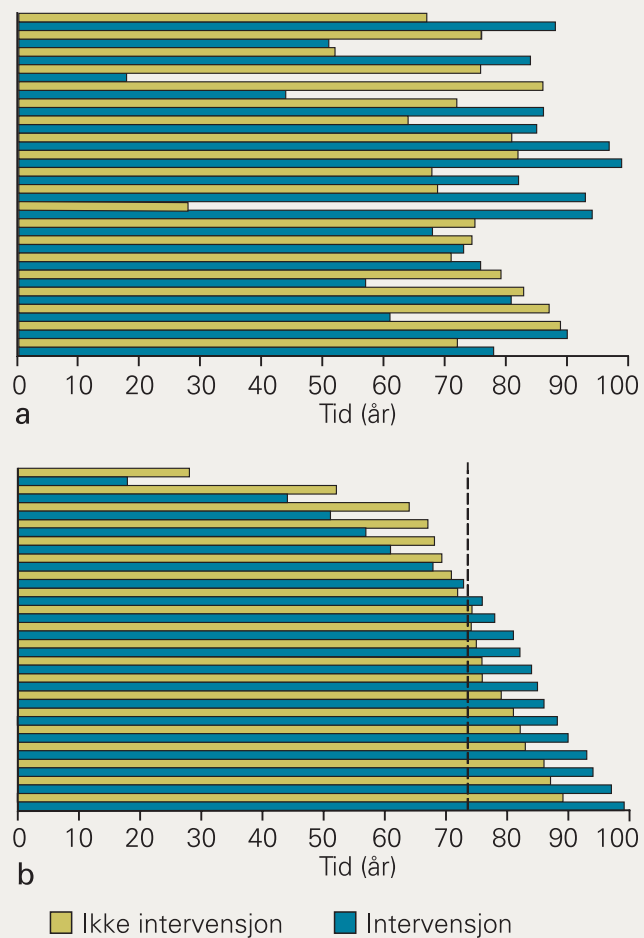
Horisontale effekt mål

Horisontale effekt mål tar utgangspunkt i tid til uønsket hendelse (f.eks. død) hos en viss andel av pasientene. Dersom levetiden var 83 år for den foreløpig siste døde når halvparten er døde i intervensjonsgruppen, er median overlevelse 83 år i denne gruppen. Dersom median overlevelse var for eksempel 75 år i kontrollgruppen, har behandlingen altså økt median overlevelse med åtte år.

Kombinerte effekt mål

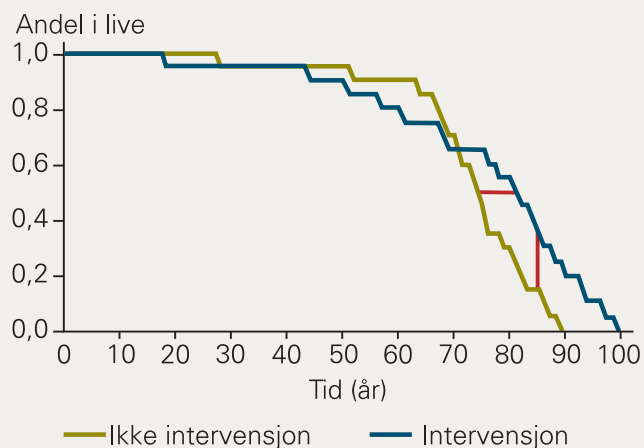
Kombinerte effekt mål gir uttrykk for effekter aggregert over tid. Mens de vertikale effekt mål kan variere mye over tid, kan de horisontale variere med hvilken andel man velger å måle etter. De kombinerte fanger opp begge, både tidsdimensjonen og andelen

Figur 1



a) Levetid for 20 pasienter som fikk behandling (intervensjon) for en dødelig tilstand og 20 identiske kontrollpasienter. b) Samme pasienter som figur 1a, men ordnet etter levetid. Den stiplede linjen angir et tidspunkt der det er ett dødsfall mindre i intervensjonsgruppen enn i kontrollgruppen

Figur 2



Samme data som i figur 1b fremstilt som en overlevelsesanalyse med levetid på x-aksen og andel overlevende på y-aksen. D_i minus D_k er tegnet inn (vertikal rød linje) for å gi et visuelt bilde av hvordan absolutt risikoreduksjon (ARR), relativ risikoreduksjon (RRR) og antall som må behandles (number needed to treat, NNT) fremkommer. Den horisontale røde linjen er lagt inn ved 50 % for å gi et bilde av hvordan man kommer frem til forskjell i median levetid

Ramme 1

Ulike fortolkninger av begrepet antall som må behandles (number needed to treat, NNT)

- En vanlig tolking av NNT: Man skal behandle 17 pasienter i ti år for å unngå ett dødsfall
- En mer korrekt tolking av NNT: Man skal i gjennomsnitt behandle 17 pasienter i ti år for å observere ett dødsfall mindre på tiårsdagen fra behandlingsstart
- Et eksempel på en gjengivelse av NNT hvor utgangsrisiko er tatt med: Man skal i gjennomsnitt behandle 17 pasienter i ti år for å observere ett dødsfall mindre på tiårsdagen fra behandlingsstart for pasienter som ubehandlet har en utgangsrisiko svarende til at sju av 17 pasienter er døde etter ti år dersom de ikke får behandling

døde på hvert tidspunkt. Det vanligste kombinerte effektmål er endring i forventet levetid. Denne kan beregnes på grunnlag av overlevedesdataene eller leses direkte som arealet mellom de to overlevelseskurvene. I figur 2 har dette arealet to deler – den første representerer et initialt leveårstap, den største en leveårsgevinst. Netto leveårsgevinst tilsvarende i dette tilfellet det største arealet minus det minste.

Dersom den uønskede hendelsen ikke er død, men f.eks. hjerneslag, er prinsippet at man avslutter registreringen for den enkelte pasient når pasienten enten opplever hendel-

sen eller av andre årsaker ikke kan være med i studien lenger (f.eks. død). Arealet mellom kurvene beregnes på samme måte som ved fatale hendelser, men det får nå en annen tolking: gjennomsnittlig utsettelse av hendelsen. I en dansk undersøkelse fant man at osteoporoseintervensjoner som reduserer bruddrisikoen til det halve i høyrisikogrupper, gir en gjennomsnittlig utsettelse av hoftebrudd på fra en uke til 12 måneder, avhengig av intervensjonsvarighet og alder ved intervensjonsstart (24).

Hvordan skal NNT tolkes?

«Man skal behandle 17 pasienter i ti år for å unngå ett dødsfall» lyder ikke spesielt komplisert, men tolkingen av NNT er problematisk av flere årsaker.

Tilfeldigheter vil gjøre at man neppe unngår én hendelse hver gang man har behandlet 17 pasienter i ti år. I praksis vil antall unngåtte hendelser kunne være null så vel som to, tre eller mer av ren tilfeldighet. Av tilsvarende grunner får man langt fra alltid en sekser på hvert sjette terningskast selv om sannsynligheten for en sekser er eksakt 1/6.

Dødsfall kan ikke unngås i det lange løp, bare utsettes i tid. Det samme gjelder i varierende grad andre uønskede hendelser som f.eks. hjerteinfarkt eller hoftebrudd. Kun når intervensjonen utsetter infarkt eller brudd så lenge at pasienten rekker å dø av andre årsaker, kan man med rette si at hendelsene er unngått.

NNT betyr *ikke* at 16 pasienter ikke har vært påvirket av behandlingen når NNT er 17. Ramme 1 angir både en populær, en mer korrekt og en mer informativ tolking av NNT.

NNT varierer i forløpet av en intervensjon. Straks etter behandlingsstart er det ingen forskjell i andelen døde mellom gruppene, og NNT er uendelig stor. Det samme gjelder om man venter med å måle NNT til alle pasienter i begge grupper er døde. I perioden mellom disse ytterpunktene vil NNT variere.

NNT varierer med den absolutte risiko for den uønskede hendelse i kontrollgruppen. Dersom for eksempel den relative risikoreduksjon er 50 % og den absolutte risiko for død etter tre år er 2 %, er ARR 1 % og NNT 100 (1/0,01). I en høyrisikogruppe der 10 % er døde etter tre år, blir ARR 5 % og NNT 20 (1/0,05) med samme RRR. Generelt angir absolutt risiko en øvre grense for hvor stor absolutt risikoreduksjon som er mulig og dermed hvor liten NNT kan bli.

Når man f.eks. velger å intervenere ved ti års absolutt risiko for koronar hendelser på 20 %, kan absolutt risikoreduksjon ikke bli større enn 20 % og NNT ikke lavere enn 5. Dette innebærer at inntil én av fem, men ikke flere, kan ha fått del i den målbare nytten av intervensjonen ti år etter behandlingsstart, selv om NNT på tiårsdagen er høyere enn 5. Derimot er det ikke utelukket at flere enn én av fem vil få utsatt koronar hendelse etter

tiårsdagen, selv i de tilfeller der intervensjonen avsluttes etter ti år (25).

Hvis ingen hadde nytte etter tiårsdagen, ville overlevelseskurvene umiddelbart bli sammenfallende på dette tidspunkt. Ved intervensjoner mot eksempelvis hypertensjon og hyperkolesterolemi er sistnevnte scenario neppe plausibelt.

Mange medisinske intervensjoner påvirker flere enn én uønsket hendelse. Om fettinntaket i kostholdet reduseres, kan risikoen for så vel hjerteinfarkt som hjerneslag reduseres. Behandler man osteoporose, kan risikoen for både hoftebrudd og underarmsbrudd reduseres. Spørsmålet blir da for hvilken hendelse man skal beregne og presentere NNT. Disse begrensningene gjelder imidlertid ikke bare NNT, men alle tverrsnittsmål (ARR, RRR, økt median levetid).

Blir NNT korrekt forstått?

Til tross for mange anbefalinger av NNT, finner vi bare indirekte og svake indikasjoner på at begrepet blir korrekt forstått (26). Derimot finnes det flere indikasjoner på at det blir misforstått. Som nevnt hevder enkelte forfattere at ti av 11 pasienter ikke har nytte av en behandling eller ikke trenger den når NNT er 11 (1), hvilket ikke er tilfellet for kroniske sykdomstilstander (fig 1a/1b). Tilsvarende fortolkninger er påvist blant legfolk (27, 28) og leger (20, 29).

I andre studier ble legfolk spurt om de ville ønske en hypotetisk behandling for hyperkolesterolemi eller osteoporose (27, 28). Respondentene ble randomisert til et vidt spekter av effektstørrelse (NNT fra 10 til 400). NNT på 10 må anses som en betydelig effekt, mens 400 tilsvarer en ARR på 0,0025. Det trengs kliniske studier med ca. 10 000 pasienter for å oppdage så små effekter. Imidlertid var andelen legfolk som sa ja til behandling praktisk talt upåvirket av effektstørrelse målt som NNT. I en studie vedrørende medikamentell intervensjon mot osteoporotiske brudd (28) ble effekten også presentert som antall vunne leveår uten hoftebrudd. Da ble det påvist en sammenheng mellom effektstørrelsen og andelen som stilte seg positiv til intervensjonen. Oppslutningen om intervensjonen var størst når NNT ble brukt, enten NNT var 10, 50, 100 eller 400. Sheridan og medarbeidere (30) viste at kun 30 % av pasientene klarte å identifisere den beste av to behandlinger når effektene ble uttrykt som NNT, mot 42 % for ARR og 60 % for RRR. Enda færre klarte å regne ut risikoreduksjonen på bakgrunn av oppgitte data (NNT = 6 %, ARR = 17 % og RRR = 21 %). Disse studiene understreker for det første at NNT ikke nødvendigvis er så lett å forstå, dessuten at man er sensitiv til hvilket effektmål som brukes.

Konfidensintervaller for NNT

I praksis vet man ikke den sanne verdi av absolutt risikoreduksjon og dermed heller ikke den sanne NNT. Usikkerheten uttrykkes med

Ramme 2

Kontrollspørsmål når intervensjonseffekter er angitt som antall som må behandles (number needed to treat, NNT)

- Påvirker intervensjonen én eller flere hendelser (endepunkter)?
- Hvilken hendelse er NNT beregnet for?
- På hvilket tidspunkt er NNT beregnet?
- Angis det hva NNT er på andre tidspunkter?
- Er konfidensintervallet for NNT korrekt beregnet når konfidensintervallet for absolutt risikoreduksjon inneholder null?
- Er NNT tolket som uttrykk for andel som har nytte av intervensjonen?
- Er det angitt NNT for ulike risikogrupper?
- Er NNT benyttet til å beregne kostnad per unngått hendelse?

konfidensintervaller. Dette gjøres oftest ved å ta den inverse (resiproke) av konfidensgrensene for ARR – hvis ARR er 0,05 (95% KI 0,02–0,08), blir NNT 20 (95% KI 12,5–50). Denne enkle metoden gir noe skjevhet i estimatene for konfidensgrensene for NNT, men det har trolig liten praktisk betydning (31). Viktigere er det at metoden kan gi feilaktige konfidensintervaller om man ikke er vaksom (8): Konfidensintervaller for NNT er ikke definert mellom –1 og 1, siden $NNT = 1/ARR$ og ARR kun kan anta verdier i intervallet [–1, 1]. NNT gir for øvrig ingen mening i området mellom –1 og +1 fordi man må behandle minst én pasient for å kunne oppnå effekt. Således vil NNT ha to konfidensintervaller når et konfidensintervall for ARR inneholder 0. Om for eksempel ARR har konfidensintervallet [–0,01, +0,02], vil konfidensintervallene for NNT bli [–∞, –100] og [50, ∞].

NNT og prioritering

I økonomisk evaluering av helsetiltak er det vanlig å summere opp resultatene i et kostnad-effekt-forhold som for eksempel kostnad per vunnet leveår (32). Det finnes teoretiske begrunnelser for hvordan slike forholdstall kan gi veiledning i prioriteringsdiskusjoner, og det finnes veiledende normer for hvor mye samfunnet med rimelighet bør akseptere av kostnad per leveår (33). Dersom man benytter NNT og kostnad per unngått hendelse (død, hoftebrudd, hjerneslag etc.), blir analysen tilsynelatende mye enklere å utføre, men faren for feiltolkning er betydelig (25). Om medikamentkostnadene for ett års kardioprotektiv behandling er 5 000 kroner og absolutt risikoreduksjon for hjerteinfarkt er 1% (0,01) etter tre års behandling, kan det være fristende å gjøre følgende enkle regnestykke: «Når NNT er 100 (1/0,01) og det koster 15 000 kroner å behandle i tre år, blir kostnaden per unngått hjerteinfarkt $100 \cdot 15\ 000$ kroner = 1 500 000 kroner.» Dette regnestykket kan av flere grunner være meget misvisende. Det tas ikke hensyn til at NNT kan variere i forløpet og at behandlingen kanskje utsetter snarere enn forhindrer hjerteinfarkt (fig 1b). Ei heller tar man i betraktning kostnader og effekter utover treårsperioden og kostnader utover medikamentkostnader eller sparte behandlingskostnader når hjerteinfarkt faktisk forebygges/utsettes. Videre gjøres ikke diskontering av fremtidige kostnader og gevinster, og man fanger ikke opp effekt på andre typer hjerte- og karsykdom enn hjerteinfarkt (ramme 2).

Hvilket effektmål er best?

Effektmål for intervensjoner mot kroniske sykdoms- eller risikotilstander bør ideelt fange opp andelen med uønsket hendelse på hvert tidspunkt slik at man får et aggregert mål for effekten. Økning i forventet levetid før uønsket hendelse er eksempel på et aggregert effektmål som fanger opp tidsdimensjonen i effekten. Det samme gjør effektmål basert på Cox-regresjon, hvis alle forutsetningene er til stede.

NNT og andre tverrsnittsmål er basert på to andeler målt på ett tilfeldig tidspunkt og fanger ikke opp tidsdimensjonen, hvilket er en vesentlig begrensning. Når vi legger hovedvekt på NNT, er det fordi det uten nærmere dokumentasjon fremstilles som bedre enn andre effektmål. Særlige problemer ved NNT er at begrepet kan forlede oss til å tro at sykdomshendelser blir helt unngått selv om de i realiteten bare utsettes. Videre har NNT fått et navn som inviterer til å tro at andelen som har nytte av intervensjoner er $1/NNT$. At dette ikke er korrekt ved kroniske sykdomsprosesser, kan være vanskelig å forstå, endog når det blir forklart med enkle numeriske eksempler (34). Mange mennesker har problemer med å tolke tall og statistikk korrekt (35, 36). Man må derfor regne med at ethvert effektmål, det være seg NNT, ARR eller RRR, kan misforstås og bli misbrukt.

Etter vår oppfatning er det verken teoretiske eller empiriske argumenter som taler for at NNT bør foretrekkes i den kliniske hverdag. Det som i dag finnes av teoretiske og empiriske argumenter, trekker heller i motsatt retning. NNT egner seg antakelig best til å beskrive effekter av intervensjoner mot akutte tilstander, der «utsettelsesdimensjonen» er mindre relevant.

Litteratur

1. Laupacis A, Sackett DL, Roberts RS. An assessment of clinically useful measures of the consequences of treatment. *N Engl J Med* 1988; 318: 1728–33.
2. McCormack JP, Levine M. Meaningful interpretation of risk reduction from clinical drug trials. *Ann Pharmacother* 1993; 27: 1272–7.
3. Cook RJ, Sackett DL. The number needed to treat: a clinically useful measure of treatment effect. *BMJ* 1995; 310: 452–4.
4. Jaeschke R, Guyatt G, Shannon H et al. Basic statistics for clinicians: 3. Assessing the effects of treatment: measures of association. *CMAJ* 1995; 152: 351–7.
5. Schulzer M, Mancini GB. «Unqualified success» and «unmitigated failure»: number-needed-to-treat-related concepts for assessing treatment efficacy in the presence of treatment-induced adverse events. *Int J Epidemiol* 1996; 25: 704–12.
6. Fahey T, Griffiths S, Peters TJ. Evidence based purchasing: understanding results of clinical trials and systematic reviews. *BMJ* 1995; 311: 1056–9.
7. McQuay HJ, Moore RA. Using numerical results from systematic reviews in clinical practice. *Ann Intern Med* 1997; 126: 712–20.
8. Altman DG. Confidence intervals for the number needed to treat. *BMJ* 1998; 317: 1309–12.
9. Altman DG, Andersen PK. Calculating the number needed to treat for trials where the outcome is time to an event. *BMJ* 1999; 319: 1492–5.
10. Massel D, Cruickshank MK. The number remaining at risk: an adjunct to the number needed to treat. *Can J Cardiol* 2002; 18: 254–8.
11. Sackett DL, Richardson WS, Rosenberg W et al. Evidence-based medicine. How to practice and teach EBM. New York: Churchill Livingstone, 2000.
12. Skolbekken JA. Om leger, journalister og plikten til å kjenne sin dødsrisiko. *Tidsskr Nor Lægeforen* 2003; 123: 1548–9.
13. Wik L, Hansen TB, Fylling F et al. Delaying defibrillation to give basic cardiopulmonary resuscitation to patients with out-of-hospital ventricular fibrillation: a randomized trial. *JAMA* 2003; 289: 1389–95.
14. Forrow L, Taylor WC, Arnold RM. Absolutely relative: how research results are summarized can affect treatment decisions. *Am J Med* 1992; 92: 121–4.

15. Bobbio M, Demichelis B, Giustetto G. Completeness of reporting trial results: effect on physicians' willingness to prescribe. *Lancet* 1994; 343: 1209–11.
16. Bucher HC, Weinbacher M, Gyr K. Influence of method of reporting study results on decision of physicians to prescribe drugs to lower cholesterol concentration. *BMJ* 1994; 309: 761–4.
17. Hux JE, Naylor CD. Communicating the benefits of chronic preventive therapy: does the format of efficacy data determine patients' acceptance of treatment? *Med Decis Making* 1995; 15: 152–7.
18. Cranney M, Walley T. Same information, different decisions: the influence of evidence on the management of hypertension in the elderly. *Br J Gen Pract* 1996; 46: 661–3.
19. Misselbrook D, Armstrong D. Patients' responses to risk information about the benefits of treating hypertension. *Br J Gen Pract* 2001; 51: 276–9.
20. Nexoe J, Gyrd-Hansen D, Kragstrup J et al. Danish GPs' perception of disease risk and benefit of prevention. *Fam Pract* 2002; 19: 3–6.
21. Fretheim A, Bjørndal A, Oxman AD et al. Retningslinjer for medikamentell primærforebygging av hjerte- og karsykdommer – hvem bør behandles? *Tidsskr Nor Lægeforen* 2002; 122: 2277–81.
22. Meland E, Ellekjaer H, Gjelsvik B et al. Medikamentell forebygging av hjerte- og karsykdommer i allmennpraksis. *Tidsskr Nor Lægeforen* 2000; 120: 2643–7.
23. San Laureano JA, Briganti EM, Colville DJ. Number needed to treat: a useful new method of assessing the magnitude of treatment effect and its application to the management of diabetic retinopathy. *Aust N Z J Ophthalmol* 1999; 27: 137–42.
24. Christensen PM, Brosen K, Brixen K et al. Expressing effects of osteoporosis interventions in terms of postponing of fractures. *Eur J Clin Pharmacol* 2002; 58: 629–33.
25. Kristiansen IS, Gyrd-Hansen D. Cost-effectiveness analysis based on the number-needed-to-treat: common sense or non-sense? *Health Economics* 2004; 13: 9–19.
26. Laupacis A, Sackett DL, Roberts RS. Therapeutic priorities of Canadian internists. *CMAJ* 1990; 142: 329–33.
27. Kristiansen IS, Gyrd-Hansen D, Nexoe J et al. Number needed to treat: easily understood and intuitively meaningful? Theoretical considerations and a randomized trial. *J Clin Epidemiol* 2002; 55: 888–92.
28. Christensen PM, Brixen K, Brosen K et al. A Randomized trial of laypersons' perception of the benefit of osteoporosis therapy: number needed to treat versus postponement of hip fracture. *Clin Ther* 2003; 25: 2575–85.
29. Halvorsen PA, Kristiansen IS, Aasland OG et al. Medical doctors' perception of the «number needed to treat» (NNT). A survey of doctors' recommendations for two therapies with different NNT. *Scand J Prim Health Care* 2003; 21: 162–6.
30. Sheridan SL, Pignone MP, Lewis CL. A Randomized comparison of patients' understanding of number needed to treat and other common risk reduction formats. *J Gen Intern Med* 2003; 18: 884–92.
31. Hutton JL. Number needed to treat: properties and problems. *J R Statist Soc* 2000; 163: 403–19.
32. Kristiansen IS, Stavem K, Linnestad K et al. Evaluering av medisinske metoder – kan vi stole på kostnad-effekt-analyser? *Tidsskr Nor Lægeforen* 2003; 123: 657–60.
33. Kristiansen IS. Hvor mye bør samfunnet være villig til å betale for helseforbedringer? Verdien av liv og helse. Oslo: Universitetet i Oslo, 2003: 55–68.
34. Kristiansen IS, Nexoe J, Gyrd-Hansen D et al. NNT is not easily understood. *Fam Pract* 2002; 19: 566–7.
35. Sheridan SL, Pignone M. Numeracy and the medical student's ability to interpret data. *Eff Clin Pract* 2002; 5: 35–40.
36. Schwartz LM, Woloshin S, Black WC et al. The role of numeracy in understanding the benefit of screening mammography. *Ann Intern Med* 1997; 127: 966–72.