

Kritisk søkelys på norsk screeningstudie

Våren 2007 publiserte forskere fra Universitetssykehuset Nord-Norge, Ullevål universitetssykehus og Rikshospitalet resultater fra en stor norsk screening- og intervensjonsstudie. I kjølvannet av dette har det kommet kritiske innspill til valg av studiedesign og validitet av resultatene. Dreier debatten seg egentlig om forskjellig kunnskapssyn?

Heidi Tiller

heidi.tiller@unn.no

Avdeling for immunologi og transfusjonsmedisin
Universitetssykehuset Nord-Norge
9038 Tromsø

Avdeling for immunologi og transfusjonsmedisin ved Universitetssykehuset Nord-Norge er nasjonalt referanselaboratorium for blodplateimmunologi. Spesielt interesserer vi oss for blodplateantistoffer hos gravide. Disse antistoffene kan passere gjennom morkaken over til barnet og ødelegge blodplatene til barnet i magen. Tilstanden kalles neonatal alloimmun trombocytopeni, og oppstår hos 1 av 1 000–2 000 fødte barn. I alvorlige tilfeller kan barnet få hjerneblødning, enten i mors liv eller i nyfødtpærioden. Komplikasjoner til hjerneblødning kan for eksempel være cerebral parese eller død. Tilstanden kan oppstå dersom mor og barn har ulike blodplattetyper. I den kaukasiske populasjonen er det først og fremst de 2 % kvinner med blodplattetyper HPA1bb som er utsatt for å få barn med neonatal alloimmun trombocytopeni (1).

Hvorfor ikke randomisering?

I perioden 1995–2004 ble 100 000 gravide kvinner fra helseregion 1, 2 og 5 inkludert i en screening- og intervensjonsstudie. Screening gikk ut på å analysere blodplattetyper i blodprøver tatt ved første svangerskapskontroll. De 2 100 gravide som hadde blodplattetype HPA1bb ble fulgt videre opp ved å undersøke om de utviklet antistoff mot barnets blodplater. Intervensjonen besto i at barnet ble tatt med keisersnitt 2–3 uker før termin dersom mor utviklet antistoff over et visst nivå. I tillegg hadde man klargjort forlikelige blodplater til barnet som ved behov ble transfundert straks etter forløsning. Resultater fra studien viste at neonatal morbiditet og mortalitet ble redusert med nesten 75 % som følge av screening og intervensjon (1, 2).

I screeningstudien valgte man å gjøre en

enarmet ikke-randomisert kohortstudie med en kontrollgruppe fra et historisk datamateriale. Kontrollgruppen besto av flere prospektive studier, noen var intervensjonsstudier, mens andre beskrev såkalt naturlige forløp av neonatal alloimmun trombocytopeni.

I diskusjonsdelen av artikkelen der screening- og intervensjonsstudien presenteres, ble det i hovedsak brukt to argumenter for hvorfor randomisert kontrollert studie ikke ble valgt som studiedesign. Forfatterne anså det som urealistisk å skulle gjennomføre en randomisert kontrollert studie, siden det ville innebære at 1 000 allmennpraktikere måtte foreta randomisering og gi informasjon om studien til 2 000 gravide. Kvaliteten på informasjonen gitt til den gravide ville neppe bli tilfredsstillende. Videre mente man at det ville være etisk betenkelig å randomisere kvinner med høyt nivå av antistoff til å skulle føde vaginalt, ettersom vaginal fødsel sannsynligvis gir økt risiko for hjerneblødning hos et barn med neonatal alloimmun trombocytopeni sammenliknet med keisersnitt.

I kostnad-nytte-studien brukte man et beslutningstre for å analysere tre ulike screeningstrategier. Hovedendepunkter var kvalitetsjusterte leveår og kostnader. Konklusjonen av analysen var at antenatal screening for neonatal alloimmun trombocytopeni i Norge er mulig å gjennomføre kostnadseffektivt (2). I denne studien brukte man en litteraturstudie der også retrospektive studier var inkludert som kontrollpopulasjon (3). I artikkelen med kostnad-nytte-analysen diskuteres det at manglende randomisering er et metodisk problem som gir studien begrensninger. Forfatterne innser at bruk av retrospektive studier i kontrollgruppen gir fare for overestimering av antall tilfeller av alloimmun trombocytopeni, og at man dermed risikerer seleksjonsskjevhet i retning av falskt positivt resultat av screening.

Kritiske røster skaper debatt

En tverrfaglig gruppe møtte høsten 2007 Sosial- og helsedirektoratet og presenterte

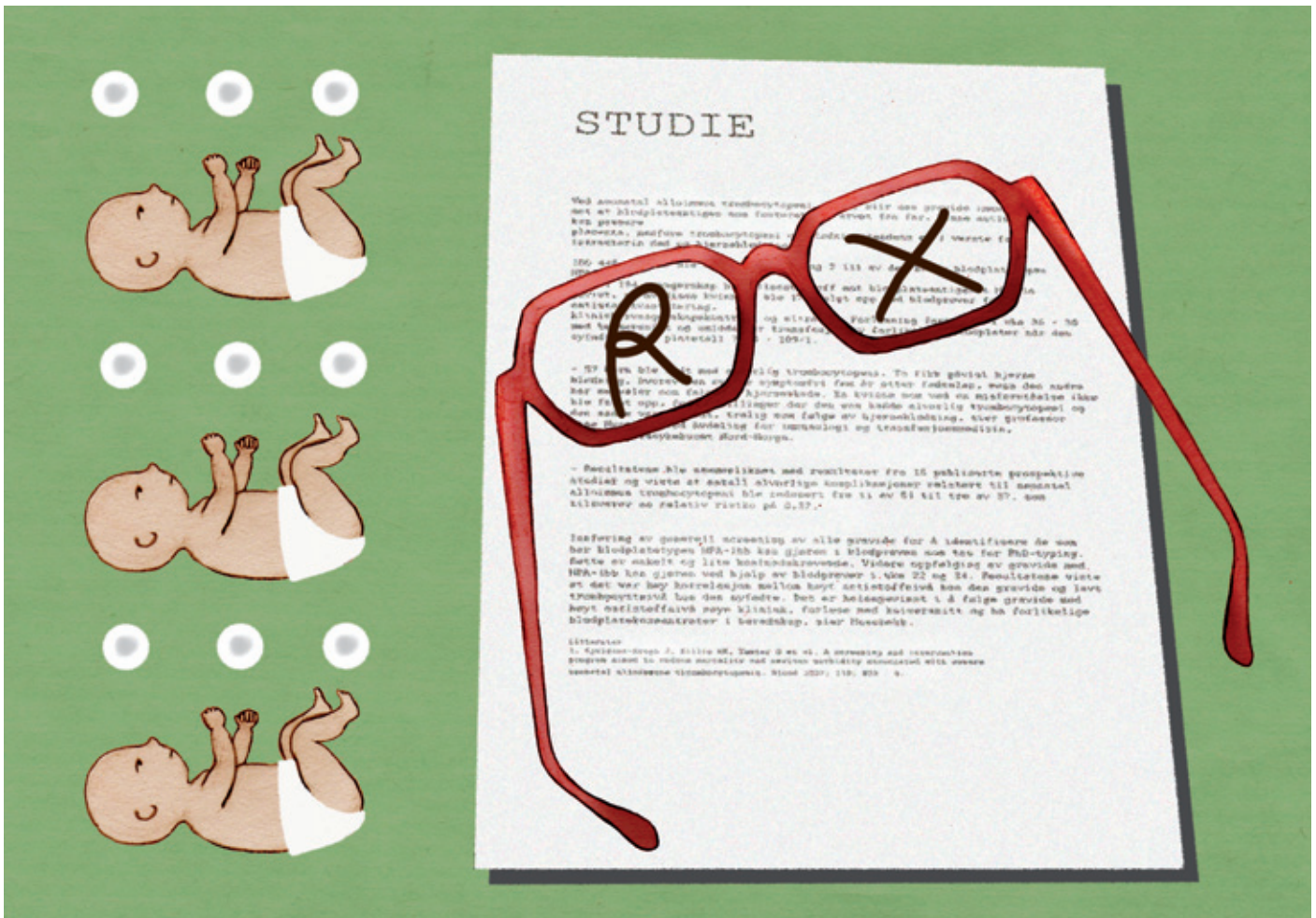
resultatene med tanke på implementering av ny nasjonal screening. I kjølvannet av dette møtet har det kommet kritiske innspill til bl.a. valg av studiedesign og validitet av resultatene. I et innlegg i *BJOG – An International Journal of Obstetrics and Gynaecology* ble det hevdet at kostnad-nytte-analysen av screening for neonatal alloimmun trombocytopeni var basert på ugyldige forutsetninger (4). Hovedargumentet var at bruk av historiske kontrollgrupper ikke er tilfredsstillende som beslutningsgrunnlag. Forfatteren kritiserer også bruk av ulike kontrollgrupper i screeningstudien og kostnad-nytte-analysen og mener det i seg selv gjør at kostnad-nytte-analysen ikke er gyldig. Han hevder videre at sensitivitetsanalysen ikke har tatt høyde for usikkerheten omkring kontrollgruppen.

Forfatterne av kostnad-nytte-studien hevder i et motinnlegg at det er tatt hensyn til usikkerheten omkring kontrollgruppen ved at man har brukt 95 % konfidensintervall både for screeningpopulasjonen og for den historiske kontrollgruppen i sensitivitetsanalysen. Resultatene fra studien viser at screening er kostnadseffektivt selv om man setter opp «høyeste og laveste tall» fra begge gruppene mot hverandre. Forfatterne understreker dog at de er fullt klar over usikkerhetene i data når det gjelder effekt av screening og intervensjon. Grunnen til at man brukte en noe annen kontrollgruppe i kostnad-nytte-analysen, var at de fleste prospektive studiene manglet opplysning om hvordan det hadde gått med barna som fikk hjerneblødning (outcome). Forfatterne påpeker at det er usikkerhet også ved randomiserte, kontrollerte studier, og at større kliniske og politiske beslutninger ofte tas på et svakere kunnskapsgrunnlag enn dette (5).

«Vi gjorde det som var best mulig, både ut fra hva som var praktisk gjennomførbart og etisk forsvarlig.» har en av forfatterne hevdet til undertegnede i en personlig henvendelse. Hva er det best mulige? Og er det godt nok i denne sammenhengen? Eller er resultatene ugyldig kunnskap som ikke kan danne grunnlag for å vurdere innføring av screening? Jeg ser flere problemer knyttet til kontrollpopulasjonen. Imidlertid vil jeg hevde at denne debatten hovedsakelig angår ulikt kunnskapssyn.

Equipoise – hva slags og for hvem?

Hvis vi likevel tenker oss at screeningstudien skulle vært gjennomført som en randomisert studie, blir det sentralt å vurdere det



Illustrasjon Katrine Kallekleiv

såkalte equipoise-begrepet. Brukt i forskningssammenheng brukes begrepet oftest om at det foreligger usikkerhet vedrørende hvilket behandlingsalternativ som er det beste (6). Det finnes ikke et etablert norsk ord for equipoise i fagmiljøet.

Faglig hersker det i hovedsak enighet om at equipoise eller usikkerhet er en forutsetning for at randomisering skal være etisk forsvarlig. Dersom behandling A og B i utgangspunktet oppfattes som like gode, kan man si at det foreligger equipoise som berettiger randomisering (7). Dersom det motsatte er tilfelle, at behandling A anses som klart bedre enn B, foreligger ikke slik usikkerhet – og randomisering anses som uetisk.

Men equipoise er mer komplisert enn som så. Før Benjamin Freedman introduserte begrepet klinisk equipoise i 1987, var det usikkerhetsprinsippet hos den enkelte lege/forsker som ble regnet som etisk gullstandard. Ut fra eget skjønn og kunnskap om den enkelte pasient kunne legen bestemme om det forelå usikkerhet eller ikke omkring behandlingsalternativene, og dermed hvorvidt randomisering av pasienten skulle forkastes eller aksepteres. Freedman mente dette var utilfredsstillende og lanserte derfor et mer kollektivt equipoisebegrep. Dersom det foreligger equi-

poise i et medisinsk fagmiljø – selv om enkeltindivider i miljøet ikke nødvendigvis er i equipoise – er kravet om klinisk equipoise oppfylt. Freedmans begrep har vært den ledende og bredt ut aksepterte doktrinen de senere år.

Klinisk equipoise er ikke nødvendigvis det samme som vitenskapelig equipoise (8). Vitenskapelige hypoteser kan være uavklart og må testes. Imidlertid kan det kliniske fagmiljøet ha en klar oppfatning om hva som er for eksempel best behandling. Best behandling for pasienten er ikke alltid synonymt med det som gir størst klinisk effekt. Ubegrep og potensiell risiko ved en behandling må alltid balanseres mot troen på antatt effekt av en behandling eller intervensjon.

Hvem må være i equipoise for at en randomisert kontrollert studie skal være etisk? I mitt eksempel mente et tverrfaglig miljø bestående av både klinikere og forskere med spisskompetanse på neonatal alloimmun trombocytopeni at planlagt intervensjon ville gi færre komplikasjoner enn ikke-intervensjon. Det rådet ikke equipoise i denne gruppen, derfor var ikke randomisering etisk forsvarlig.

Veatch lanserer imidlertid tanken om at det ikke trenger å herske equipoise verken hos klinikere eller forskere for at randomi-

sering skal være etisk forsvarlig, så lenge pasienten er tilstrekkelig informert og samtykker uten noen form for tvang (9). Debatten omkring begrepet synes å være tilbake ved start. Det individuelle equipoisebegrepet har fått sin renessanse. Men i sin resirkulerte versjon er det nå pasienten som er i fokus. Det medisinskfilosofiske miljøet er altså i god takt med samfunnets øvrige helsepolitiske utvikling – eller er det kanskje motsatt?

Ulikt kunnskapssyn?

Equipoise for hva og hvem og hvorfor er viktige spørsmål. Et flertall vil være enige i at en eller annen form for usikkerhet vedrørende behandlingsalternativene er en forutsetning for å foreta randomisering. Men randomisering er ikke et mål i seg selv, kun et redskap for å oppnå ny kunnskap.

Randomisering vil alltid innebære skjevhet (bias) med hensyn til hva slags personer som inkluderes. Hvordan man vurderer utfallet av forsøket vil alltid være uttrykk for verdibaserte preferanser hos hver enkelt person, og personlige vurderinger avgjør om man er villig til å delta i en randomisering eller ikke. For eksempel vil det være en overvekt av mennesker som ønsker eksperimentell behandling som velger å delta i en

randomisert studie. Vi kan altså ikke anta at vi inkluderer «typiske» subjekter. Dette reiser i seg selv spørsmål om validiteten av et studium – randomisert eller ikke (6).

Randomisering gir ingen garanti for sikker og tilstrekkelig kunnskap. Tilhengere av evidensbasert kunnskap vil kanskje mene at randomisering er den eneste og absolutte forutsetning for gyldig kunnskap. Kritiske røster til randomisering har muligens et kunnskapsbegrep som omfatter mer enn statistisk signifikans, og man kan spekulere i om equipoise for denne gruppen er mer klinisk basert?

Klinisk skjønn er kunnskap som er vanskelig å måle i p-verdier og «power», men de færreste vil vel protestere mot at dette er en viktig dimensjon ved den samlede kunnskap. Debatten kan dermed illustrere et eksempel på hvordan tilsynelatende uenighet om metode i realiteten er uttrykk for verdibasert uenighet (10).

Det er u diskutabelt at en randomisert, kontrollert studie ville gitt bedre data for å vurdere effekten av screening og intervensjon. Spørsmålet er om de dataene som foreligger, likevel er gode nok som kunnskapsgrunnlag for å vurdere om screening for neonatal alloimmun trombocytopeni bør innføres i Norge. Dataene fra kohortstudien er prospektive, godt kvalitetssikret og kommer fra et stort og representativt materiale.

Svaret på om kunnskapen fra screening- og kostnad-nytte-studiene om neonatal alloimmun trombocytopeni er gyldig eller ubrukelig, avhenger av hva slags kunnskapsbriller man leser med. Utgangspunktet for debatten er ny og banebrytende kunnskap med eklatant potensial i vestlig nyfødtnet-

disin. Når validitet av resultater og valg av metode diskuteres, er det viktig at metodekritikk ikke forveksles med paradigmatisk diskrepans.

Oppgitte interessekonflikter: Heidi Tiller arbeider ved samme avdeling som flere av forskerne bak screeningstudien.

Litteratur

1. Kjeldsen-Kragh J, Killie MK, Tomter G et al. A screening and intervention program aimed to reduce mortality and serious morbidity associated with severe neonatal alloimmune thrombocytopenia. *Blood* 2007; 110: 833–9.
2. Killie MK, Kjeldsen-Kragh J, Husebekk A et al. Cost-effectiveness of antenatal screening for neonatal alloimmune thrombocytopenia. *BJOG* 2007; 114: 588–95.
3. Spencer JA, Burrows RF. Feto-maternal alloimmune thrombocytopenia: a literature review and statistical analysis. *Aust N Z J Obstet Gynaecol* 2001; 41: 45–55.
4. Fretheim A. Cost-effectiveness analysis of screening for neonatal alloimmune thrombocytopenia was based on invalid assumption. *BJOG* 2008; 115: 412–3.
5. Killie MK, Kjeldsen-Kragh J, Husebekk A et al. Author response to: Cost-effectiveness analysis of screening for neonatal alloimmune thrombocytopenia was based on invalid assumption. *BJOG* 2008; 115: 413–4.
6. Kørner H. Nye helseteknologier – fra utvikling til innføring. *Tidsskr Nor Lægeforen* 2003; 123: 340–3.
7. Freedman B. Equipoise and the ethics of clinical research. *N Engl J Med* 1987; 317: 141–5.
8. Gifford F. Freedman's «clinical equipoise» and sliding-scale all-dimensions-considered equipoise. *J Med Philos* 2000; 25: 399–426.
9. Veatch RM. The irrelevance of equipoise. *J Med Philos* 2007; 32: 167–83.
10. Ashcroft R. Equipoise, knowledge and ethics in clinical research and practice. *Bioethics* 1999; 13: 314–26.

Manuskriptet ble mottatt 26.3. 2008 og godkjent 3.7. 2008. Medisinsk redaktør Anne Kveim Lie.