

Släng inte ut observationsstudier med badvattnet

Bedöm deras kvalitet i stället



MÅNS ROSÉN, professor, direktör
rosen@sbu.se
SUSANNA AXELSSON, odont dr,
projektledare
JONAS LINDBLOM, farm dr, pro-

jektledare; samtliga
Statens beredning för medicinsk
utvärdering (SBU)

Randomiserade kontrollerade studier (RCT) har metodologiska fördelar, eftersom eventuella systematiska fel mellan försöks- och kontrollgrupperna görs om till ett slumpfel där osäkerheten hanteras med statistisk metodik. Icke-experimentella studier där försöksledaren inte har kunnat kontrollera för vilka som hamnat i försöks- respektive kontrollgruppen kallas för observationsstudier.

Observationsstudierna kan ha som mål att identifiera riskfaktorer för en sjukdom eller att studera effekter och risker av en viss åtgärd som kan vara såväl preventiv som behandlande. I alla dessa typer av observationsstudier jämförs en grupp med en annan.

Den typ av observationsstudier som är fokus i denna artikel är kontrollerade kohortstudier där man jämför en grupp som fått behandling med en grupp som fått alternativ behandling eller ingen behandling alls.

Syftet med denna artikel är att diskutera och belysa för- och nackdelar med kontrollerade observationsstudier i relation till RCT och diskutera vilka krav man ska ställa på kontrollerade observationsstudier.

Randomiserade studier inte alltid bäst

Fördelarna med RCT har med rätta resulterat i att allt fler behandlingsinsatser utvärderas med denna typ av studieuppläggning. Av olika skäl är det inte alltid möjligt att göra RCT, och man får då basera sina slutsatser på observationsstudier, oftast kohort- eller fall-kontrollstudier. Det kan t ex av etiska skäl vara svårt att göra randomiserade försök. Pap smear-screening är ett sådant exempel där screeningen startade utan en RCT, men där nedgången i incidens i cervixcancer efter att screeningen startat varit så kraftfull även på befolkningsnivå att ingen kan förneka effekterna. Det vore därför oetiskt att nu starta en RCT.

När det gäller samhällsinriktade förebyggande insatser är det inte heller möjligt att randomisera på individnivå. I vissa fall kan observationsstudier vara bättre än RCT, t ex för att bedöma långsiktiga risker av olika åtgärder där det krävs stora patienturval och lång uppföljningstid. RCT kan även ha andra begränsningar som bristfällig randomisering och snäva inklusionskriterier för de patienter som ingår i studien. Det finns

»Tyvärr finns en spridd uppfattning att observationsstudier generellt är mindre tillförlitliga än RCT.«

också en risk för systematiska fel (bias) på grund av ekonomiska intressen när RCT finansieras av industrin. Det finns naturligtvis metodologiska problem också med observationsstudier, och tillförlitligheten i RCT och observationsstudier måste bedömas enbart utifrån kvaliteten i studierna.

Nyanserad debatt behövs om rangordning av studier

Utgångspunkten vid en systematisk granskning av studier som undersöker medicinska metoder har ofta varit att man rangordnar efter studietyp, dvs RCT antas generellt ha högre kvalitet än kontrollerade kohort- och fall-kontrollstudier. Längst ned på rangordningen finns fallstudier eller s k eminensbaserade uttalanden.

Det är dock viktigt att påpeka att denna rangordning kan förändras vid en noggrann kvalitetsgranskning av studierna. Tyvärr finns en spridd uppfattning att observationsstudier generellt är mindre tillförlitliga än RCT. Att kategoriskt påstå detta är en felsyn, och organisationer som Cochrane Collaboration och SBU kan med sina strikta kriterier ha bidragit till denna uppfattning, även om det inte varit meningen.

Vi hoppas med denna artikel nyansera debatten också kring observationsstudiers värde.

RCT versus observationsstudier

Under lång tid har debatten pågått om värdet av RCT jämfört med observationsstudier.

En tidig studie av Sacks och medarbetare från 1982 jämförde 50 RCT med 56 studier där man använt historiska kontroller [1]. Med historiska kontroller avsåg man patienter som hade behandlats tidigare och som hade samma hälsoproblem. Åtgärderna som studerades bedömdes effektiva i 20 procent av de randomiserade studierna och i 79 procent av studierna med historiska kontroller. Författarna drog slutsatsen att risken för bias var större i observationsstudier med historiska kontroller.

I New England Journal of Medicine publicerades år 2000 två studier av Benson och Hartz [2] respektive Concato och medarbetare [3] som väckte stor uppmärksamhet för att de kom fram till slutsatsen att valgjorda observationsstudier inte systematiskt överskattade behandlingseffekterna jämfört med RCT.

Ioannidis och medarbetare fann en hög statistisk korrelation på 0,75 mellan resultaten mätt som oddskvoter i RCT och ob-

SAMMANFATTAT

Det finns en felaktig, men ofta upprepad uppfattning att man bara kan lita på resultaten från randomiserade studier och att observationsstudier inte är tillförlitliga.

Analysen visar att det är kvaliteten som avgör. Det finns många exempel på att såväl randomiserade studier som observationsstudier visat felaktiga resultat om de har varit små eller kvalitetsmässigt undermåliga.

Välgjorda observationsstudier ger resultat som nästan alltid

överensstämmer med resultaten från randomiserade studier. Undantagen har oftast gällt preventiva och besvärslindrande insatser där man inte kontrollerat för socioekonomiska förhållanden. I mindre än 30 procent av alla behandlingsinriktade observationsstudier har man i ledande medicinska tidskrifter kontrollerat för socioekonomi.

Det behövs en mer nyanserad inställning till observationsstudiers värde baserad på deras metodologiska kvalitet.

servationsstudier. Skillnaderna minskade ytterligare om man bara inkluderade prospektiva studier [4].

I en nyligen publicerad Cochranerapport fann man att effekten ofta var likvärdig mellan studietyperna, men att effekten var större i observationsstudier än i RCT [5]. Författarna hade av någon inte preciserad anledning exkluderat »klassiska observationsstudier« och studier som använt stora administrativa databaser. I övrigt hade alla studier inkluderats oberoende av om de uppfyllde vissa kvalitetskrav. Många av dessa observationsstudier hade använt historiska kontroller.

Syftet med de flesta av dessa jämförande studier mellan RCT och observationsstudier är att analysera skillnader i effektstorlek mellan olika behandlingar eller preventiva insatser. I många, men långtifrån alla, av dessa studier har effekten varit större i observationsstudier än i RCT [5-7]. McKee och medarbetare fann dock inte att observationsstudier gav systematiskt större effekter än randomiserade försök [8]. De fann vidare att deltagare i RCT som utvärderade behandlingseffekter var sjukare och lägre utbildade än de som inte deltog. I RCT som utvärderade preventiva insatser var förhållandet det motsatta. Deltagare var friskare och högre utbildade än de som inte deltog.

De som funnit att observationsstudier gett högre effektstorlek än RCT har tagit det som intäkt för att observationsstudierna överskattat effekterna. Detta gäller dock enbart under förutsättningen att resultaten från RCT är det sanna värdet [9].

Vår bedömning av litteraturen är att man ska vara särskilt försiktig med att tolka resultaten från observationsstudier som utvärderar behandlingsinsatser när det gäller studier som är små, baserade på historiska kontroller, där försöksledaren gjort uppföljningen, där utfallsmåttet är självrapporterade besvär och där kontroll av förväxlingsfaktorer (confounders) saknas eller varit bristfällig.

Många problem med observationsstudier är gemensamma med dem för RCT, men det kritiska problemet i observationsstudier är naturligtvis att få likartade försöks- och kontrollgrupper. Indikations- och selektionsbias är det största specifika problemet i observationsstudier jämfört med RCT. I övrigt har de likartade metodproblem. Vägjorda observationsstudier visar dock nästan alltid resultat som är likartade med RCT. Det finns dock ett fåtal undantag från denna regel, som ofta tas upp i debatten kring observationsstudiernas kvalitet.

Fel i enskilda studier – oberoende av studietyp

Det finns många exempel på att enskilda studier oberoende av studietyp kan visa resultat som senare visar sig vara felaktiga. De kan ha stora metodologiska brister eller vara så små att slumpen spelat ett spratt. Detta gäller även randomiserade studier. En metaanalys av 14 RCT för hypertoni behandling och kranskärlssjukdom visade att enskilda studier varierade i sina punktestimat från 0,49 till 1,60 med poolade resultat kring 0,86 [10]. Benson och Hartz redovisar 16 RCT som jämför skillnader i infektionsrisk för laparaskopi med öppen kirurgi vid appendektomi [2]. Tre av 16 randomiserade studier visar punktestimat över 1, även om de poolade resultaten pekar på en klart reducerad risk för laparaskopi.

LeLorier och medarbetare visade på stor heterogenitet mellan metaanalyser utförda på små RCT och resultaten från stora RCT [11]. Mulrow visade att 8 av 33 RCT som analyserade intravenös streptokinasbehandling efter hjärtinfarkt pekade på förhöjda dödlighetsrisker, trots att den samlade metaanalysen visade på en klart lägre dödlighetsrisk [12].

Aprotinin har använts i 14 år för att minska blodförlusterna efter bypass-kirurgi. I januari 2006 publicerades en observationsstudie som visade på en högre dödlighetsrisk för patienter

som fått detta medel än för patienter som fått alternativ behandling [13]. Ytterligare en observationsstudie som publicerades i mars samma år fann en ökad risk. Det amerikanska läkemedelsverket, FDA, fann dock ingen anledning vid det tillfället att skicka ut en varningssignal, eftersom RCT inte pekade på en överrisk.

En senare publicerad Cochrane-rapport med en metaanalys av 52 små RCT visade inte heller någon ökad risk [14]. Det var först när en större RCT stoppats i förtid på grund av att aprotinin verkade ha en överrisk som läkemedlet drogs in.

I New England Journal of Medicine redovisas ytterligare två vägjorda observationsstudier som visar ökad dödlighetsrisk [15, 16]. Det verkar alltså ha funnits en övertro på RCT som förenat beslutet att dra in aprotinin. Redaktören i New England Journal of Medicine drar flera viktiga lärdomar och understryker bl a problemen med att slå ihop data från små studier och behovet av att jämföra en metod med en annan i stället för med placebo (»head to head«-jämförelser) [17].

Exemplen kan mångfaldigas och gäller för alla typer av studier. Detta är inte förvånansvärt när studierna varit små, och det pekar mer på behovet av stora studier och mer samlade analyser.

Systematiska fel i observationsstudier

Observationsstudier kan liksom RCT visa resultat som senare visar sig vara felaktiga. Det är ytterst sällsynt att många studier pekar i samma felaktiga riktning, men det har förekommit i enstaka fall. De två exempel som ofta anförs som exempel på observationsstudiernas bristande tillförlitlighet är östrogenbehandling och risken för hjärt-kärlsjukdom samt antioxidanter och vitaminers inverkan på hjärt-kärlsjukdom och lungcancer. I båda fallen pekade flera stora observationsstudier på att riskerna för hjärt-kärlsjukdomar respektive lungcancer skulle minska av östrogenbehandling eller med ett högt intag av vitamin C och E. Senare gjorda RCT visade inte på någon riskminskning.

Problemet i båda fallen var att man inte tillräckligt kontrollerat för att försöks- och kontrollgruppen hade likartad sammansättning avseende viktiga bakgrundsfaktorer. I en del studier kontrollerades för rökning, fysisk aktivitet och andra riskfaktorer för hjärt-kärlsjukdomar, men inte för socioekonomiska faktorer. Många studier har visat att kvinnor som tagit östrogen lever sundare och är mer välutbildade än de som inte tagit östrogen. Humphrey och medarbetare respektive Lawlor och medarbetare visade att den relativa riskminskningen för hjärt-kärlsjukdom försvann när man kontrollerade för socioekonomi [18, 19]. Detsamma gäller för antioxidanter och vitaminer, där studier också visat att riskminskningen försvunnit när man kontrollerat för socioekonomi [20]. Intaget av vitaminer är högre hos socioekonomiskt väletablerade grupper.

Vad kan vi lära oss av detta?

Den primära svårigheten med observationsstudier är att få jämförbara grupper och/eller att kunna kontrollera för tänkbara olikheter. I en del av dessa studier hade man kontrollerat för riskfaktorer för hjärt-kärlsjukdomar utan att de relativa riskerna förändrades märkbart. Sociala och ekonomiska faktorer är sannolikt förutom ålder de mest betydelsefulla riskfaktorerna för förtida död och sjuklighet när det gäller många sjukdomar. Socioekonomi är såväl en oberoende riskfaktor för sjukdom och död som en indirekt riskfaktor, eftersom levnads-

»Vägjorda observationsstudier visar dock nästan alltid resultat som är likartade med RCT.«

»Vår hypotes är att det är särskilt viktigt att kontrollera för socioekonomiska faktorer ...«

vanor som rökning, alkohol, kost och fysisk aktivitet varierar mellan olika socioekonomiska grupper.

Vår hypotes är att det är särskilt viktigt att kontrollera för socioekonomiska faktorer när man studerar preventiva, symptomlindrande och icke-akuta insatser. Välutbildade personer med god social position har mer kunskap om vilka möjligheter som står till buds och bättre förmåga att framföra och få igenom sina krav. I de fall åtgärderna kostar pengar har de också med större sannolikhet nödvändiga ekonomiska förutsättningar. Det här gäller särskilt preventiva insatser eller insatser av mer symptomlindrande karaktär.

Hypotesen stöds av många epidemiologiska studier, t ex den ovan citerade studien av McKee och medarbetare [8]. Preliminära resultat från läkemedelsregistret visar t ex att välutbildade kvinnor och män hämtar ut östrogen respektive sildenafil i större utsträckning än lågutbildade. Det har sannolikt inte med behoven att göra. Lågutbildade som har högre risk för hjärt-kärlsjukdomar får däremot också i större utsträckning hjärt-kärlmedel än högutbildade, utom när det gäller dyrare läkemedel [21].

Ett annat exempel är användning av demensläkemedel, som visar sig vara vanligare bland högutbildade än lågutbildade, trots att förekomsten av demens är större bland lågutbildade [22]. Invandrare får inte rekommenderade läkemedel efter hjärtinfarkt i samma utsträckning som svenskfödda, vilket kan ha med språksvårigheter eller ekonomi att göra [23].

Ett liknande mönster kan sannolikt ses för planerade och oplanerade operationer. En studie visade att välutbildade män får tillgång till planerade kranskärlsinterventioner i högre utsträckning än lågutbildade män [24]. När det gäller akuta och livräddande insatser spelar sannolikt socioekonomisk position mindre roll. Hälso- och sjukvården måste då snabbt fatta beslut och ta hand om alla som kommer till vården.

Socioekonomi borde kontrolleras betydligt mer

Mot bakgrund av att socioekonomi är en mycket stark riskfaktor för sjukdom och död borde denna faktor kontrolleras när man jämför effekterna av olika behandlingsinsatser. I etiologiskt inriktade studier är det också relativt vanligt att kontrollera för socioekonomi när det gäller att förstå orsaksmekanismerna. När det gäller behandlingsinriktade, kontrollerade observationsstudier är det sällan förekommande. Det finns undantag, men de är få.

Ett utmärkt exempel på en observationsstudie där man kontrollerat för socioekonomi är den danska studien kring autism och vaccination mot MPR (mässling, påssjuka och röda hund) [25]. En mindre del av alla behandlingsinriktade observationsstudier som publiceras har dock kontrollerat för socioekonomi. En analys av behandlingsinriktade observationsstudier som publicerades under 2006 i de fyra högrankade tidskrifterna *New England Journal of Medicine*, *Lancet*, *JAMA* och *BMJ* visade att över 70 procent inte kontrollerat för socioekonomi. I mindre högrankade tidskrifter är siffran säkert betydligt högre.

Här finns goda möjligheter att öka kvaliteten i observationsstudier.

Studietyper räcker inte för rangordning av studiekvalitet

En genomgång av litteraturen visar att man inte a priori kan rangordna studiers kvalitet efter studieuppläggning, dvs att

RCT alltid är bättre än observationsstudier osv. Utgångspunkten kan vara att RCT principiellt har stora metodologiska fördelar, men det behövs en noggrann analys av andra kvalitetskriterier för att bedöma enskilda studiers sammanlagda kvalitet. Vägjorda observationsstudier visar nästan utan undantag resultat som är relativt lika de som fås i RCT. Vägjorda och stora RCT ger dock de säkraste resultaten när det gäller att uppskatta behandlingseffekter.

Observationsstudier som senare visar sig uppvisa »felaktiga« resultat är oftast små, baseras på historiska kontroller eller har dålig kontroll av förväxlingsfaktorer. Många observationsstudier har dock kontrollerat för riskfaktorer till den sjukdom de studerar, men sällan för socioekonomiska förhållanden. Socioekonomi är en oerhört stark riskfaktor för många sjukdomar, och det borde vara regel att man tar med denna faktor i sin kontroll av förväxlingsfaktorer när man gör behandlingsinriktade observationsstudier.

Stora observationsstudier, oftast registerbaserade studier, har en del klara fördelar. Slumpvariationerna blir små eftersom de baseras på totalundersökningar eller stora patientpopulationer. En fördel är att utfallet mäts i ett senare skede utan att försöksledaren kan påverka vilka som följs upp.

En stor fördel är också att de kan användas för att upptäcka långsiktiga risker eller bieffekter även för sällsynta utfall. Det är mindre risk för indikationsbias när man mäter oväntade utfall [26]. En läkare som skriver ut ett läkemedel vet vad medlet ska åstadkomma, men oftast inte vilka bieffekter det kan leda till för den enskilda patienten. För okända effekter som inte har påverkat forskrivningsbeslutet är risken för bias mycket mindre än när man mäter förväntade utfall. När det gäller positiva effekter har registerstudier samma problem som andra observationsstudier när det gäller att kontrollera för förväxlingsfaktorer.

Samkörningar med andra register ger dock möjlighet att kontrollera för en del bakgrunds faktorer som socioekonomi, läkemedelskonsumtion, tidigare sjukhusvistelser etc.

Våra slutsatser

Våra huvudbudskap kan sammanfattas i följande punkter:

- Randomiserade försök har stora metodologiska fördelar, och det är viktigt att fler stora och vägjorda RCT görs på såväl nya som etablerade metoder där vi har bristfällig kunskap. Tro inte att vi menar något annat.
- Var kritisk mot små studier, särskilt om de baseras på jämförelser med historiska kontroller, om försöksledaren gjort uppföljningen, om utfallsmåttet är självrapporterade besvär eller om kontroll av förväxlingsfaktorer saknats eller varit bristfällig.
- Underkänn inte observationsstudier med automatik och med argumentet att det finns en hierarkisk ordning där RCT alltid är bättre än observationsstudier.
- Vägjorda och stora observationsstudier med lång uppföljning där man har »hårda« resultatmått som död eller allvarlig sjuklighet och kontrollerat för tänkbara förväxlingsfaktorer har en god evidens som med stor sannolikhet överensstämmer med »sanningen«. Vid preventiva och besvärslindrande åtgärder är det särskilt viktigt att studierna kontrollerat för socioekonomiska förhållanden.
- Om du planerar att göra en observationsstudie, gör en studie med tillräcklig styrka och kontrollera för tänkbara förväxlingsfaktorer. Försök alltid att kontrollera för socioekonomi.

Internationella riktlinjer ger hjälp

På flera håll i världen pågår ett viktigt arbete med att utveckla riktlinjer för att rapportera och kvalitetsgranska observations-

studier [27-30]. Internationella grupper har formulerat riktlinjer för observationsstudier, benämnda STROBE, Strengthening the reporting of observational studies in epidemiology [27], respektive TREND, Transparent reporting of evaluations with nonrandomized designs [28]. För förebyggande arbete har riktlinjer utvecklats av bl a US Preventive Services Task Force [29]. GRADE-systemet (Grading of recommendations assessment, development and evaluation) borde också ha bättre möjligheter att nyansera evidensgraderingen kring observationsstudier [30].

Ingen av dessa grupper har dock ännu uppmärksammat den stora betydelsen av att kontrollera för socioekonomiska förhållanden.

■ *Potentiella bindningar eller jävsförhållanden: Inga uppgivna.*

Kommentera denna artikel på lakartidningen.se

REFERENSER

1. Sacks H, Chalmers TC, Smith H Jr. Randomized versus historical controls for clinical trials. *Am J Med.* 1982;72:233-40.
2. Benson K, Hartz AJ. A comparison of observational studies and randomized controlled trials. *N Engl J Med.* 2000;342:1878-86.
3. Concato J, Shah N, Horwitz RI. Randomized, controlled trials, observational studies, and the hierarchy of research designs. *N Engl J Med.* 2000;342:1887-92.
4. Ioannidis JP, Haidich AB, Pappa M, Pantazis N, Kokori SI, Tektonidou MG, et al. Comparison of evidence of treatment effects in randomized and nonrandomized studies. *JAMA.* 2001;286:821-30.
5. Kunz R, Vist G, Oxman AD. Randomisation to protect against selection bias in healthcare trials. *Cochrane Database Syst Rev.* 2007;(2): MR000012.
6. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ.* 1999; 319:312-5.
7. Worall J. Why there's no cause to randomize. *Br J Philos Sci.* 2007; 58:451-88.
8. Mulrow CD. Systematic reviews: rationale for systematic reviews. *BMJ.* 1994;309:597-9.
9. Mangano DT, Tudor JC, Dietzel C. The risk associated with aprotinin in cardiac surgery. *N Engl J Med.* 2006;354:353-65.
10. Henry DA, Carless PA, Moxey AJ, O'Connell D, Stokes BJ, McClelland B, et al. Anti-fibrinolytic use for minimising perioperative allogenic blood transfusion. *Cochrane Database Syst Rev.* 2007;(4):CD001886.
11. Schneeweiss S, Seeger JD, Landon J, Walker AM. Aprotinin during coronary-artery bypass grafting and risk of death. *N Engl J Med.* 2008; 358:771-83.
12. Shaw AD, Stafford-Smith M, White WD, Phillips-Bute B, Swaminathan M, Milano C, et al. The effect of aprotinin on outcome after coronary-artery bypass grafting. *N Engl J Med.* 2008;358:784-93.
13. Humphrey LL, Chan BKS, Sox HC. Postmenopausal hormone replacement therapy and the primary prevention of cardiovascular disease. *Ann Intern Med.* 2002;137:273-84.
14. Lawlor DA, Davey Smith G, Brucksdorfer KR, Kundu D, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet.* 2004;363:1724-7.
15. Ringbäck Weitoft G, Ericsson Ö, Löfroth E, Rosén M. Equal access to treatment? Population-based follow-up of drugs dispensed to patients after acute myocardial infarction in Sweden. *Eur J Clin Pharmacology.* 2008;64(4):417-24.
16. Haglund B, Köster M, Rosén M. Inequality in access to coronary revascularisation in Sweden. *Scand Cardiovasc J.* 2004;38:334-9.
17. Madsen KM, Hviid A, Vestergaard M, Schendel D, Wohlfahrt J, Thorsen P, et al. A population-based study of measles, mumps, and rubella vaccination and autism. *N Engl J Med.* 2002;347:1477-82.
18. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet.* 2004; 363:1728-31.
19. Des Jarlais DC, Lyles C, Crepaz N; TREND group. Improving the reporting quality of nonrandomized evaluations of behavioral and public health interventions: the TREND statement. *Am J Public Health.* 2004;94:361-6.
20. Atkins D, Best D, Briss PA, Eccles M, Falck-Ytter Y, Flottorp S, et al; GRADE Working Group. Grading quality of evidence and strength of recommendations. *BMJ.* 2004;328: 1490.