

Hvordan gjøre opp et klinisk materiale?

Epidemiologiske metoder i pasientnær forskning

Per G. Farup

*Forskningsenheten, Sykehuset Innlandet HF, Gjøvik, og
Enhet for anvendt klinisk forskning, Det medisinske fakultet, NTNU, Trondheim*

Korrespondanse: Forskningsjef, prof. Per G. Farup, Forskningsenheten, Sykehuset Innlandet HF, Gjøvik, N-2819 Gjøvik, Norway
Telephone: +47 61157000 Telefax: +47 61157439 E-mail: Per.Farup@ntnu.no

SAMMENDRAG

Bakgrunn. Korrekt forskningsdesign er avgjørende for forskningens kvalitet og må tilpasses forskningsspørsmålet. Bearbeiding av kliniske materialer er observasjonsstudier med epidemiologiske metoder, også kalt klinisk epidemiologi.

Eksempler på klinisk epidemiologi. De vanligste epidemiologiske design er tverrsnittsstudier, pasientkontrollstudier og kohortstudier. Egne eksempler på pasientnær forskning med disse metodene presenteres, resultatene vurderes, og studienes sterke og svake sider relateres til design.

Diskusjon. De forskjellige metoder kompletterer hverandre, det er ikke ett design som er egnet for alle problemstillinger. Selv om design er avgjørende for forskningskvaliteten, er god metode langt mer enn bare design i snever forstand og inkluderer alt fra planlegging til publikasjon. Gjennomføringen og generaliserbarheten er sammen med valg av design avgjørende for kvaliteten og anvendelsen av resultatene.

Konklusjon. Kunnskap om design og metode er nødvendig for å bearbeide et klinisk materiale og for kritisk lesning av medisinsk litteratur.

INTRODUKSJON

Det finnes ikke ett design som er egnet for alt, forskningsdesignet må tilpasses forskningsspørsmålet, og resultater fra studier med forskjellige metoder supplerer hverandre (1). Det randomiserte, kontrollerte og ofte blinde forsøk (RCT) er gullstandarden for noen kliniske problemstillinger, men ikke for alle. Selv ved problemstillinger hvor RCT er å betrakte som gullstandarden kan raskere, enklere og billigere metoder gi pålitelige resultater. Observasjonsstudier slik som tverrsnittundersøkelser, kohortstudier, pasientkontrollstudier og varianter av disse epidemiologiske design er også gullstandard for enkelte kliniske problemstillinger. "Klinisk epidemiologi" er et begrep som anvendes, selv om det språklig er ulogisk, når kliniske problemstillinger studeres med observasjonelle metoder. Retrospektive metoder (pasientkontrollstudier) tar utgangspunkt i syke personer. I befolkningsepidemiologiske studier anvendes metoden for å finne årsak til sykdom, mens den i klinisk epidemiologi kan studere den diagnostiske verdi av tidligere eller aktuelle symptomer og tegn (2,3). Prospektiv befolkningsepidemiologi (kohortstudier) studerer årsak til utvikling av sykdom i en frisk populasjon, mens klinisk epidemiologi kan ta utgangspunkt i en syk populasjon og studere forløpet og betydningen av prognostiske faktorer (2,4).

Bearbeiding av kliniske materialer er observasjonsstudier – man observerer hvordan en tilstand var, er eller utviklet seg. Man skal være seg bevisst hvilket design man bruker og designets sterke og svake sider.

Artikkelen viser egne eksempler på kliniske studier med observasjonelt design, og diskuterer prosjektens sterke og svake sider.

EKSEMPLER PÅ KLINISK EPIDEMIOLOGI

Vi har studert funksjonelle mage-tarmplager, kosthold og matintoleranse i flere observasjonsstudier (5-7).

Tverrsnittundersøkelse

Åttifire pasienter med irriterbar tarm syndrom (IBS) ble inkludert i en tverrsnittundersøkelse (6). Vi kartla matintoleranse, kostvaner, muskel-skjelettplager og psykiske faktorer, og gjorde objektive tester med tanke på matvareallergi/malabsorpsjon. Sytti prosent av personene anga relasjon mellom mageplagene og ett eller flere matslag (gjennomsnitt 4,8 matslag) og 62% reduserte eller utelot ett eller flere matslag fra kosten (gjennomsnitt 2,5 matslag). Samsvaret mellom pasientenes angivelse av hvilke matslag de reagerte på og objektive tester for matallergi/malabsorpsjon var ikke signifikant. Det var heller ingen signifikant assosiasjon mellom matintoleranse og psykiske symptomer eller muskel-skjelettplager. Derimot var det en signifikant assosiasjon mellom muskel-skjelettplager og psykiske symptomer, slik det er kjent fra andre undersøkelser.

Pasientkontrollstudie

Deretter sammenlignet vi et utvalg pasienter med IBS med en gruppe friske frivillige i en pasientkontrollstudie (5). Vi samlet anamnesticke opplysninger om

melkeintoleranse og utførte laktosemalabsorpsjonstest med registrering av symptomer i relasjon til testen. Melkeintoleranse og symptomer etter inntak av laktose var signifikant hyppigere hos personer med IBS, mens prevalensen av laktosemalabsorpsjon var den samme i de to gruppene. Sensitivitet og spesifisitet av melkeintoleranse og symptomer etter inntak av laktose med henblikk på laktoseintoleranse ble beregnet. Det viste seg vanskelig å finne gode symptombaserte diagnostiske kriterier for laktosemalabsorpsjon.

Kohortstudie

Pasientene ble fulgt opp i en prospektiv kohortstudie (7). Alle pasientene ble tilbudt kostveiledning av dietetikere som ga konkrete råd om kostendringer der det ble funnet indisert. Ca. halvparten fikk konkrete råd om kostendringer. Ved oppfølging etter seks måneder vurderte vi forløpet av IBS og faktorer som påvirket forløpet. Det var en moderat bedring i symptomscore i gruppen som helhet, og størst bedring hos pasienter uten psykisk komorbiditet. Bedringen var signifikant større i gruppen med diare-dominert IBS som hadde fått konkrete kostråd enn i de andre gruppene. Vi konkluderte derfor med at prognosen for IBS var god med tendens til bedring over tid, spesielt for personer uten psykisk komorbiditet, og at kostholdsveiledning er nyttig for gruppen av IBS-pasienter med diaredominerte symptomer.

DISKUSJON

Tverrsnittundersøkelser

Tverrsnittundersøkelser er øyeblikksbilder som beskriver prevalensen av en eller flere tilstander eller risikofaktorer på et gitt tidspunkt og tillater analyse av assosiasjoner mellom de målte variabler. Designet gir raskt svar (ingen tidsaspekter slik som oppfølging), er relativt billig, metodene kan enkelt standardiseres, det er ingen manglende oppfølging og lite problemer med manglende data (3). Fordi tidsaspektet mangler gir designet ingen informasjon om årsak til sykdom, sykdomsvarighet, sykdomsforløp eller årsakssammenhenger, men beskriver assosiasjoner mellom de observerte variablene. Svakheten ved det manglende tidsperspektivet forsøker man å redusere ved å innhente opplysninger om fortiden gjennom spørsmål om tidligere levevaner, kosthold, røykevaner etc. med den svakhet ("recall bias") som ligger i denne type data. Dessuten må man alltid vurdere om gruppen som deltar i tverrsnittundersøkelsen er representativ for den populasjon man vil ha informasjon om (seleksjonsbias). Utgangspunktet for vår studie var personer med IBS og alarmsymptomer (vekttap, blod i avføringen etc.) som først ble utredet for å utelukke organisk sykdom og deretter inkludert i studien. Muligheten for seleksjonsbias er til stede fordi vi inkluderte bare pasienter med alarmsymptomer, men vi mener likevel at gruppen er representativ for personer med IBS og at studien gir et godt

bilde av prevalensen av matintoleranse hos personer med IBS. Assosiasjonsanalysene er interessante, men tolkningen av assosiasjoner i tverrsnittstudier er bare hypotesedannende og sier intet om årsak. Våre negative funn utelukker hverken at matallergi, psykiske faktorer eller muskel-skjelettplager er korrelert med eller årsak til matintoleranse, og assosiasjonen mellom psykiske faktorer og muskel-skjelettplager sier intet om årsakssammenheng. Slike tverrsnittundersøkelser er pålitelige prevalensundersøkelser men utover dette ofte bare hypotesedannende og et godt utgangspunkt for videre forskning slik som RCT (intervensjonsstudier) eller prospektive kohortundersøkelser.

Pasient-kontrollstudier

Pasient-kontrollstudier har et dårlig rykte pga. fare for bias, men designet er ofte overraskende godt og har sine sterke sider. Det kan gi nyttig informasjon med få pasienter og det er relativt enkelt å gjennomføre. Det er raskt, billig, spesielt godt egnet for sjeldne tilstander, kan undersøke mange etiologiske faktorer, har ingen problemer med oppfølging, og er ofte hypotese-genererende (3). Laktosemalabsorpsjon er en relativt sjelden tilstand i Norge (prevalens ca. 4%) og kan derfor egne seg for et slikt design, og gjennomføringen med friske sykehusansatte som kontroller var grei og billig i forbindelse med tverrsnittstudien. Designet har sine begrensninger og svakheter. Det er et retrospektivt design som hverken kan måle prevalens eller insidens, og det måler bare ett endepunkt. Bias er et spesielt stort problem for pasient-kontrollstudier. Det anvendes informasjon om tidligere hendelser som kan inneholde systematiske feil pga. måten data samles på og personenes hukommelse. I vår undersøkelse er det svært sannsynlig at pasientene i større grad enn de friske frivillige hadde merket seg tidligere perioder med melkeintoleranse ("recall bias"). Derimot er tidligere biologiske data, f.eks. blodprøver, og aktuelle data slik som testen for laktosemalabsorpsjon, mer pålitelig. Det er viktig at pasientgruppen og kontrollgruppen er sammenlignbare og representative for den samme befolkning. Kontrollene kan være andre pasientgrupper, kollegaer, venner, naboer, et tilfeldig utvalg av befolkningen etc., og kan bli tilordnet ("matched") med pasientgruppen (2,3,8). Man kan bruke flere kontrollgrupper og variere forholdet mellom pasienter og kontroller. Historiske kontroller anbefales ikke (9). Noen av svakhetene ved designet unngås i en "innkapslet" (eng. "nested") pasient-kontrollstudie hvor data hentes fra en prospektiv kohortundersøkelse. I vår studie tror vi ikke at kontrollgruppen var selektert selv om vi ikke kan se bort fra at friske personer som selv hadde mistanke om laktoseintoleranse kan være overrepresentert fordi de ønsket å få testet seg selv. Materialet tillot beregning av den diagnostiske nøyaktighet av symptomet "melkeintoleranse". Det er gode retningslinjer for gjennomføring og rapportering av diagnostiske studier (The STARD Initiative) (10).

Kohortundersøkelse

En kohortundersøkelse følger en gruppe personer over tid og relaterer endepunktene i oppfølgingsperioden til variabler ved inklusjon (2,4). Mens man i befolknings-epidemiologi tar utgangspunkt i den friske befolkningen, er pasientgrupper ofte utgangspunkt for klinisk epidemiologi. Designet tillater beregning av insidens og studier av årsakssammenheng fordi personene følges over tid. Det er en styrke at mange variabler (både ved oppstart og avslutning) kan undersøkes og at metodene kan standardiseres og måles nøyaktig. Kohortstudier er tidkrevende, ofte kostbare, uhensiktsmessige ved sjeldne tilstander, og utsatt for bias pga. manglende oppfølging. Å bruke designet til å vurdere effekten av en intervensjon, slik vi gjorde, er langt fra ideelt pga. bias både i seleksjon til behandling og i vurderingen av resultat (1,11,12). Resultatene våre, som viser at prognosen er god med tendens til bedring over tid, tror vi er holdbare, men om dietetisk veiledning er nyttig for personer med diareddominert IBS er langt mer usikkert. En RCT med randomisering til dietetisk veiledning eller tradisjonell oppfølging ville gitt et langt mer pålitelig svar. Generelt er kohortundersøkelser velegnet til studier av prognose, mens randomiserte studier er langt å foretrekke for vurdering av en intervensjon (1,11,12). En retrospektiv kohortundersøkelse (også betegnet som historisk prospektiv eller historisk kohortundersøkelse fordi man i ettertid undersøker forløpet hos en gruppe pasienter) vil stort sett ha de samme problemer med standardisering av målemetoder og usikre angivelser som pasientkontrollundersøkelser, men er langt billigere og enkle å utføre enn en prospektiv kohortundersøkelse.

Planlegging og valg av design

Oxford Centre of Evidence-based Medicine har en anbefaling om det ideelle design for forskjellige problemstillinger (12). Et RCT av høy kvalitet og med smale konfidensintervaller er gullstandarden for vurdering av etiologi, terapi, prevensjon og hyppige bivirkninger i studier hvor randomisering er mulig, etterfulgt av kontrollerte observasjonsstudier og ukontrollerte studier. På bunnen av den hierarkiske listen finnes opinionsundersøkelser og ekspertuttalelser (11,12). En vel gjennomført prospektiv kohortstudie er derimot overlegen et RCT for vurdering av prognose, diagnose, differensial diagnose, symptom prevalens og sjeldne bivirkninger (1,12). Selv om pasient-kontrolldesign ikke får toppkarakter ved noen problemstilling kan det være overlegent andre metoder, spesielt ved sjeldne tilstander (3).

Metode er mer enn bare et design i snever forstand og inkluderer alt fra planlegging til publisering, slik som målsetning, forskningshypotese, valg av pasienter

og kontroller, statistiske metoder, klinisk relevante forskjeller, beregning av studiestørrelse, studiestyrke etc. (1,13). Har man, eller kan man skaffe, det nødvendige pasientmateriale med kvalitativt gode nok data for å trekke pålitelige og klinisk signifikante konklusjoner?

Generelt

Selv om de fleste undersøkelser viser at både RCT og epidemiologiske studier er gode og gir sammenfallende resultater hvis studiene er av høy kvalitet, settes metodene ofte opp mot hverandre i et gradert hierarki, og diskusjonene om hva som er best diskuteres også i de mest prestisjetunge tidsskriftene (10,14-16). Vi har eksempler også fra den senere tid på at metodene gir motstridende resultater. Observasjonelle studier har konkludert med at menopausal hormonbehandling og antioksidierende vitaminer beskytter mot hjerte/kar-sykdommer, mens RCT konkluderer med at hormonbehandling øker insidensen av hjerte/kar-sykdommer og at antioksidanter mangler beskyttende effekt i behandlingsstudier (17,18). Det er også reist tvil om små mengder alkohol virkelig har en helsefremmende effekt, slik det er fremkommet i observasjonsstudier (19). Diskrepansen mellom studieresultatene forklares ved at konfundere i større grad påvirker resultatene i observasjonsstudier enn i RCT.

Design er viktig for kvaliteten, men for kvaliteten på sluttproduktet er kvaliteten på gjennomføringen og generaliserbarheten kanskje like viktig som designet (1,13). En dårlig gjennomført studie med upålitelige data men riktig design gir ikke et riktigere bilde av sannheten enn en mindre ideelt designet studie med god gjennomføring (20). Den eksterne validiteten (generaliserbarheten) er like viktig som den interne validiteten (21). Generaliserbarheten er kanskje bedre i observasjonsstudier hvor man i større grad tar hensyn til pasientens ønsker, enn i RCT hvor pasientens behandling er tilfeldig og hvor en større andel pasienter ekskluderes (22). Trolig kompletterer resultatene fra observasjonsstudier og RCT hverandre (1,22). I mange studier er nettopp "mine" pasienter (gamle, barn, kvinner i fertil alder etc.) utelukket fordi inklusjons- og eksklusjonskriteriene er så spesielle at de fleste pasientene faller utenfor, eller metodene for behandling eller oppfølging så avvikende fra egne rutiner at resultatene har liten overføringsverdi.

Kunnskap om design og metode er nødvendig for å bearbeide et klinisk materiale og for kritisk lesning av medisinsk litteratur. Er kvaliteten tilfredsstillende og resultatene generaliserbare? Systematiske oversiktsartikler og resultater fra pragmatiske kliniske studier med tilstrekkelig antall pasienter, høy kvalitet og et vel tilpasset design er den beste veiledning for daglig klinisk praksis (1).

REFERANSER

1. Glasziou P, Vandenbroucke JP, Chalmers I. Assessing the quality of research. *BMJ* 2004; **328** (7430): 39-41.
2. Jacobsen G. Analytisk epidemiologi – pasient-kontroll- og kohortstudier [Analytical epidemiology – case-control and cohort studies]. *Tidsskr Nor Laegeforen* 2002; **122** (27): 2636-2640.
3. Newman TB, Browner WS, Cummings SR, Hulley SB. Designing an observational study: Cross-sectional and case-control studies. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, editors. *Designing clinical research*. Philadelphia: Lippincott Williams & Wilkins, 2001: 107-124.
4. Cummings SR, Newman TB, Hulley SB. Designing an observational study: Cohort studies. In: Hulley SB, Cummings SR, Browner WS, Grady D, Hearst N, Newman TB, editors. *Designing Clinical Research*. Philadelphia: Lippincott Williams & Wilkins, 2001: 95-106.
5. Farup PG, Monsbakken KW, Vandvik PO. Lactose malabsorption in a population with irritable bowel syndrome: prevalence and symptoms. *A case-control study*. *Scand J Gastroenterol* 2004; **39** (7): 645-649.
6. Monsbakken KW, Vandvik PO, Farup PG. Perceived food intolerance in subjects with irritable bowel syndrome – etiology, prevalence and consequences. *Eur J Clin Nutr* 2006; **60** (5): 667-672.
7. Monsbakken KW, Vandvik PO, Farup PG. The value of a general therapeutic approach in subjects with irritable bowel syndrome. *Aliment Pharmacol Ther* 2005; **21** (1): 21-27.
8. Grimes DA, Schulz KF. Compared to what? Finding controls for case-control studies. *Lancet* 2005; **365** (9468): 1429-1433.
9. Sacks H, Chalmers TC, Smith H, Jr. Randomized versus historical controls for clinical trials. *Am J Med* 1982; **72** (2): 233-240.
10. Bossuyt PM, Reitsma JB, Bruns DE, Gatsonis CA, Glasziou PP, Irwig LM, Lijmer JG, Moher D, Rennie D, de Vet HC. Towards complete and accurate reporting of studies of diagnostic accuracy: The STARD initiative. *Ann Intern Med* 2003; **138** (5): 40-44.
11. Barton S. Which clinical studies provide the best evidence? The best RCT still trumps the best observational study. *BMJ* 2000; **321** (7256): 255-256.
12. Phillips B, Ball C, Sackett D, Badenoch D, Straus S, Haynes B, Dawes M. Levels of evidence. *Oxford – Centre for Evidence Based Medicine* 2006, http://www.cebm.net/levels_of_evidence.asp.
13. Concato J, Horwitz RI. Beyond randomised versus observational studies. *Lancet* 2004; **363** (9422): 1660-1661.
14. Vandenbroucke JP. When are observational studies as credible as randomised trials? *Lancet* 2004; **363** (9422): 1728-1731.
15. Benson K, Hartz AJ. A comparison of observational studies and randomized, controlled trials. *N Engl J Med* 2000; **342** (25): 1878-1886.
16. Pocock SJ, Elbourne DR. Randomized trials or observational tribulations? *N Engl J Med* 2000; **342** (25): 1907-1909.
17. Lawlor DA, Davey SG, Kundu D, Bruckdorfer KR, Ebrahim S. Those confounded vitamins: what can we learn from the differences between observational versus randomised trial evidence? *Lancet* 2004; **363** (9422): 1724-1727.
18. Col NF, Pauker SG. The discrepancy between observational studies and randomized trials of menopausal hormone therapy: did expectations shape experience? *Ann Intern Med* 2003; **139** (11): 923-929.
19. Jackson R, Broad J, Connor J, Wells S. Alcohol and ischaemic heart disease: probably no free lunch. *Lancet* 2005; **366** (9501): 1911-1912.
20. Kunz R, Oxman AD. The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. *BMJ* 1998; **317** (7167): 1185-1190.
21. Rothwell PM. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *Lancet* 2005; **365** (9453): 82-93.
22. McKee M, Britton A, Black N, McPherson K, Sanderson C, Bain C. Methods in health services research. Interpreting the evidence: choosing between randomised and non-randomised studies. *BMJ* 1999; **319** (7205): 312-315.