
FORELESNINGSNOTATER I KLINISK STATISTIKK:

DET RANDOMISERTE KLINISKE FORSØK

KULL 95B

4. april – 23. mai 2000

Strategier for kvantitative studier av sykdom og helse
Det randomiserte kliniske forsøk
Systematiske oversikter / Meta-analyser

Stein Emil Vollset
Seksjon for medisinsk statistikk
Institutt for samfunnsmedisinske fag
Universitetet i Bergen

KAPITTEL 1: STRATEGIER FOR KVANTITATIVE STUDIER AV SYKDOM OG HELSE	5
Det medisinske eksperiment	5
Observasjonsstudier	7
Hvilke strategier kan vi stole på?	10
KAPITTEL 2: DET RANDOMISERTE FORSØK	11
Eksempler på randomiserte studier	12
Etikk og det randomiserte forsøk	14
Randomisering	16
Valg av endepunkt	18
Valg av kontrollgruppe	19
Blinding	20
Analyse etter behandlingshensikt	20
Protokollen	20
Beregning av forsøksstørrelse	24
Feilkilder ved randomiserte forsøk	25
Rapportering av randomiserte forsøk	26
KAPITTEL 3: META-ANALYSER	28
Eksempler på meta-analyser	28
Protokoll for meta-analyse	31
Statistiske metoder for meta-analyse	32
Feilkilder ved meta-analyser	35
Litteraturliste	38

Kapittel 1: Strategier for kvantitative studier av sykdom og helse

Vi er alle, i større eller mindre grad, opptatt av egen og andres helse og sykdom. Friske ønsker kunnskap om risiko for å rammes av sykdom og ulykker og videre om faktorer som påvirker denne risikoen. Epidemiologien har som mål nettopp å kunne fortelle friske mennesker om risiko for sykdom: Om hvordan for eksempel kjønn, alder, leveste, omgivelser og arvelige forhold påvirker sykdomsrisiko og hvordan endringer i levevis kan minske eller øke fremtidig sykdomsrisiko.

Den som allerede er rammet av en mer eller mindre alvorlig sykdom ønsker kunnskap om hvordan sykdommen vil forløpe om man lar den gå sin egen gang uten inngripen av noe slag (sykdommens naturlige forløp), og om hvordan man kan intervensere for å påvirke forløpet i gunstig retning. Det siste er legens og den kliniske medisins fremste oppgave. Den kliniske epidemiologien har som mål å kartlegge faktorer som kan hjelpe pasienten og legen med å forutsi sykdommens forløp.

Vi skal i det følgende gjennomgå strategier, epidemiologiske og statistiske metoder som er nyttige i studier som søker å vinne kunnskap om sykdom og helse. Et viktig skille i studietyper går mellom observasjonsstudier og eksperimenter. I det kliniske eksperimentet manipuleres behandlingstilvalgene slik at tildeling av behandling skal være uavhengig av pasientens prognose. Dette oppnås vanligvis ved randomisering (tilfeldigheter avgjør hvilken av et sett adekvate behandlinger den enkelte pasient eller friske deltaker utsettes for). For observasjonsstudier griper man ikke aktivt inn, men nedtegner forhold av interesse og relaterer disse til senere sykdomsutviklingen eller sykdomsforløp.

Det medisinske eksperiment

Medisinsk eksperimentering har en dårlig klang i manges ører. Man assosierer gjerne med en kynisk vitenskapsmann (eller kvinne) som uten hensyntagen til pasientenes ve og vel vil kartlegge nye behandlingers virkninger og bivirkninger. Eksperimenter med radioaktive stoffer uten forsøkspersonenes vitende i USA og kanskje Norge på 1940, 1950-tallet og kanskje senere (66,71,78) er eksempler som undergraver legitim medisinsk forskning til pasientens beste. Mange vil hevde prinsipielt at man skal ikke eksperimentere med mennesker og enda mindre med pasienter som kan være svekket i sin evne til å ivareta egne interesser.

Er man kjent med det spinkle erfaringsgrunnlag mange medisinske prosedyrer hviler på og tenker nøye igjennom problemene forbundet med å vinne ny medisinsk kunnskap, vil det kliniske eksperiment fremstå som et viktig hjelpemiddel i situasjoner hvor det kan benyttes uten å bryte det fundamentale prinsipp i all klinisk virksomhet, nemlig at behandlingen ikke skal være til

skade og at alle pasienter skal gis den beste tilgjengelige behandling. Prinsipper for gjennomføring av medisinsk forskning som involverer mennesker er nedfelt i Helsinkideklarasjonen som vil bli nærmere gjennomgått i kapittelet om randomiserte forsøk. Ofte er nemlig situasjonen den at man ikke vet hvilken av flere tilgjengelige behandlinger som er den beste. For sykdommer hvor det ikke finnes noen kurativ behandling, må nye behandlinger, nær sagt hvor lovende de måtte være, etableres gjennom randomiserte kliniske forsøk. På den annen side er det ikke vanskelig å gi eksempler på situasjoner hvor det ville være fullstendig etisk uakseptabelt å gjennomføre et randomisert forsøk. Med en gang en ny behandling etableres som effektiv kan man ikke senere gjennomføre utprøvinger hvor ikke alle deltagere får lik tilgang på den effektive behandlingen.

Det randomiserte kliniske forsøk

Den viktigste kilde til sikker ny klinisk viten innen vestlig medisin er det randomiserte kliniske forsøk. Nå rapporteres årlig resultater fra mellom ni og ti tusen slike forsøk fra alle medisinske fagfelter. Det er særlig de siste 10-15 år det er blitt et selvsagt krav at nye medikamenter og behandlinger skal prøves under strenge betingelser. Det tok lang tid fra de første randomiserte forsøk ble gjennomført mot slutten av 1940-tallet til metoden ble anerkjent. Prinsippet for det randomiserte kliniske forsøk er enkelt. Man ønsker å undersøke om en ny behandling for en gitt sykdom er effektiv. Blant en gruppe pasienter som har gitt sitt skriftlige samtykke til å delta i eksperimentet, trekkes halvparten tilfeldig ut og får den nye behandlingen. Den resterende halvparten får tradisjonell eller ingen behandling. Forløpet av sykdommen sammenlignes så for de to gruppene. Det fundamentalt viktige prinsipp er at det skal være uforutsigbart, og helst ukjent så lenge forsøket pågår, hvilken av de alternative behandlingene den enkelte pasient får. Det første oppnås ved at tildeling av behandling er tilfeldig eller randomisert. Randomisert tildeling av behandling gir den store fordel at man kan forvente at de to gruppene blir like i alle andre henseender enn selve behandlingen. Vi velger å illustrere prinsippene for det randomiserte forsøket med en av de første slike studier som ble gjennomført.

I 1946 satte det britiske Medical Research Council ned en komité for å studere effekten av streptomycin på forløpet av tuberkulose. Bakgrunnen var at tidligere utprøvinger av behandling for tuberkulose var gjort uten kontrollgrupper. Et resultat av dette var den seiglivet myten om effekt av gull i tuberkulosebehandling. Mellom januar og september 1947 ble 109 pasienter rekruttert til studien. Behandlingen var enten 6 måneders sengeleie (C, for "control"), eller streptomycin og 6 måneders sengeleie (S). Tildeling av behandling C eller S skjedde ved at man ved hvert deltakersykehus, etter at en pasient var innlemmet i studien, åpnet en nummerert forseglett konvolutt som inneholdt et kort som foreskrev enten behandling C eller S. Kortene og konvoluttene var laget ferdig på forhånd av Professor Bradford Hill som lot en liste over tilfeldige tall avgjøre om det skulle stå S eller C på kortet. Det ble laget separate sett av konvolutter for hvert deltaker sykehus og for mannlige og kvinnelige pasienter. På denne måten forsikret man seg om at tilfeldighet alene bestemte om en pasient

fikk behandling S eller C. Femtifem pasienter ble fordelt til S-gruppen og 52 til C-gruppen. Pasientenes sykdomsforløp ble så fulgt i 6 måneder og er rapportert til stor detalj i en meget velskrevet artikkel i British Medical Journal i 1948 (1). Pasientene ble klassifisert etter 6 måneder i en av følgende 6 grupper:

- (1) Betydelig forbedring,
- (2) Moderat eller lett forbedring,
- (3) Ingen endring av betydning,
- (4) Moderat eller lett forverring,
- (5) Betydelig forverring og
- (6) Død.

Denne klassifiseringen ble gjort uten kjennskap til om den enkelte pasient tilhørte S eller C gruppen. Blant 32 pasienter hvor tilstanden ble beskrevet som betydelig forbedret viste det seg av 28 hadde fått streptomycinbehandling og 4 var i kontrollgruppen. Dette betyr at 51 % av streptomycinpasientene og 8% av kontrollgruppen opplevde betydelig forbedring etter 6 måneder. Sannsynligheten for at en så stor forskjell skulle fremkomme ved tilfeldighet er mindre enn en på en million.

Bradford Hill skriver senere at det var store problemer med å få akseptert prinsippet om at tildeling av behandling skulle skje ved en tilfeldig mekanisme, og han mente at man ikke hadde fått gjennomslag for studien om ikke leveringssituasjonen for streptomycin hadde vært slik at man bare kunne behandle ca. 50 pasienter (47). Det er langt lettere å akseptere randomisering til en lovende behandling når man ikke har nok til alle.

Observasjonsstudier

Åpenbart i studier av helseskadelige forhold eller faktorer, kan man ikke gjennomføre randomiserte forsøk. Et eksempel er studier av sigarettøyking. Her kan man ikke randomisere utdeling av sigaretter, men man må nøye seg med å studere forekomst av sykdom og død hos personer som selv har valgt å røyke og sammenligne med lignende personer som ikke røyker.

Kohort studien

Prinsippet for kohortstudien er like enkelt som for det randomiserte forsøket. Man kartlegger utbredelsen av den eller de faktorer man er interessert i en befolkningsgruppe. Befolkningen overvåkes fra dette tidspunktet og man ser om sykdomshyppigheten er avhengig av faktoren som studeres.

Vi vil nå illustrere prinsippene for denne studietypen med en viktig kohortstudie fra begynnelsen av 1950-tallet. I 1951 sendte to engelske epidemiologer, Richard Doll og Austin Bradford Hill, ut et spørreskjema til 59,600 britiske leger (31). Spørreskjemaet var enkelt. Legene ble spurt om:

1. navn, adresse, alder

2. om de var tobakksrøykere,
3. om de hadde røykt tidligere, men sluttet,
4. om de aldri hadde røykt så mye som en sigarett daglig i minst ett år.
5. Man spurte i tillegg røykerne om mengder røykt, alder ved røykestart, type tobakk og eventuell alder ved røykeslutt.

Man fikk tilbake 40,564 tilfredsstillende utfylte skjemaer. Blant 24,389 mannlige leger over 35 år var det 12.7 % ikke-røykere, prosentene som røykte 1-14 g, 15-24 g og 25 g og mer daglig var 34.6, 31.4 og 21.3.

I løpet av de første 29 måneder etter at spørreskjemaene var sendt ut døde 789 blant de mannlige legene over 35 år. Trettifem av disse dødsfallene skyldtes lungecancer. Alder-standardisert lungecancer-dødelighet var 0.0, 0.48, 0.67 og 1.14 pr. 1000 pr. år blant henholdsvis ikke-røykere, røykere av 1-14, 15-24 og 25+ g tobakk daglig. For koronarsykdom var dødelighetstallene 3.89, 3.91, 4.71 og 5.15 pr. 1000 pr. år. Man kunne altså allerede i 1954 se en klar relasjon mellom lungecancer og koronarsykdom på den ene siden og økende forbruk av tobakk på den andre siden.

Nylig er 40-års oppfølging av disse legene publisert (32). Førsti-års oppfølgingen gir viktig ny informasjon om røykingens skadevirkninger. Man kunne for eksempel direkte beregne sannsynlighetene for å overleve fra 35 til 70 år for ikke-røykere (ca. 80 %) og stor-røykere (ca. 50 %). Dette betyr at tre av ti røykere dør mellom 35 og 70 år på grunn av røykingen. Å vinne slik kunnskap kan ikke ta mindre enn 35 år og illustrerer viktigheten av å ha et generasjonsperspektiv når medisinsk viten skal innhentes.

Kasus-kontroll studien

Vi har sett at kohortstudien og det randomiserte forsøket ofte er langvarige og ressurskrevende operasjoner. En mer kostnadseffektiv måte å studere sykdommers årsaker er kasus-kontroll studien.

Forut for kohort studien som er beskrevet ovenfor hadde Doll og Bradford Hill gjort en kasus-kontroll studie (30).

Mellom april 1948 og oktober 1949 meldte 20 sykehus i London alle pasienter med cancer i lunge, mage (ventrikkel), tykktarm (colon) og endetarm (rectum). Studien rapporterer 649 tilfeller av lungekreft blant menn og 60 blant kvinner. Alle disse pasientene ble intervjuet om røykevaner. For hver pasient med lungekreft ble det intervjuet en pasient uten kreftsykdom av noe slag av samme kjønn, i samme 5-års aldersgruppe og innlagt i samme sykehus på samme eller omtrent samme tid.

Blant de 649 mannlige lungekreft pasientene var det 2 ikke-røykere (0.3%), mens det blant de 649 kontroll-pasientene var 27 ikke-røykere (4.2 %). Blant de 60 kvinnelige lungekreftpasientene var det 19 ikke-røykere (31.7%) og 32 blant de 60 kontrollene (53.3%). Sett i historisk lys er det meget skrem-

mende at over 95% av pasientene i Londonsykehus på slutten av 1940-tallet var røykere.

Om vi antar at kontrollpasientenes røykevaner er representative for den generelle befolkningen antyder disse tallene at lungekreft insidensraten (antall nye tilfeller lungekreft pr. 100,000 friske pr. år) er 14 ganger høyere blant menn som røyker enn blant menn som ikke røyker. For kvinner er det tilsvarende tall, som også kalles insidens-rate ratio eller rate ratio, 2.4. Disse tallene er ikke langt fra de anslag vi opererer med i dag. En tilsvarende kasus-kontroll studie ble gjennomført samtidig i USA (112).

Tverrsnittsstudien

Tverrsnittsstudien kartlegger status på et gitt tidspunkt og er nyttig når man for eksempel skal kartlegge forekomsten (prevalensen) av en sykdom eller planlegge helsetjenester. Det kan være spørreskjema- eller intervjuundersøkelser som har til hensikt å kartlegge levevis, helsetilstand, forbruk av medikamenter og helsetjeneste på et visst tidspunkt. Som eksempel kan nevnes den stor legekårsundersøkelsen som ble gjennomført 1993 i regi av den norske lægeförening. Ellers gjennomfører Statistisk sentralbyrå stadig undersøkelser av denne typen. Tverrsnittsstudien har mindre verdi når det gjelder å finne faktorer som disponerer for sykdom eller har prognostisk verdi. Grunnen til dette er at tverrsnittsstudien kun gir et øyeblikksbilde som ikke inneholder informasjon om relasjoner som utvikler seg over tid.

Den økologiske studien

En type tverrsnittsundersøkelse som ofte benyttes er å sammenligne sykdomshyppighet og forekomst av sykdomsrelaterede faktorer i forskjellige geografiske områder eller land. Som eksempel skal vi se på en studie som sammenholdt forekomst antistoffer mot bakterien *Helicobacter pylori* og hyppigheten av kreft i magesekken blant 17 befolkningsgrupper i 13 land (97).

Helicobacter-studien viste at om man ordnet befolkningene etter økende forekomst eller dødelighet av ventrikk-cancer fant man en klar tendens til økende forekomst av antistoffer mot *Helicobacter pylori* i de samme befolkningene. Fordi økologiske studier er beheftet med alvorlige feilkilder og man lett finner assosiasjoner som ikke gjenspeiler årsaksforhold, gir denne studien kun svak støtte til hypotesen om at *Helicobacter pylori* forårsaker kreft i magesekken. Hypotesen må er derfor studert i mer pålitelige studietyper som kasus-kontroll- eller kohort-studier. Endelig bevis får man kanskje først om man kan påvise at vaksinasjon mot bakterien reduserer forekomst av ventrikk-cancer. Om en studie av sistnevnte type noensinne vil kunne gjennomføres er usikkert - og kanskje heller ikke ønskelig.

I en økologisk studie, basert på data fra 11 land, viste Doll en sammenheng mellom sigarettforbruk pr. innbygger i 1930 og dødelighet av lungecancer i 1950 (29).

Hvilke strategier kan vi stole på?

Om vår hensikt er å få kunnskap om årsaker til sykdom eller se på ulike behandlings effektivitet er det stor forskjell i pålitelighet blant de forskjellige studietypene. Det randomiserte kliniske forsøket vil når det er adekvat gjennomført gi den langt sikreste viten. Generelt har vi en pålitelighets-gradient fra den randomiserte forsøket til kohort-studien til kasus-kontroll-studien og til tverrsnitts- og økologiske studier. Disse to siste må aldri tillegges alt for stor vekt i utsagn om sykdommers årsaker. Avhengig av hvordan kasus-kontroll eller pasient-kontroll studien gjennomføres kan den variere i pålitelighet fra den man får fra en kohort-studie til det helt upålitelige.

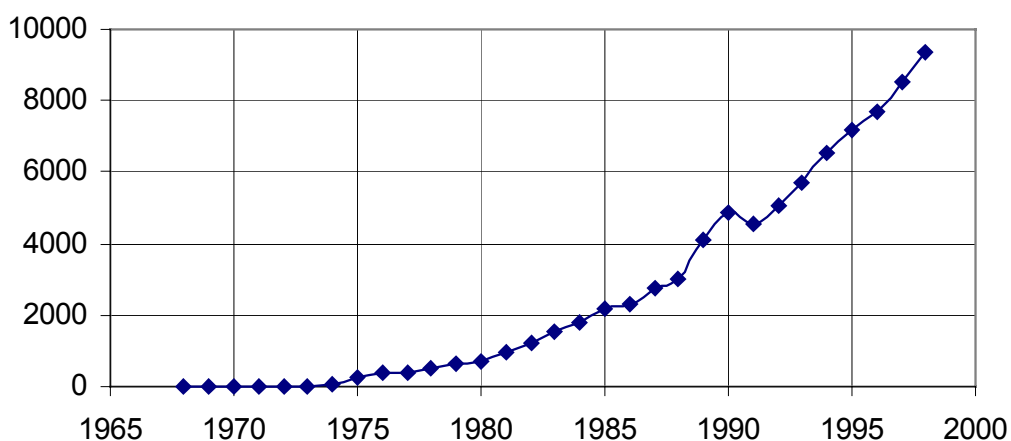
Kapittel 2: DET RANDOMISERTE FORSØK

I dette kapitlet skal vi i detalj gjennomgå det randomiserte forsøket, som er det kanskje viktigste instrumentet vi har for å evaluere medisinsk behandling og forebyggende tiltak. Hovedprinsippet for det randomiserte kliniske forsøket er at en tilfeldig og uforutsigbar mekanisme skal bestemme hvilken behandling den enkelte pasient skal få. Randomisering gjennomføres for å sikre at det behandlingsalternativet, som gis den enkelte pasient, tildeles uavhengig av den enkeltes prognose. Behandlingsalternativene, som oftest to, er bestemt på forhånd. Om mulig skal hverken pasienten eller helsepersonellet som har kontakt med pasienten vite hvilken av de alternative behandlingene den enkelte pasient får. Studien sies i så fall å være dobbel-blind.

Fra de første randomiserte kliniske forsøk på slutten av 1940-tallet og frem til midten av 1970-tallet var randomiserte studier ikke vanlige innen klinisk forskning. Siden den gang er antallet randomiserte forsøk som rapporteres økt fra noen hundre til anslagsvis nær 10,000 årlig i 1998/99. Figuren viser antall tidskriftartikler i MEDLINE hvor "randomized eller randomised" forekommer som tekstord. Dette er ikke noe nøyaktig mål for hvor mange randomiserte forsøk som gjennomføres, men gir en god pekepinn. Nå klassifiserer MEDLINE artikler direkte som "randomized-controlled-trials" og opptelling på denne publikasjonstypen gir 6737, 7159, 7873, 9131, 9938, 9651, 9209 og 9656 slike artikler for årene 1991-98. Dette antyder at tallene i figuren heller er for lave enn for høye.

Etterhvert er det naturlig nok blitt vanskelig å beholde oversikten over

Årlig antall artikler i PubMed med "randomi[z/s]ed" som tekstord



resultater fra randomiserte forsøk. Derfor har sammenfattende analyser eller meta-analyser av serier av randomiserte forsøk for samme problemstilling blitt

stadig vanligere de senere årene. Metoder for og problemer forbundet med meta-analyser vil vi gjennomgå i kaptittel 3.

Med et så stort volum av randomiserte forsøk og den store innflyelse disse har eller burde ha på medisinsk praksis(75,76) er det viktig å kjenne til de mange problemer disse forsøkene er forbundet med. Først og fremst er om problemstillingen er av en slik art at det er riktig og forsvarlig å anbefale eller anmode pasienter eller friske mennesker å delta i studien. Ofte er dette kontroversielt. En pågående studie gir tamoxifen til friske kvinner for å se om denne substansen beskytter mot utvikling av brystkreft. Samtidig har man lenge visst at tamoxifen langt fra er noe uskyldig stoff. Blant annet er tamoxifen mistenkt for å øke risikoen for kreft i livmoren (53,107).

Derneft er det svært kostbart og ressurskrevende å gjennomføre et randomisert klinisk forsøk. Det er ikke uvanlig at et randomisert klinisk forsøk koster flere 10-talls millioner kroner å gjennomføre. Forsøk med prislapper på flere hundretalls millioner kroner har vært gjennomført og flere er under planlegging. Man bør ikke la seg skremme av store prislapper. I langt større grad enn hva som er vanlig i dag burde randomisering inkorporeres i medisinsk praksis. Uten videre ekstra kostnader kunne man da få viktig kunnskap der er det tvil om terapeutiske alternativets effektivitet (20).

Videre må vi føle oss noenlunde trygge på at intervensjonen har effekt i aktuelle doser og at den blir gitt over et tilstrekkelig langt tidsrom. En nylig gjennomført randomisert studie hvor man ga vitaminene A og E til 29,000 finske røykere konkluderte med at disse stoffene ikke beskyttet mot lungekreft gitt over en 6 års periode relativt sent i livet (94). Man hadde særlig store forhåpninger til vitamin A og kanskje hadde man funnet en effekt om det hadde vært gitt tidligere i livet eller over et lengre tidsrom.

Eksempler på randomiserte studier

Det er vanskelig å trekke frem viktige studier blant de rundt 20-30,000 randomiserte studiene som er gjennomført de siste 5 årene eller blant de 30-50,000 som er gjennomført de siste 50 årene. Vi våger likevel å trekke frem noen studier som har hatt stor innflytelse og også studier som har vært gjennomført i Norge. Enger (38) gir en ypperlig innføring i det randomiserte kliniske forsøk og eksempler fra dets tidlige historie i Norge.

Et av de aller første randomiserte kliniske forsøk var streptomycin studien (1) som ble gjennomgått i stor detalj i kapittel 1. Behandling av pasienter etter hjerteinfarkt er et eksempel på et område innen medisinen hvor randomiserte forsøk har preget den terapeutiske utviklingen. Allerede på midten av 1960-tallet viste en norsk randomisert studie gunstig effekt av diettendringer på prognosen etter et hjerteinfarkt (58). I denne studien ble 206 kontrollpasienter og 206 diett-pasienter fulgt i fem år og 80 pasienter i diettgruppen fikk ny episode med hjertesykdom (nytt infarkt, angina pectoris eller plutselig død) mens 120 i kontroll-gruppen fikk en slik episode. Reduksjon av dødelighet etter gjennomgått hjerteinfarkt ved beta-blokkade ble tidlig vist i en stor

norsk multisenter undersøkelse (100). Senere er acetylsalicylsyre eller annen antikoagulasjonsbehandling blitt en selvsagt del av etterbehandling etter hjerteinfarkt (8). En viktig studie med warfarin-behandling etter hjerteinfarkt ble gjennomført i Norge (87). Den skandinaviske 4S-studien viste at kolesterolsenkende behandling kan bedre overlevelsen hos hjertepasienter (101). Et stor randomisert studie har også klart vist at jevnlig inntak av acetylsalicylsyre beskytter mot hjerteinfarkt hos tidligere friske menn (90). Også når det gjelder forebygging av hjertekarsykdom er pionerstudier gjennomført i Norge. I Oslo-studien ble i overkant av 1,200 menn med høy risiko for hjertesykdom randomisert til en intervensjons gruppe med diett-veiledning og røykeavvenning eller til en kontrollgruppe. Etter 5 år var det klart færre tilfeller av hjertesykdom i intervensjonsgruppen (19 av 604) enn i kontroll-gruppen (36 av 628) (48). Akutt-behandling av hjerteinfarkt har gjennomgått store endringer etter at store randomiserte studier har vist at trombolytisk behandling med for eksempel streptokinase reduserer dødelighet etter hjerteinfarkt (40). Ved disse studiene er reduksjonen så liten at et adekvat forsøk krever flere 10,000 pasienter. Til tross for disse studienes store størrelse er det fremdeles kontroversielt hvilken trombolytisk behandling som er å foretrekke.

Ved hjelp av randomiserte forsøk er behandlingen av tynntarmssår under radikal omveltning. Mange av disse pasientene kan nemlig effektivt behandles med en antibiotikakur mot bakterien *Helicobacter Pylori*. Innen all behandling av kreftpasienter er randomiserte forsøk en selvsagt del av evaluering og nyervervelse av kunnskap. Det samme gjelder alle terapeutiske fremskritt i behandling av AIDS, og nye vaksiner må etableres gjennom randomiserte forsøk (polio, hepatitt A, B, meningococc etc).

Et av de største og mest kostbare randomiserte forsøk som noen gang er gjennomført var utprøvingen av polio-vaksinen i USA på femtitallet (17,42). Denne studien hadde en randomisert komponent hvor 200,745 skolebarn ble vaksinert mens 201,229 fikk placebo-vaksine fra slutten av april til midten av juni 1954. Hvert barn fikk tre injeksjoner fra en eske som inneholdt 3 glass (à 10 doser) med aktiv vaksine og 3 glass (à 10 doser) med placebo-vaksine. Glassene var forsynt med et nummer og det var ikke mulig å vite hvem som fikk aktiv vaksine og hvem som fikk placebo. Eskene, hver med doser til 20 skolebarn, ble brukt på klassenivå slik at man sørget for balanse mellom vaksine og placebo i hver enkelt skoleklasse. Antallet poliomyelitt-tilfeller i perioden 1. mai til 31. desember 1954 blant de vaksinerte var 57 og antallet blant skolebarn som fikk placeboinjeksjon var 142. I tillegg fulgte man utviklingen av poliomyelitt blant rundt 220,000 personer som ble vaksinert uten randomisering og blinding og blant 1,200,000 kontroller som ikke fikk vaksine. Således var den totale studiepopulasjonen over 1.8 millioner skolebarn.

Randomiserte forsøk har vist at folsyre beskytter mot medfødt ryggmargsbrokk og kanskje andre misdannelser (26,65). Dette har ført til at helsemyndighetene i mange land nå anbefaler gravdide kost rik på folinsyre (Norge) eller kjemisk substitusjon (USA). Andre store vitaminforsøk har ikke vist gunstig effekt av vitaminer for å forebygge sykdom (46,94) mens en studie antyder gunstig effect av vitamin E hos pasienter med koronar sykdom (91). En

stort forsøk med stoffet interferon- β har for første gang forandret forløpet av multippel sklerose(99).

Langt fra alle randomiserte forsøk er vellykket i den forstand at de etablerer en ny effektiv behandling. Mange forsøk har vist at den aktive behandlingen har dårligere effekt enn placebo-gruppen. I enkelte tilfeller finner påviser man en ugunstig effekt av den nye behandlingen. Dette er viktige studier iden forstand at om ikke forsøket hadde vært gjennomført kan vi regne med at den uheldige behandlingen hadde blitt brukt på et langt større antall pasienter, kanskje langt inn i fremtiden.

Et dramatisk eksempel på denne siste type studier er CAST og CAST-II studiene som viste at ny antiarrytmisk behandling hos pasienter med hjertesykdom ga en betydelig overdødelighet til tross for at man oppnådde regularisering av hjerterytmen (64,95,96).

En høyaktuell problemstilling i 1999 er laserbehandling av myokard hos pasienter med behandlingsrefraktær angina pectoris(67,105). I dette feltet har det de siste årene vært publisert resultater fra en rekke studier, også flere randomiserte(2,18,19,43,63), som viser en klar bedring i subjektive mål (f.eks. angina funksjonsklasse). Ser man på mer objektive mål for behandlingseffekt (overlevelse, ejectivesjonsfraksjon, arbeids-EKG etc.) er resultatene langt fra entydige. Manglende blinding av behandlingen og av personell som evaluerer pasientene gjør at mange fortsatt er sterkt skeptiske til denne behandlingsformen. Ved Haukeland sykehus pågår et dobbelblindt forsøk med perkutan myokardial laserbehandling(67).

Etikk og det randomiserte forsøk

Helsinki-deklarasjonen

Det 18ende World Medical Assembly vedtok i et møte i Helsinki i juni 1964 et sett retningslinjer for forskning på pasienter. Denne såkalte Helsinki-deklarasjonen er siden revidert flere ganger, sist i forbindelse med World Medical Assembly's møte i Sør-Afrika i 1996.

Helsinki-deklarasjonen består av tre deler: Først gis 12 grunnleggende prinsipper som slår fast at biomedisinsk forskning på mennesker må foregå etter aksepterte vitenskapelige prinsipper og grundig kjennskap til vitenskapelig litteratur. En forskningsprotokoll, som alltid skal inneholde en vurdering av etiske problemer med forsøket, skal forelegges en uavhengig komité. En klinisk kompetent person skal ha ansvaret for forsøkspersonene. Videre heter det at viktigheten av studien må stå i forhold til risikoen den enkelte utsettes for. Interessene til den enkelte deltager må alltid veie tyngre enn samfunnets eller vitenskapens interesser. Leger må ikke delta i forskning som hvor risiko for den enkelte deltager er uforutsigbar og han eller hun må innstille undersøkelsen om det viser seg at risikoen overstiger en eventuell gevinst. Vitenskapelige arbeidere som ikke følger Helsinki-deklarasjonen bør ikke aksepteres for publisering. Alle potensielle forsøkspersoner skal informeres om studiens formål, metoder, forventet gevinst og risiko. Legen må så innhente pasientens infor-

merte samtykke, helst skriftlig. Det må legges stor vekt på at den enkelte pasient ikke føler noe som helst press til å delta i studien. Videre gis retningslinjer for de tilfeller der den potensielle forsøkspersonen er mindreårig eller ute av stand til å gi et informert samtykke.

Den andre delen av Helsinki-deklarasjonen gir i seks punkter retningslinjer for klinisk forskning. Det slås fast at det er tillatelig å prøve ut nye behandlinger om legen tror at det er til pasientens beste. Hovedprinsippet er at i enhver medisinsk studie skal alle pasienter, inkludert kontrollgruppen, sikres den beste tilgjengelige behandling.

Den tredje delen gir i 4 punkter retningslinjer for ikke-klinisk biomedisinsk forskning. Her slås det fast at deltakelse skal være frivillig og forskning som kan være skadelig for den enkelte ikke må foregå. Det gjentas at vitenskapens og samfunnets interesser ikke må gå på bekostning av den enkeltes velvære.

Helsinki-deklarasjonen's uttalte og gjentatte krav om at den enkelte pasient's interesser alltid skal veie tyngre enn samfunnets har vært problematisert (44).

Debatt om Helsinkideklarasjonen

Det er blitt gitt godkjenning til et stort antall (mer enn 15) randomiserte kliniske forsøk med placebo-kontroll for å utprøve behandlingsregimer for å forhindre HIV-infeksjon fra mor til nyfødt barn etter at amerikanske helsemyndigheter i 1994 anbefalte zimovudine peroralt i siste del av svangerskapet til alle HIV-positive gravide(7).

Denne anbefalingen var basert på et randomisert klinisk forsøk(24) som viste at smittehyppigheten kunne reduseres fra 25.5 til 8.3 %. De placebo-kontrollerte forsøkene som er godkjent etter 1994 har til felles at de ble planlagt utført i utviklingsland (de fleste i Afrika, men også i Thailand og den Dominikanske republikk). Begrunnelsen er at behandlingen som ble anbefalt i USA i 1994 var så kostbar (800 US\$) at den i praksis ikke kan tas i bruk i den fattige delen av verden. I for eksempel Uganda er denne summen 400 ganger mer enn det beløpet som brukes på helse per innbygger(11).

For å gjøre relevant forskning for å hjelpe HIV-positive gravide kvinner må man gjøre forsøk med billigere behandlinger og sammenligne mot standardbehandlingen lokalt hvor forskningen utføres. Dette innebærer et klart brudd på Helsinki-deklarasjonen slik den er formulert i dag og denne forskningen har derfor vært gjenstand for en omfattende debatt(11,62,70). Det er foreslått endringer i Helsinki deklarasjonen som tar høyde for denne problemstillingen. Denne debatten pågår høsten 1999(16,59).

Etiske komitéer i Norge

I Norge skal all medikamentutprøving meldes Statens Legemiddelkontroll. Videre skal alle biomedisinske forskningsprosjekter hvor det inngår forsøk på mennesker legges frem for en av fem regionale etiske komitéer. Disse komitéene har vært i arbeid siden 1980-tallet og er bredt sammensatt av to medisinsk

fagkyndige, en sykepleier, en representant for sykehuseierne, en jurist, en fagkyndig i etikk og en lekrepresentant. Komiteenes mandat er relativt vidt og omfatter både terapeutisk og ikke terapeutisk forskning på pasienter og friske forsøkspersoner som ikke er av en slik art at det kan regnes som en del av vanlig etablert behandling (14,54). Disse komitéene har utarbeidet et skjema som må fylles ut i forbindelse med søknaden. Det kreves som en hovedregel skriftlig informert samtykke av alle deltakere (13) og man legger stor vekt på utformingen av informasjonen til deltakerne i studien.

Randomisering

Hvorfor randomisere?

Man randomiserer for å sikre at behandlingsgruppene blir så like som mulig(6). Randomisering gir en *asymptotisk* garanti for at prognose-fordelingen i de forskjellige behandlingsgruppene er lik i utgangspunktet. Igjen betyr asymptotisk at garantien om likhet mellom gruppene blir bedre jo flere som fordeles til hver gruppe. Om alle pasienter har like god eller dårlig prognose er randomisering unødvendig. Det er kun når det er variasjon i prognose mellom pasientene at randomisering er nødvendig. Uheldigvis befinner vi oss nær sagt alltid i den siste situasjonen. Jo større den prognostiske variabiliteten er, jo større pasientmaterialer kreves før randomisering har den ønskede effekt. Vi merker oss at i små studier, og særlig når den prognostiske variabiliteten er stor, kan ulikheter mellom behandlingsgruppene oppstå selv ved adekvat randomisering.

Hvordan randomisere

Prinsippet for randomisering er enkelt(5). En tilfeldig mekanisme skal avgjøre hvilket behandlingsalternativ den enkelte pasient skal få. Tidligere lot man ofte behandlingen alternere slik at annenhver pasient fikk behandling A og annenhver behandling B, eller man lot pasientens fødselsdag avgjøre. Dette siste kunne gjøres slik at om fødselsdagen var oddetall ble behandling A gitt og for liketalls fødselsdager behandling B. Disse prosedyrene tilfredsstiller ved første øyekast de krav som stilles til adekvat randomisering. Dette er ikke tilfelle. Slike prosedyrer må aldri brukes. Grunnen er at det er synlig for helsepersonellet og kanskje også for pasienten hvilken behandling vedkommende vil komme til å få. Om enkelte leger, sykepleiere eller pasienten selv har preferanser for den ene eller den andre behandlingen er det rom for manipulering med innleggesrekkefølge eller pasienten kan holdes utenfor forsøket om fødselsdagen ikke stemmer med ønsket behandling. Slike mekanismer kan føre til opphopning av pasienter med svært god eller svært dårlig prognose på det ene behandlingsalternativet og resultatene fra forsøket kan bli ende opp langt fra sannheten. Dette illustrerer godt at det tilfeldige i seg selv ikke er nok, tildelingen av behandling må også være utforutsigbar. Vi har lært at ingen, hverken pasient eller behandler, skal ha noen som helst kunnskap om hvilke av behandlingsalter-

nativene som står for tur på det tidspunkt pasienten gir sitt informerte samtykke om å delta. Så snart dette samtykket er gitt bli pasienten randomisert. Det er også viktig at pasienter med sterke behandlings- preferanser ikke innlemmes i studien. Man kan da risikere at de trekker seg om forsøket ikke er blindet eller bivirkninger røper den aktuelle behandling i et blindet forsøk.

En gjennomgang av et større antall kliniske forsøk i behandling av akutt infarkt viste at manglende eller inadekvat randomisering kan føre til store prognostiske forskjeller mellom behandlings- og kontroll-gruppen. Denne forskjellen i prognose kan være forklaringen på at ikke-randomiserte studier gjennomgående finner større behandlingseffekter enn randomiserte studier(22). Også blant randomiserte studier har man sett at manglende blinding av randomiseringsresultatet frem til behandlingen er definitiv eller må være åpenbar bidrar til å overdrive behandlingseffekten (55,86).

Enkel randomisering

Med enkel randomisering til to grupper er det for hver enkel deltager 50 % sjanse for at han eller hun skal bli fordelt til hver av behandlingene. Selv om man i praksis ikke kaster mynt og krone for hver enkel pasient ville dette være et eksempel på enkel randomisering. Vi forstår av dette at enkel randomisering ikke garanterer likt antall pasienter på hver av de to behandlingene. Faktisk vil gruppestørrelsen følge en binomisk fordeling. Om man ønsker bedre balanse i gruppestørrelse må man bruke det som kalles blokk-randomisering.

Blokk-randomisering

Blokk-randomisering får man om man lager i stand et antall kort som tilsvarer det antallet pasienter man ønsker i studien. På halvparten av kortene skriver man behandling A og på den andre halvparten behandling B. Man stokker så kortene godt og noterer rekkefølgen de ligger i. Denne blir randomiseringssekvensen. Her har man kontroll over hvor mange pasienter som havner på hvert behandlingsalternativ mens rekkefølgen er tilfeldig. Om man ikke vet hvor mange pasienter man kommer til å rekruttere til studiene kan man operere med en blokk-størrelse som er garantert mindre enn pasientantallet og så repetere prosessen når den første blokken er brukt opp. På denne måten er man garantert likt antall på hver behandling etter hver blokk. Om man gjør blokkstørrelsen liten oppstår det mønstre i randomiseringsrekkefølgen slik at man mot slutten av hver blokk kan forutsi den neste pasient's behandling. Dette er uønsket, men kan motvirkes ved å variere blokkstørrelsen.

Stratifisert randomisering

Om det er stor variasjon i prognose blant pasientene som inngår i studien er det fare for at pasienter med enten god eller dårlig prognose skal dominere en av behandlings- gruppene. I så fall er det umulig å skille effekten av prognose fra effekten av behandlingen man studerer. Om man har kunnskap om viktige prognostiske variabler på randomiseringstidspunktet, kan man randomisere

separat innen hver prognose-gruppe. Således sikrer man seg mot ubalanse mellom behandlingsgruppene på viktige prognostiske variabler som alder eller stadieinndeling av kreftpasienter. Stratifisert randomisering må suppleres med blokk-randomisering innen hvert stratum for å være effektiv (73).

Minimisering

Minimisering er en metode som benyttes i større og større grad når pasienter rekrutteres over tid. Om det iløpet at randomiseringsprosessen oppstår en ubalanse i på forhånd definerte variabler vil sannsynligheten økes for at neste pasient havner i behandlingsgruppen som vil redusere denne ubalansen (106).

Randomisering i praksis

Det er viktig at randomiseringen foregår på en enkel og sikker måte. Straks pasienten har gitt sitt skriftlige informerte samtykke til å delta i studien skal han eller hun registreres som deltaker og deretter randomiseres. I store studier foregår det gjerne ved at man kontakter et randomiseringssenter over telefon og får oppgitt behandlingen. Om behandlingen gis akutt må denne telefonen være bemannet 24 timer. Ellers kan det være hensiktsmessig at hvert senter eller lege som deltar i studien har fått tilsendt et sett nummererte konvolutter som inneholder behandlingskoden og som skal åpnes etter at pasienten er registrert. Det må redegjøres for disse konvoluttene ved studiens slutt. Eksempler på andre adekvate prosedyrer er at pasienten skriver seg inn i en studie protokoll hvor hvert nummer på forhånd er koplet til ett av behandlingsalternativene. Om studien ikke er blindet er det svært viktig å ha strenge prosedyrer omkring randomiseringen slik at hverken pasienter eller studiepersonell kan manipulere med tildeling av behandling. Det er sterke indisier for at slik manipulering, bevisst eller ubevisst, ikke er uvanlig i praksis (22,55,86).

Valg av endepunkt

Medisinsk behandling eller annen intervensjon kan ha et kortsiktig eller langsiktig siktemål. Hos en pasient med akutt hodepine er siktemålet å kurere hodepinen i løpet av 10 til 30 minutter. Hos en pasient med høyt blodtrykk uten symptomer er siktemålet å forhindre sykdom på lengre sikt, kanskje flere ti-år inn i fremtiden. Enten hovedsiktemålet er lang- eller kort-siktig må god medisinsk behandling også tilfredstille andre krav. I behandlingen av hodepinen vil man kreve at medikamentet ikke har langsiktige uheldige bivirkninger. Det smertestillende stoffet phenacetin viste seg å gir nyrekomplikasjoner. Om siktemålet er langsiktig vil kreve at behandlingen ikke gir uforholdsmessig store umiddelbare plager eller bivirkninger.

I tabellen viser vi eksempler på tilstander hvor man i kliniske forsøk benytter surrogatendepunkter for å redusere antall pasienter og tiden det tar å gjennomføre forsøket (34,36,37,74,111). Bruk av surrogatendepunkt kan medføre dramatiske besparelser. I evaluering av behandling av akutt hjerteinfarkt

kreves det titusener av pasienter og oppfølging over mange år om man vil studere reduksjon i dødelighet. Om man istedet studerer reetablering av sirkulasjon i infarktområdet eller hjertets ejsjonsfraksjon kan studien gjennomføres med kanskje et hundretall pasienter og oppfølging i bare et par uker. Tolkningen og nytten av de sistnevnte studier forutsetter imidlertid at surrogatendepunktet gir informasjon om det egentlige endepunktet. Ofte er korrelasjonen mellom surrogatendepunktet og det man egentlig er ute etter dårlig dokumentert og mindre pålitelig enn man antar.

Tilstand	Egentlig endepunkt	Surrogat endepunkt
Inoperabel coloncancer	Død	Reduksjon av tumorstørrelse
Akutt hjerteinfarkt	Død	Åpning av okkludert kar
Hypertensjon	Slag	Senkning av blodtrykk
Diabetes	Retinopati	Kontroll av glucosenivå
Hyperkolesterolemi	Arteriosclerose	Kolesterolsenkning
Vaksine-utvikling	Fravær av klinisk sykdom	Antistoffrespons
Multippel sclerose	Kur av sykdom	Redusert antall hjerneplaques
HIV infeksjon	Utvikling av AIDS	CD4 tall

Man gir blodtrykksmedisin for å hindre tidlig død eller sykdom på lengre sikt. Dette betyr at det tar mange tiår å gjennomføre adekvate blodtrykksstudier. Derfor ser man ofte kun på den umiddelbare blodtrykksenkende effekt og slutter derfra at medikamenter som effektivt senker blodtrykket i dag er gunstig på lengre sikt. Det samme gjelder lipidsenkende medikamenter. Istedenfor å studere arteriosclerose eller prematur død direkte ser man på om medikamentet påvirker pasientens lipid-profil i antatt gunstig retning. I kortvarige studier på surrogatendepunkter har man ikke anledning til å studere langsiktige bivirkninger eller komplikasjoner på andre organsystemer.

Valg av kontrollgruppe

En kritisk avgjørelse i ethvert klinisk forsøk er hvilken behandling kontrollgruppen skal få. Om det ikke finnes noen standard behandling er valget enkelt. Kontrollgruppen får da ingen behandling, helst blindet slik at behandlingen tilsynelatende er den samme for alle, men uten aktivt virkestoff i kontrollgruppen (f.eks. placebotabletter, dvs. tabletter uten aktivt virkestoff).

I andre sammenhenger finnes det en eller flere etablerte behandlinger for den aktuelle tilstanden. Om effekten av disse er godt dokumentert må kontrollgruppen få den beste alminnelig aksepterte behandlingen. Det sier seg selv at dette ofte kan være kontroversielt. Det er dokumentert at det ikke er ualminnelig at kontrollgruppen får suboptimal behandling, ofte med placebo til tross for at det finnes etablert behandling (80,83). Årsaken til dette kan være at man vil forsterke den nye behandlingens overlegenhet. En slik praksis er selvfølgelig i strid med Helsinkideklarasjoen.

Blinding

Om mulig skal alle parter i et randomisert forsøk, legen og annet medisinsk personale såvel som pasienten være blindet for behandlingen som gis. Dette vil si at ingen skal vite hvem som får aktiv behandling og hvem som får kontroll-behandling eller placebo. Grunnen til at blinding er viktig er at kunnskap om behandling kan påvirke pasientens rapportering av egen tilstand og den kan påvirke legens evaluering av pasienten eller prøveresultater som vurderes skjønnsmessig, for eksempel røntgenbilder.

Et viktig, men ofte oversett poeng er at også randomiseringen må være blind. Dette for å unngå manipulering fra deltakere eller behandlere med hensyn til deltakelse i åpne studier (55,86).

Analyse etter behandlingshensikt

Ofte vil ikke alle pasienter fullføre behandlingen som de er randomisert til. Dette kan skyldes bivirkninger, at pasienten trekker seg fra forsøket av andre årsaker eller medikamentslurv. Slike endringer etter randomisering skal man ikke ta hensyn til i analysen av forsøket, men la randomiseringen definere behandlingsgruppene. Denne fremgangsmåten kalles analyse etter behandlingshensikt (engelsk: intention-to-treat analysis)(49) og må følges om man ikke ønsker å innføre uoverskuelige feilkilder. At analysen skal gjøres på denne måten er også en spore til å både hindre frafall fra studien og legge forholdene til rette for at pasientene i størst mulig grad følger den behandlingsanvisningen de er randomisert til.

Protokollen

Forut for ethvert klinisk forsøk skal det foreligge en protokoll. Dette er slått fast i Helsinki-deklarasjonen som også eksplisitt krever at en etisk vurdering skal inngå i protokollen. Protokollen skal beskrive hvordan forsøket skal gjennomføres. Alle detaljer må fremgå slik at i prinsippet skal forsøket kunne gjennomføres uten annen rettleiding enn den skriftlige protokollen. Forslag til disposisjon for protokoll for kliniske forsøk finnes en rekke steder (3,25,73) samt i Sosialdepartementets forskrifter for klinisk utprøving av legemidler (88). Tabellen gir forslag til disposisjon av protokollen. Vi vil nå ta for oss hvert av punktene i tabellen.

PROTKOLL FOR RANDOMISERT KLINISK FORSØK

- 1 **Generell informasjon**
- 2 **Begrunnelse for og hensikten med studien**
- 3 **Etikk.**
Drøft gevinst mot risiko.

	Sikres alle beste behandling?
	Pasientinformasjon
4	Forsøksplan (Design)
	Blindingsprosedyrer
5	Utvelgelse av forsøkspersoner
	Inklusjonskriterier
	Eksklusjonskriterier
6	Randomiseringsprosedyre
7	Behandlingsregimer
	Behandlingens lengde, doser
	Hvordan behandles kontrollgruppen
8	Måling av behandlingseffekt
	Definisjon av respons/endepunkt
	Sekundære endepunkter
	Oppfølgingslengde
9	Overvåkning
	Bivirkninger
10	Forsøksstørrelse/statistikk
	Plan for analyse av studien
	Statistisk begrunnelse for antall forsøkspersoner
	Overvåkning av langvarige studier, interim-analyser
11	Kvalitetssikring av studien.
	Overvåkning av compliance
	Håndtering av protokollavvik, frafall etc.
12	Håndtering av persondata.
	Eventuell søknad til Datatilsynet
13	Administrasjon av studien
14	Finansiering, rapportering, godkjenning, forsikring, etc.
15	Sammendrag, vedlegg, litteratur

1. **Generell informasjon** Protokollens førsteside bør inneholde forsøkets tittel og navn/adresse på den/de ansvarlige for studien samt institusjonsadresser med telefonnummer, fax, elektronisk postadresse. Finansieringskilder listes også på protokollens førsteside.
2. **Begrunnelse for og hensikt med studien** Her bør man gi en gjennomgang av tidligere studier med samme behandlinger og indikasjoner. Det bør begrunnes hvordan den aktuelle studien kan tilføre ny viten eller befeste nåværende kunnskap. Hensikten med studien bør gis i få og enkle setninger.
3. **Etikk** Her skal nytten av intervensjonen for hver enkelt forsøksperson vurderes mot den risikoen vedkommende løper. Det må godtgjøres at de som ikke får den eksperimentelle behandlingen får den beste alminnelig til-

gjengelige behandling. Det må også redegjøres for hvordan potensielle deltakere informeres om studien, spesielt om mulige ulemper eller skader som kan oppstå. All informasjon til deltakerne samt samtykkeerklæringen må vedlegges. Det må redegjøres for hvorfor man eventuelt må avvike fra å innhente deltakernes skriftlige samtykke.

4. **Forsøksplan** Her skal det redegjøres for forsøkets struktur. De aller fleste kliniske forsøk og den typen vi har drøftet i dette kapittelet er kontrollerte (en kontrollgruppe inngår i forsøket) og randomiserte (fordeling til behandlingsalternativer skjer tilfeldig) forsøk. Det er viktig å redegjøre for i hvilken grad det er mulig å blinde pasienten for hvilken behandling han eller hun får (enkel blind) og å blinde alt helsepersonell som har kontakt med pasienten (dobbel-blind). Enkel-blind situasjoner omfatter både en situasjon hvor legen, men ikke pasienten er blindet og omvendt at pasienten er blindet, men ikke legen. Viktigheten av forskjellige blindingsalternativer er avhengig av arten av sykdommen, behandlingen og måling av behandlingseffekt. Man bør etterstrebe maksimal blinding. I den anledning kan det nevnes at trippel-blinding som inkluderer at statistikeren og forfatterne av studie-rapporten også har vært foreslått.
5. **Utvelgelse av forsøkspersoner** Det skal detaljert og utvetydig gjøres rede for hvem som kan rekrutteres til studien: Om den er begrenset til visse diagnoser eller under-diagnoser, om det er alders eller kjønnsrestriksjoner. I primær-preventive studier vil man ofte kreve at pasienten skal være frisk og fri for symptomer og tilhøre en viss aldersgruppe. Ofte innebærer behandlingens natur at personer med visse tilstander eller annen samtidig behandling må ekskluderes. Det er selvsagt at i utprøving av medikamenter kan ingen som oppfyller kontraindikasjoner vurderes for inklusjon i studien.
6. **Randomiseringsprosedyre** Alle detaljer i forbindelse med randomiseringen av pasientene skal beskrives. Etter at pasienten er inkludert i studien og hun eller han har gitt sitt skriftlige samtykke skal vedkommende randomiseres. Enten dette skjer ved at man tar en telefon til et sentralt randomiserings-senter, eller man åpner en konvolutt som var laget ferdig på forhånd skal dette beskrives i protokollen. Hovedpoenget er å godtgjøre at tildelingen av behandling er tilfeldig og uforutsigbar og skjer etter at pasienten er inkludert i studien. Det skal fremgå hvordan disse konvoluttene ble produsert eller hvordan den sentrale randomiserings-enheten fungerer. Det skal fremgå om randomiseringen er blokket og/eller stratifisert eller om andre randomiseringsstrategier benyttes. Det er viktig at alle involverte i størst mulig grad er blindet for randomiseringsresultatet.
7. **Behandlingsregimer** Behandlingsalternativene beskrives i detalj. For medikamentell behandling gis doser, administrasjonsmåte, tidspunkter for inntak og behandlingsperiode. Både fabrikantnavn og generisk navn på medikamentet må gis. For ikke-medikamentell behandling gis detaljert

protokoll/prosedyre-beskrivelse både for behandlingsgruppe og kontrollgruppe.

8. **Måling av behandlingseffekt** En av de viktigste avgjørelsene som tas på protokollstadiet er hvordan behandlingens effekt skal måles. Man må bestemme seg for en hovedrespons som skal rapporteres som studiens hovedresultat. Vanligvis registreres også sekundære endepunkter.
9. **Overvåkning** (monitorering) Alle deltakere i kliniske forsøk skal kontrolleres underveis. Alle bivirkninger og pasientenes generelle velvære og helsestilstand skal registreres. Biokjemiske og andre parametre på sykdommens forløp skal registreres. Om det oppstår komplikasjoner som kan være relatert til behandlingen må koden brytes om studien er blindet. Også om det på et tidlig tidspunkt blir klar at den ene behandlingen er klart bedre enn den andre må studien brytes (dette gjelder enten resultatene går i kontrollgruppens eller den nye behandlingens favør).
10. **Forsøksstørrelse/statistikk** Det skal gjøres rede for hvordan man har beregnet det nødvendige antall forsøkspersoner som må rekrutteres til forsøket for at man skal kunne forvente å påvise en eventuell behandlingseffekt av en viss størrelse. Forsøkets styrke, statistisk test og signifikansnivå, om testen er en-sidig eller to-sidig, og den minste behandlingseffekt man ønsker å kunne påvise skal angis. For detaljer vises til eget avsnitt om forsøksstørrelse. Videre skal de statistiske metoder som vil bli benyttet ved endelig analyse av studien beskrives. For langvarige studier skal det utføres interimanalyser, som regel av en egen monitoreringskomité som har kjennskap til behandlingkoden og som kun har denne oppgaven. Denne komitéen kan avbryte studien om den på et tidlig tidspunkt finner at behandlingene gir forskjellig resultat, enten ved at den ene behandlingen er klart bedre enn den andre, eller ved at bivirkninger eller andre komplikasjoner forekommer med uventet hyppighet på den ene eller begge behandlingene.
11. **Kvalitetssikring av studien** Det skal gjøres rede for hvordan man vil sikre at studien gjennomføres som planlagt. Hvordan man vil kontrollere kvaliteten på skjemaer som fylles ut, hvordan man vil sikre at helsepersonellet følger protokollens anvisninger i behandlingen av pasientene. I mange studier er det opp til pasienten selv å følge studiens anvisninger. I hvor stor grad så skjer eller ikke skjer kan i mange tilfelle overvåkes, f.eks. ved måling av evt. medikament (eller metabolitter) i blodet.
12. **Håndtering av persondata** Det må bringes på det rene om Datatilsynet må søkes om tillatelse i forbindelse med oppretting av database for studien. Videre bør det godtgjøres personopplysninger kun er tilgjengelige i den grad gjennomføringen av studien krever det.

13. **Administrasjon av studien** Ansvar for gjennomføring av alle deler av studien skal fordeles til navngitte personer; fra søknader til etisk komité, evaluering av pasienter for deltagelse, målinger og intervjuer underveis, etc.
14. **Finansiering, rapportering, forsikring, etc.** Her skal budsjett med fordeling av utgifter på forskjellige parter fremgå. Det skal vurderes å tegne forsikring for deltakerne. Planer for fremtidig publisering og rapportskrivning skal nedtegnes; inkludert forfatterrekkefølge eller eventuelt gruppeforfatterskap. Det må fremgå om den eller de studieansvarlige har økonomisk interesse forbundet med utfallet av studien. Økonomiske forbindelser mellom studieansvarlige og eventuelle sponsorer bør beskrives.
15. **Sammendrag, vedlegg, litteraturliste** Studien skal beskrives kort i en sammendrag. Alle registrerings-skjemaer, informasjon til deltagerne, samtykke-formular skal vedlegges komplett. Liste over relevant litteratur inngår i protokollen.

Beregning av forsøksstørrelse

Vi skal her gjennomgå beregning av forsøksstørrelse for situasjonen hvor endepunktet kan uttrykkes som en proporsjon (prosent). Det kan være at vi vil sammenligne dødelighet i behandlingsgruppen mot dødelighet i kontrollgruppen, eller prosentene som blir friske, etter et gitt kriterium, i de to behandlingsgruppene.

Formelen gir nødvendig antall pasienter i hver av to behandlingsgrupper når den endelige analysen utføres med en to-sidig Pearson's χ^2 test med signifikans-nivå, α , og styrke $1-\beta$. Husk at både styrke og signifikansnivå er størrelser som er knyttet til den statistiske testens egenskaper. I beregning av forsøksstørrelse er det vanlig å benytte en styrke på 80 (eller 90) % og et to-sidig signifikansnivå på 0.05. Dette betyr at om det ikke er noen forskjell mellom behandlingene er sannsynligheten for å få et falskt signifikant funn lik 5 % (signifikansnivået). Om det foreligger en behandlings-effekt av spesifisert størrelse er sannsynligheten for at man skal oppnå et (korrekt) signifikant funn 80 % (styrken). Det finnes mange varianter av denne beregningen som alle gir nøyaktig samme svar. En vanlig brukt formel er (81),

$$n_1 = n_2 = \frac{\left[z_{1-\alpha/2} \sqrt{2p(1-p)} + z_{1-\beta} \sqrt{p_1(1-p_1) + p_2(1-p_2)} \right]^2}{(p_2 - p_1)^2}$$

hvor n_1 og n_2 er antallet pasienter i hver av de to behandlingsgruppene, p_1 og p_2 er forventete respons-proporsjoner, $p=(p_1+p_2)/2$, α er signifikansnivået, $1-\beta$ er styrken, og z_p er z-verdien for en standard normal-fordeling som har et areal p til venstre.

La oss se på et eksempel. Hvor mange pasienter må innlemmes i en studie som vil sammenligne to trombolytiske behandlinger av akutt hjerteinfarkt. Man regner at med standard behandling er dødeligheten de første 30 dager 7 % og at denne kan forbedres med et prosentpoeng til 6 %. Om vi bruker formelen og krever en styrke på 80 % når testen utføres to-sidig med et signifikansnivå på 5 % finner vi at det kreves 9,542 pasienter i hver behandlingsgruppe, altså et forsøk i størrelsesorden 20,000 pasienter. Eksempelet her er inspirert av GUSTO studien som sammenlignet fire trombolytiske strategier på i alt 41,021 pasienter i 1,081 sykehus i 15 land (98). Tilsynelatende små endringer i antatt behandlingseffekt kan gi store utslag i beregnet forsøks-størrelse. Om vi for eksempel hadde regnet med at den aktive behandlingen hadde redusert dødeligheten fra 12 til 6 % ville 356 pasienter i hver behandlings-gruppe vært tilstrekkelig.

Styrke	$z_{1-\beta}$	Signifikansnivå	$z_{1-\alpha/2}$
0.50	0	0.10	1.645
0.80	0.842	0.05	1.960
0.90	1.282	0.01	2.576
0.95	1.645	0.001	3.291
0.99	2.326	0.0001	3.891

Feilkilder ved randomiserte forsøk

Det randomiserte kliniske forsøk er uten sammenligning den mest pålitelige studietypen innen klinisk forskning. Om alle involverte er innstilt på å respektere de data som måtte fremkomme og studien gjennomføres forskriftmessig vil vi få så sikker viten som det er mulig å fremskaffe. Randomiseringen sørger for at behandlingsgruppene blir like i utgangspunktet og den viktigste feilkilden er at tilfeldigheter skulle gi overvekt av pasienter med dårlig prognose i en av behandlingsgruppene. For å dokumentere at så ikke er tilfelle, eller i hvilken grad det eventuelt har skjedd, innledes rapporteringen av et klinisk forsøk alltid med fordelingen av prognostiske variabler etter randomiseringsgruppe. Vi har tidligere sett at mangler ved selve randomiserings-prosedyren kan gi alvorlig prognostisk skjevfordeling mellom behandlingsgruppene(22,55,86).

Valg av endepunkt er kritisk. Endepunktet må ha klinisk og praktisk relevans. Om surrogatendepunkter benyttes må man ha sterke holdepunkter for at dette er nært relatert til den effekten man egentlig vil måle.

Manglende blinding i studier hvor enten pasientens eller undersøkers skjønsmessige vurderinger inngår i endepunktet kan føre til at resultatene

dreies i den ene eller andre behandlingens favør. Det er viktig at randomiseringen er blind i åpne studier.

Ofte finansieres randomiserte kliniske utprøvinger fra kilder som har interesse i utfallet av studien. Rochon og medarbeidere (80) tok for seg studier vedrørende medikamentell behandling av visse rheumatiske lidelser og viste at resultatene av disse nær sagt aldri gikk i finansieringskildens disfavør. Slike forhold spiller sannsynligvis en større rolle ved små studier enn ved store studier. Videre ble det vist at kontrollgruppen i en del tilfeller fikk suboptimal behandling med den følge at effekten av den nye eller alternative behandlingen overdrives (80).

Mange kliniske studier gjennomføres med et for lite antall forsøkspersoner til at man kan forvente å påvise den forventede behandlingseffekt med statistisk signifikans. En følge av dette kan være at man ikke kan ta i bruk nye adekvate behandlinger siden det ikke med sikkerhet er påvist at de er effektive. En annen uheldig konsekvens av små studier er at de i visse tilfeller kan overdrive behandlingseffekten. Dette har sammenheng med publikasjonsfordreining (engelsk: publication bias) og vi vil forklare det nærmere.

Det er vist at blant publiserte studier på samme problemstilling er det en tendens til at små studier viser en større behandlingseffekt enn store studier. Dette kan forklares ved at en liten studie som ikke finner noen statistisk signifikant effekt ofte vil bli avvist i vurderingsprosessen på grunn av manglende styrke. Motsatt vil en like liten studie som viser en signifikant effekt ikke møte den samme innvendingen. Man tenker at her er jo studieresultatet et bevis på at styrken var tilstrekkelig. Hva man ofte ikke tenker er at signifikansen kan ha to komponenter, en moderat behandlingseffekt bidrar til signifikans men må hjelpes av tilfeldigheter for å nå den formelle signifikansgrensen. Dette har ingen konsekvenser om man kun ser på resultatet av hypotesetesten. Et signifikant resultat som er fremkommet på denne måten vil derimot være ledsaget av et overdrevet estimat av behandlingseffekten.

Rapportering av randomiserte forsøk

Resultater fra ethvert randomisert forsøk skal rapporteres uavhengig av studiens utfall. Ofte mangler slike rapporter kritisk informasjon, for eksempel om randomiseringsprosedyren. Det er utarbeidet sjekklistene og forslag til strukturert rapportering av randomiserte forsøk (102). The Journal of the American Medical Association publiserte nylig et eksempel på strukturert rapportering (109). Viktigheten av grundig rapportering av ikke bare resultatene, men også design og gjennomføring av studien har fått fornyet oppmerksomhet gjennom CONSORT retningslinjene (4,12,85). Bruk av subgruppeanalyser er også nylig drøftet(10).

Kapittel 3: Meta-analyser

Ofte er samme medisinske problemstilling undersøkt i et stort antall studier. Det kan dreie seg om både om observasjonsstudier og randomiserte studier. Om man kvantitativt summerer den samlede erfaring fra disse studiene kalles sammenfatningen gjerne en meta-analyse. Slike kvantitative oversikts-analyser var relativt sjeldne før 1980, mens det siden 1985 er publisert minst 50-100 meta-analyser årlig (23). Meta-analyser er forbundet med alvorlige feilkilder som den enkelte studien er fri for. For eksempel vil selektiv publisering av studier med gunstig resultat i favør av en ny behandling føre til at sammenfatningsmålet blir fordreid i den nye behandlingens favør. Dette problemet kalles publikasjons-fordreining (engelsk: publication bias). Særlig er faren for dette stor dersom det blir gjort et stort antall mindre forsøk. Problemer oppstår også på grunn av at en serie studier aldri er identiske. Man finner forskjeller i pasientpopulasjonen som er undersøkt, både behandlingsregimer og den totale terapisisituasjonen endrer seg over tid slik at effekten av en bestemt behandling kan avta eller nye bivirkninger kan oppstå. Til tross for problemer med meta-analysen er den allerede veletablert innen klinisk medisin og en det finnes en stor litteratur på metodologiske aspekter ved meta-analysen(15,45,56,68,72,84,93,103,104) (21).

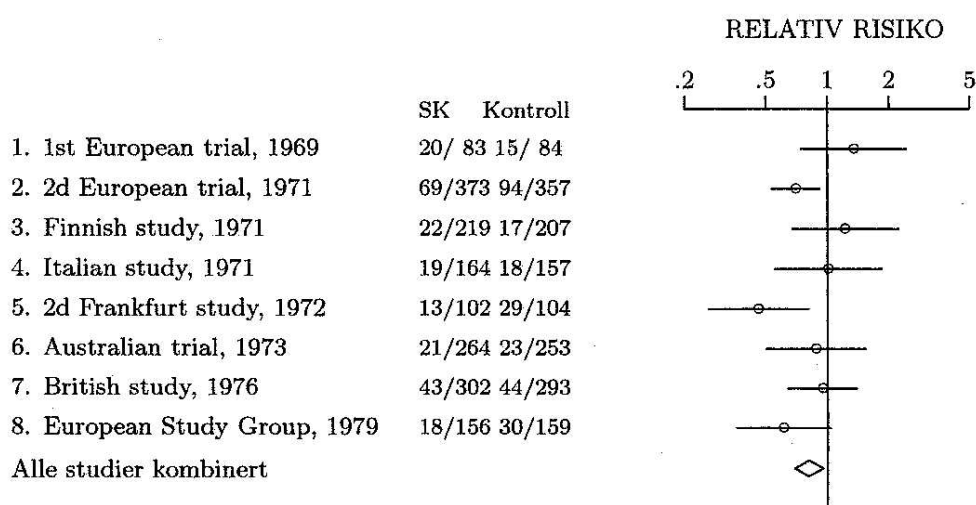
Eksempler på meta-analyser

Meta-analyser er gjennomført innen de forskjelligste medisinske disipliner fra behandling av brystkreft til kroniske otitter (33,110). De kanskje mest innflytelsesrike meta-analysene har vært innen behandling av kardiovaskulære sykdommer (8,40,89,114). Eksempler fra epidemiologien er studier av langtidsvirkninger av østrogen-terapi og alkohol som en mulig årsaksfaktor til brystkreft. Vi skal i nå illustrere meta-analysen med tre eksempler.

Intravenøs streptokinase mot akutt hjerteinfarkt. Stampfer og medarbeidere gjennomgikk i 1982 (89) alle publiserte randomiserte studier som undersøkte effekten av intravenøs streptokinase på dødelighet opptil de første 40 dager etter et akutt hjerteinfarkt. Det var dengang gjennomført 8 slike studier fra 1969 til 1979. Antall dødsfall og total størrelse i henholdsvis streptokinase- og placebo-gruppene er vist i figuren. Figuren viser også grafisk punkttestimatet for behandlingseffekten (åpen sirkel) med tilhørende 95% konfidensintervall. Vi ser at om alle studiene kombineres er den relative risikoen 0.78 i streptokinases favør. Dette betyr at streptokinase reduserer dødeligheten med 22 %. Vi ser at behandlingseffekten varierer ganske mye fra studie til studie. En test for heterogenitet i behandlingseffekt støtter denne oppfatningen ($P=0.095$). Ved slik heterogenitet bør man prøve å finne forklaringer. De to studiene som viser en ikke ubetydelig uheldig effekt av streptokinase skiller seg fra de øvrige studiene ved at de inkluderte pasienter inntil 72 timer etter symptomdebut. De øv-

rige studiene ga streptokinase innen de første 12 eller 24 timer. I de to studiene med klarest gunstig effekt av streptokinase fikk kontrollgruppen placebo, mens i de øvrige studiene med ett unntak ble kontrollgruppen antikoagulert. For denne problemstillingen kan vi føle oss noenlunde trygge på at seleksjonsbias ikke fordreier resultatene i alvorlig grad. En større randomisert studie av streptokinase i behandling av akutt hjerteinfarkt vil bli publisert uansett resultat.

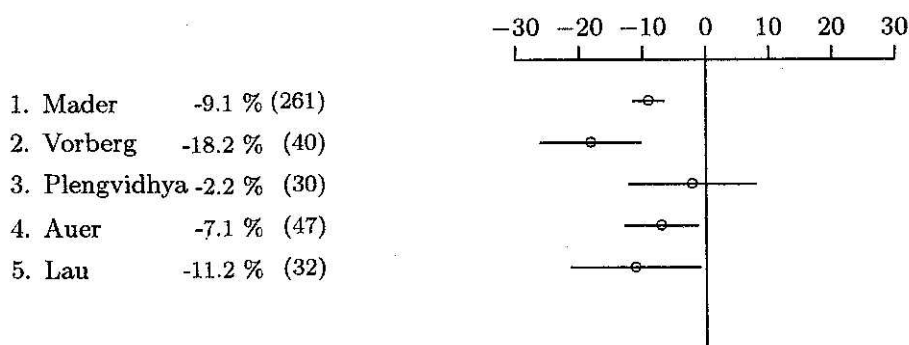
Videre antydet resultatene at streptokinase har ugunstig effekt om det gis senere enn 24 timer etter symptom-debut samt at antikoagulasjon også reduserer dødeligheten siden studier med antikoagulert kontrollgruppe viser mindre behandlingseffekt enn placebo-kontrollerte studier.



Kolesterolsenkende effekt av hvitløk. Som et eksempel på en meta-analyse hvor seleksjons-fordreining kan tenkes å spille en større rolle vil vi se på en meta-analyse av randomiserte studier av hvitløkspreparater på serum kolesterol (108). Fra i utgangspunktet 28 kandidat-studier viser figuren prosentvis forskjell i endring (fra før til etter behandling) mellom hvitløk- og placebo-gruppen i 5 randomiserte studier blant pasienter med hyperkolesterolemi. I alle disse studiene gikk effekten i favør av hvitløk. Analysen antyder at man kan forvente en reduksjon i kolesterol på ca 10 % med hvitløksbehandling.

I motsetning til meta-analysen på streptokinase i behandling av akutt infarkt kan vi her lett forestille oss at publikasjonsfordreining kan spille inn. Det kan være vanskelig å få publisert en studie som viser at hvitløk ikke har noen effekt på kolesterol. årsaken er at mange vil stille seg likegyldig til problemstillingen og vil innvende at man ikke kan publisere hver eneste studie som viser fravær av sammenheng mellom hvitløk eller andre kostfaktorer og biokjemiske parametre. Videre kan vi lett tenke oss at studier som ikke finner noen

effekt blir lagt i arkiv-skuffen allerede av forskerne selv. Derfor må konklusjonen bli at denne meta-analysen ikke slår fast at hvitløk senker kolesterol. Mange små negative studier kan være gjennomført uten å nå frem til publisering. Størrelsen av en eventuell publikasjonsfordreining er det ikke mulig å beregne. Problemstillingen bør derfor belyses i en større ny studie med flere hundre pasienter i hver behandlingsgruppe



Nitrater i behandling av akutt infarkt. I en oversikt over 10 randomiserte forsøk, publisert i perioden 1979-1985, som studerte effekten av intravenøse nitrater på dødelighet etter akutt infarkt ble det påvist en klar samlet gunstig effekt av behandlingen. Blant 1020 kontroll-pasienter døde 18.9 % mens fataliteten var 13.3 % blant 1021 pasienter som fikk enten nitroglyserin eller nitroprussid. Dette antydte at en tredjedel av dødsfallene kunne forhindres ved bruk av nitrater (113). Senere er det gjennomført en stor randomisert studie (ESPRIM) hvor 2010 infarkt-pasienter uten tegn på hjertesvikt fikk placebo og 2007 pasienter fikk molsidomin, som er nært beslektet i effekt med tidligere brukte nitrater (39). I denne store studien var dødeligheten de første 35 døgn etter infarkt praktisk talt identisk i behandlings- og kontroll-gruppen med 168 (8.4%) og 176 (8.8%) dødsfall. årsakene til den store diskrepansen mellom meta-analysen og denne studien kan være flere: medikamentet var ikke det samme, i ESPRIM-studien var hjertesviktpasienter ikke inkludert, i ESPRIM-studien ble nitrater gitt til ca. en tredjedel av pasientene før randomisering, dvs. også til kontrollgruppen, den totale terapi-situasjonen var betydelig endret fra tidlig 1980-tall til tidlig 1990-tall med utstrakt bruk av aspirin, heparin, trombolyse, nitrater til alle pasienter før randomisering. Publikasjonsfordreining er en mindre sannsynlig forklaring da det er lite trolig at det var gjennomført så mange upubliserte forsøk som systematisk avvek fra studiene som ble inkludert i meta-analysen at samleresultatet skulle bli påvirket i nevneverdig grad.

Protokoll for meta-analyse

Alle meta-analyser bør ta utgangspunkt i en protokoll for analysen. Vi viser nå et forslag til slik protokoll. Vi vil så drøfte hvert av punktene i tabellen.

1. Definer problemstilling:

- Behandling(er): doser, administrasjon, substans, prosedyrer
- Kontrollgruppe: placebo og/eller aktiv behandling
- Mål for behandlingseffekt
- Indikasjonstilling, pasient-populasjon, diagnoser, stadier

2. Søking etter kandidat-studier

- Avgjør om randomiserte og/eller ikke-randomiserte og/eller upubliserte studier skal tas med
- MEDLINE, referanselister, ekspertråd

3. Seleksjon

- Kandidatstudiene gjennomgås for å avgjøre inklusjon.
- Gjennomgangen bør være blindet for resultater og opprinnelse og helst gjøres uavhengig av to eksperter.
- Studiene gis kvalitetspoeng etter bestemte retningslinjer.
- I den endelige rapporten skal det redegjøres for seleksjonsprosessen med antall kandidatstudier og antall og grunner for eksklusjon.

4. Statistisk analyse

5. Drøfting av feilkilder

Problemstilling

Først må man definere problemstillingen i stor detalj. Hvilken behandling eller behandlinger ønsker man å studere. I fall det dreier seg om medikamenter må man ta stilling til om man skal ha snevre eller vide kriterier for behandlingen som skal studeres. Er man snever vil man kun studere en enkelt substans i en bestemt dose administrert på en gitt måte (f. eks. per os eller intravenøst) på et veldefinert stadium i sykdomsforløpet. Med vide kriterier vil man inkludere en serie beslektede substanser, tillate variasjoner i doser og administrasjonsmåte og være liberal med hensyn til når i sykdomsforløpet medikamentet ble gitt. Tilsvarende betraktninger gjør seg gjeldende for ikke-medikamentelle behandlinger.

Også pasient-målgruppen for behandlingen må defineres. Her må det tas stilling til i hvor stor grad man ønsker like inklusjons- og eksklusjonskriterier i studiene som skal sammenfattes. På samme måte som man definerer pasientpopulasjonen i en enkelt randomisert studie må man definere meta-analysens pasient-populasjon. For hvilke diagnoser, hvilken alvorlighetsgrad av sykdommen vil man sammenfatte kunnskap om behandlingen? I hvilken grad vil man godta eventuell annen samtidig behandling? Som for behandlingen kan man her ha vide eller snevre kriterier. Disse avgjørelsene får konsekvenser for hvor mange studier som analysen vil omfatte og for konklusjonene man kan trekke.

Søking etter kandidat-studier

Den uten tvil mest kritiske fasen av en meta-analyse er datainnsamlingen, dvs prosessen som finner frem kandidatstudier og som avgjør hvilke av kandidatstudiene som skal inngå i den endelige meta-analysen. Enkelte vil hevde at meta-analyser bare kan gjøres på store publiserte randomiserte studier. Årsaken er at for store studier (med stort antall pasienter) er publikasjonsfordreining et mindre problem enn for små studier. Videre regner man ikke-randomiserte studier for så upålitelige at man ikke vil risikere å basere terapeutiske anbefalinger på slike studier. Andre vil hevde at meta-analysen nettopp er nyttig når det gjelder å summere mange små studier, både observasjonsstudier og randomiserte forsøk. I fall man gjør en meta-analyse på observasjonsstudier må man være klar over slike studiers upålitelighet. Publikasjonsfordreining spiller også sterkere inn for observasjonsstudier og små studier av alle typer. Noen vil hevde at det, for å unngå publikasjonsfordreining, er svært viktig å inkludere studier som ikke er publisert. Om en slik innsats er anstrengelsen verdt vil avhenge av feltet man arbeider i og i hvilken grad resultater fra upubliserte studier i det hele tatt er tilgjengelige.

Når problemstillingen er veldefinert og man har tatt stilling til hvilke studietyper man vil inkludere søker man etter kandidat-studier. Dette gjøres idag enkelt gjennom MEDLINE (61). Man skal imidlertid være klar over den elektroniske litteratursøkingens begrensninger. Ved MEDLINE søk mister man ofte opptil en tredjedel av relevante studier. Derfor er det viktig å nøye gjennomse referanselister, og konsultere eksperter i feltet. Innen visse disipliner er det opprettet databaser over kliniske forsøk.

Endelig seleksjon

Den endelige seleksjonen av studier for meta-analysen bør helst foregå blindet for resultatet i hver enkelt studie. Selv for en meta-analytiker som legger vekt på å opptre uhildet kan egne preferanser for terapivalg farge inklusjon eller eksklusjon av studier. Studiene bør også kvalitetsvurderes slik at studier som åpenbart ikke holder kvalitetsmessige mål kan holdes utenfor. Et alternativ er å supplere samleresultatene med separate resultater for *gode* og *dårlige* studier.

Statistiske metoder og feilkilder

Valg av statistisk metode for beregning av samlet behandlingseffekt er mindre kritisk enn utvelgesprosessen beskrevet ovenfor. Slike metoder og en drøfting av feilkilder blir beskrevet i de to siste delene av dette kapitlet.

Statistiske metoder for meta-analyse

Den statistiske analysen av en meta-analyse har flere komponenter. Man vil estimere en gjennomsnittlig behandlingseffekt og angi usikkerheten rundt den-

ne med et konfidensintervall; man supplerer også dette estimatet med en test for behandlingseffekt. Ofte er det stor variasjon i behandlings-effekt mellom forskjellige studier og det kan være ønskelig å utføre en formell statistisk test på om dataene er forenlig med forutsetningen om at variasjonen i behandlings-effekt kun skyldes tilfeldig avvik fra en felles-effekt. Om variasjonen er større enn at den kan tilskrives tilfeldigheter må man søke årsaker til variasjonen og vurdere om det er riktig å estimere en felles-effekt. En test for dette formålet kalles gjerne en test for homogenitet eller heterogenitet, eller en test for interaksjon.

Det finnes to hyppig brukte metoder for summering av resultater fra en serie studier. Det ligger to ulike filosofier bak, men resultatene ofte ikke svært forskjellig. Den ene skolen (fast effekt modellen) benytter i prinsippet Mantel-Haenszel metoden som ble gjennomgått tidligere. Her forutsetter man de inkluderte studier utgjør alle studier av interesse og metoden har som mål å estimere en felles behandlingseffekt. Den andre skolen (tilfeldig effekt modellen) tenker seg at studiene vi har samlet er et utvalg fra et større univers av studier. I dette hypotetiske universet av studier er varierer behandlingseffekten. I tillegg til at man estimerer en gjennomsnittlig behandlingseffekt søker man å estimere hvor stor variasjonen er (27,41).

Uansett skole kan meta-analysen gjøres på forskjellige effektmål som odds ratio, risiko ratio, rate ratio eller forskjellsmål som risiko-forskjell eller rate-forskjell. Vi gir her kun metoden for odds ratio som er den samme som brukes innen analyse av epidemiologiske data.

Fast effekt: Yusuf-Peto metoden

I fast effekt meta-analyse kan man bruke nøyaktig de samme metoder som i analyse av stratifiserte epidemiologiske data. Mantel-Haenszel odds ratio estimatet benyttes til å estimere felles-effekten og Mantel-Haenszel testen for å beregne P verdien. En adekvat meta-analyse krever et konfidensintervall rundt estimatet for den samlede effekt. Til dette formål anbefaler vi Robins-Breslow-Greenland metoden (79,82). Her vil vi imidlertid gjennomgå en variant av Mantel-Haenszel metoden som benyttes hyppig i presentasjonen av meta-analyser.

		Behandling		
		Aktiv	Kontroll	
Vellykket	a_i	b_i	M_{1i}	
Mislykket	c_i	d_i		M_{0i}
	N_{1i}	N_{0i}	T_i	

Denne metoden kalles Yusuf-Peto metoden og er beskrevet av Yusuf og medarbeidere (114). Fordelen med denne metoden er at den beregningsmessig er meget enkel og også lett gir en test for heterogenitet i effekt blant studiene. Disse beregningene kan lett gjøres med en lommekalkulator. Ulempen med

metoden er at den er noe mer unøyaktig enn de mer kompliserte alternativene og derfor blir frarådet av enkelte.

Yusuf-Peto metoden baserer seg på forskjellen mellom den observerte og forventede verdien i hver enkel a -celle i 2×2 tabellen. Notasjonen for studie i er som vist i Tabell x. Ved hjelp av størrelsene som inngår i Mantel-Haenszel testen kan vi nå lett finne både et estimat for den felles odds ratio, konfidensintervall for denne størrelsen og en test for heterogenitet i odds ratio blant de forskjellige studiene.

Den forventede verdien i hver a_i -celle, e_i , beregnes fra kantene i 2×2 tabellen slik,

$$e_i = \frac{N_i M_i}{T_i}$$

Vi betegner forskjellen mellom observert og forventet a -celle verdi, $s_i = a_i - e_i$ og summen av forskjellen mellom observert og forventet over de forskjellige studiene er $s = \sum s_i$. Variansen til a -cellen er

$$c_i = \frac{N_{1i} N_{0i} M_{1i} M_{0i}}{T_i^2 (T_i - 1)}$$

og summen over studiene betegnes $c = \sum c_i$. Mantel-Haenszel test-størrelsen blir da,

$$X^2 = s^2 / c$$

Denne størrelsen er χ^2 -fordelt med en frihetsgrad og P verdien finnes i χ^2 -tabellen. Metodens estimat av den felles odds ratio er,

$$OR = \exp(s/c)$$

Vi benytter det naturlige logaritmesystemet slik at formelen betyr at grunntallet i det naturlige logaritmesystemet skal opphøyes i s/c . Et 95 % konfidensintervall beregnes fra følgende formel,

$$\exp(s/c \pm 1.96/\sqrt{c}).$$

Test-størrelsen for heterogenitetstesten er

$$X_{N-1}^2 = \sum (s_i^2 / c_i) - s^2 / c$$

Den er χ^2 -fordelt med $N-1$ frihetsgrader, hvor N er antall studier som inngår i analysen.

De samme formlene gir estimater for hver enkelt studie.

Regneeksempel: Yusuf-Peto metoden

Vi skal bruke Yusuf-Peto-metoden på dataene fra de 8 studiene som er vist i figuren. De nødvendige mellomregningene er vist i følgende tabell.

A_i	e_i	s_i	c_i	s_i^2/c_i	$\exp(s/c)$	95%KI	2P
20	17.395	2.605	6.958	0.975	1.45	0.69-3.06	0.32
69	83.286	-14.286	31.679	6.442	0.64	0.45-0.90	0.01
22	20.049	1.951	8.871	0.429	1.25	0.65-2.41	0.51
19	18.903	0.097	8.205	0.001	1.01	0.51-2.01	0.97
13	20.796	-7.796	8.399	7.236	0.40	0.20-0.78	0.01
21	22.468	-1.468	10.079	0.214	0.86	0.47-1.60	0.64
43	44.158	-1.158	18.597	0.072	0.94	0.60-1.48	0.79
18	23.771	-5.771	10.203	3.264	0.57	0.31-1.05	0.07
225	250.826	-25.826	102.991	18.634	0.78	0.64-0.94	0.01

De åtte første linjene i tabellen er beregnet fra hver enkelt studie isolert. Vi ser at i første kolonne står a -cellen fra hver enkelt studie. Videre står den forventede verdien til a -cellen i kolonne to og i kolonne fire a -cellens varians. I linje ni er disse størrelsene summert. Estimert og konfidensintervall for den felles odds ratio samt Mantel-Haenszel testen beregnes fra disse summerte størrelsene. $OR = \exp(-25.826/102.991) = 0.778$ og 95% konfidensintervall er $\exp(0.778 \pm 1.96\sqrt{\{102.991\}}) = (0.642, 0.944)$ mens Mantel-Haenszel teststørrelsen er $(-25.826)^2/102.991$ som gir P -verdien 0.011 ved å slå opp i χ^2 -tabell på en frihetsgrad.

For å beregne teststørrelsen for heterogenitetstesten må man i tillegg beregne s_i^2/c_i for hver enkelt studie og summere. Teststørrelsen blir

$$18.634 - (-25.826)^2/102.991 = 12.158$$

som gir $P = 0.095$ (χ^2 -fordelt med 7 frihetsgrader).

Vi merker oss at vi kan beregne odds ratio, 95 % konfidensintervall og en test for assosiasjon ved samme formler anvendt på hver enkelt studie. Disse beregningene fremgår av tabellen.

Feilkilder ved meta-analyser

Uforklarlige uoverenstemmelser mellom meta-analyser og store enkeltforsøk

Ved flere anledninger har man opplevd at tilsynelatende meget godt gjennomførte meta-analyser har funnet en klar behandlingseffekt som senere er tilbakevist i store definitive og godt gjennomførte enkelstudier. Ett eksempel er bruk av magnesium i akuttbehandling av hjerteinfarkt (50,92). Her velger kanskje de fleste å tro på den største studien (ISIS-4), men sannheten ligger kanskje ikke alltid i store tall (9), for eksempel var tidspunktet for magnesiumadministrasjon forskjellig i disse studiene (57). Tilsvarende kontroverser har man sett ved bruk av nitrater i infarktbehandling (35).

Seleksjonsfordreining

Feilkilden ved meta-analyser som har fått mest oppmerksomhet er publikasjonsfordreining (eng: publication bias). En hovedforutsetning for at en meta- eller oversikts-analyse skal gi en gyldig oppsummering av akkumulert viten er at studiene som analyseres er representative for alle gjennomførte studier. Om publiserte studier avviker systematisk fra upubliserte studier vil samleeffekten basert på publiserte studier være misvisende. Publikasjonsfordreining kan oppstå på flere nivåer. Det kan være at studien avbrytes og aldri fullføres, den kan fullføres uten at man søker å publisere den, forsøk på publikasjon kan mislykkes og sist artikkelen kan være publisert i en slik form at den ikke finnes ved standard søkeprosedyrer. Om noen eller alle disse fenomenene er relatert til resultatet av hver enkelt studie inntreffer publikasjons-fordreining.

Selvsensur – arkivskuffproblemet. Det er dokumentert at forskere som finner "ikke-signifikante" resultatet pålegger seg selv selvsensur. Blant 200 randomiserte kliniske forsøk som ble fulgt fra planleggingsstadiet var det 9 ganger større sjanse for at en "signifikant" sammenlignet med en *ikke-signifikant* studie skulle nå frem til publisering. Ikke i noe tilfelle ble de ikke-signifikante studiene stoppet av tidsskrifter. årsaken til at de ikke ble publisert lå hos initiativtakerne selv(28).

Tidsskriftsensur. Vi tenker oss en situasjon hvor et nytt medikament er lansert i behandling av en gitt sykdom. Videre forutsetter vi at medikamentet har en moderat gunstig effekt. Om det gjennomføres 100 små studier av effekten av et medikament på forløpet av sykdommen vil resultatene svinge fra ingen effekt til en sterk effekt. Vi tenker oss videre at alle studiene skrives sammen til artikler og sendes til vurdering for publisering i vitenskapelige tidsskrifter. La oss så operere med at 50 % av artiklene aksepteres for publikasjon. Noen av artiklene vil publiseres i anerkjente tidsskrifter med høyt opplag og stor innflytelse mens de fleste vil publiseres i tidsskrifter med begrenset leserkrets og kanskje på et språk som ytterligere begrenser spredningen. Om denne vurderingsprosessen påvirkes av resultatet av hver enkelt studie oppstår det publikasjons-fordreining. Da vil for eksempel positive studier bli over-representert i publiserte litteraturen og kanskje endog i velrenommerte tidsskrifter. På denne måten vil litteraturoversikter kunne overdrive behandlingseffekten.

Bortgjemte studier. Studier som er publisert på andre språk enn engelsk vil få mindre oppmerksomhet og vil være vanskeligere å finne enn engelskspråklige studier. Studier som publiseres i tidsskrifter som ikke indekseres i MEDLINE vil man ikke finne ved standard søkeprosedyrer. I tillegg kommer at klassifisering i MEDLINE ikke alltid skjer på en slik måte at det fremgår hvilken studietype, hvilken sykdom og hvilke behandlings-regimer som er studert. Kvaliteten på MEDLINE klassifikasjonen bedres stadig, men samtidig som det gjøres endringer vanskeligjør dette søk over lengre tidsrom.

Periode-effekter

Om studiene som inngår i en metaanalyse spenner over et langt tidsrom kan endringer i den totale terapisisituasjonen forvanske tolkningen av resultatene. Det kan da være nyttig å gjøre separate meta-analyser i forskjellige tidsepoker, eller fremstille behandlingseffekten over tid.

Liberale behandlings- og pasient-kriterier

Om man liberalt inkluderer beslektede behandlinger og forskjellige pasientgrupper i meta-analysen blir resultatene vanskeligere å benytte i klinisk sammenheng når de skal appliseres på den enkelte pasient.

Statistisk metode

Valg av statistisk metode vil sjelden påvirke resultatene i betydelig grad. Man bør merke seg at om behandlingseffekten er relativt konstant vil fast effekt og tilfeldig effekt modellene gi nokså like resultater. Ved betydelig variasjon (heterogenitet) vil tilfeldig-effekt metoden gi videre konfidens-intervall for samle-effekten og den vil tillegge små studier større vekt en fast-effekt metoden (Mantel-Haenszel). Dette er en ulempe ved den førstnevnte metoden siden kvaliteten av studien kan være korrelert med størrelsen. DerSimonean-Laird metoden kan da vekke små lav-kvalitetsstudier for tungt.

Svindel

En mulig feilkilde i meta-analyser er dobbelpublisering slik at de samme pasientene teller med flere ganger (52,69). Randomiserte studier kan også, forhåpentligvis sjelden, være fabrikerte (51,60,77).

Litteraturliste

1. **A Medical Research Council Investigation.** Streptomycin treatment of pulmonary tuberculosis. **BMJ** 1948;2:769-782.
2. **Allen KB, Dowling RD, Fudge TL, Schoettle GP, Selinger SL, Gangahar DM, Angell WW, Petracek MR, Shaar CJ, O'Neill WW.** Comparison of transmyocardial revascularization with medical therapy in patients with refractory angina. **N Engl J Med** 1999;341:1029-36.
3. **Altman DG.** Practical statistics for medical research. London: Chapman and Hall, 1991.
4. **Altman DG.** Better reporting of randomised controlled trials: the CONSORT statement. **BMJ** 1996;313:570-1.
5. **Altman DG, Bland JM.** How to randomise. **BMJ** 1999;319:703-704.
6. **Altman DG, Bland JM.** Statistics notes. Treatment allocation in controlled trials: why randomise? **BMJ** 1999;318:1209.
7. **Annas GJ, Grodin MA.** Human rights and maternal-fetal HIV transmission prevention trials in Africa. **Am J Public Health** 1998;88:560-3.
8. **Antiplatelet Trialists' Collaboration.** Collaborative overview of randomised trials of antiplatelet therapy--I: Prevention of death, myocardial infarction, and stroke by prolonged antiplatelet therapy in various categories of patients. [published erratum appears in *BMJ* 1994 Jun 11;308(6943):1540]. **BMJ** 1994;308:81-106.
9. **Antman EM.** Randomized trials of magnesium in acute myocardial infarction: big numbers do not tell the whole story. **Am J Cardiol** 1995;75:391-3.
10. **Assmann SF, Pocock SJ, Enos LE, Kasten LE.** Subgroup analysis and other (mis)uses of baseline data in clinical trials [In Process Citation]. **Lancet** 2000;355:1064-9.
11. **Bayer R.** The debate over maternal-fetal HIV transmission prevention trials in Africa, Asia, and the Caribbean: racist exploitation or exploitation of racism? **Am J Public Health** 1998;88:567-70.
12. **Begg C, Cho M, Eastwood S, Horton R, Moher D, Olkin I, Pitkin R, Rennie D, Schulz KF, Simel D, Stroup DF.** Improving the quality of reporting of randomized controlled trials. The CONSORT statement. **JAMA** 1996;276:637-9.
13. **Bergsjø P.** Etikkomiteéene og det informert samtykke. **Tidsskr Nor Lægeforen** 1991;111:558-9.
14. **Bergsjø P.** Biomedisinske forskningsprosjekter hvor forsøk på mennesker inngår - hva er nå det? **Tidsskr Nor Lægeforen** 1993;113:1443-4.
15. **Boissel JP, Blanchard J, Panak E, Peyrieux JC, Sacks H.** Considerations for the meta-analysis of randomized clinical trials. Summary of a panel discussion. **Controlled Clin Trials** 1989;10:254-81.

16. **Brennan TA.** Proposed revisions to the Declaration of Helsinki--will they weaken the ethical principles underlying human research? **N Engl J Med** 1999;341:527-31.
17. **Brownlee.** **JASA** 1955.
18. **Burkhoff D, Schmidt S, Schulman SP, Myers J, Resar J, Becker LC, Weiss J, Jones JW.** Transmyocardial laser revascularisation compared with continued medical therapy for treatment of refractory angina pectoris: a prospective randomised trial. ATLANTIC Investigators. Angina Treatments-Lasers and Normal Therapies in Comparison. **Lancet** 1999;354:885-90.
19. **Burns SM, Sharples LD, Tait S, Caine N, Wallwork J, Schofield PM.** The transmyocardial laser revascularization international registry report. **Eur Heart J** 1999;20:31-7.
20. **Cargill V, Cohen D, Kroenke K, Neuhauser D.** Ongoing patient randomization: an innovation in medical care research. **Health Serv Res** 1986;21:663-78.
21. **Chalmers I, Altman DG.** Systematic reviews. London: BMJ Publishing Group, 1995.
22. **Chalmers TC, Celano P, Sacks HS, Smith H, Jr.** Bias in treatment assignment in controlled clinical trials. **N Engl J Med** 1983;309:1358-61.
23. **Chalmers TC, Lau J.** Meta-analytic stimulus for changes in clinical trials. **Stat Methods Med Res** 1993;2:161-72.
24. **Connor EM, Sperling RS, Gelber R, Kiselev P, Scott G, O'Sullivan MJ, VanDyke R, Bey M, Shearer W, Jacobson RL, et al.** Reduction of maternal-infant transmission of human immunodeficiency virus type 1 with zidovudine treatment. Pediatric AIDS Clinical Trials Group Protocol 076 Study Group. **N Engl J Med** 1994;331:1173-80.
25. **CPMP Working Party on Efficacy of Medicinal Products.** EEC note for guidance: good clinical practice for trials on medicinal products in the European Community. **Pharmacol Toxicol** 1990;67:361-72.
26. **Czeizel AE, Dudas I.** Prevention of the first occurrence of neural-tube defects by periconceptional vitamin supplementation [see comments]. **N Engl J Med** 1992;327:1832-5.
27. **DerSimonian R, Laird N.** Meta-analysis in clinical trials. **Controlled Clin Trials** 1986;7:177-88.
28. **Dickersin K, Min YI.** Publication bias: the problem that won't go away. **Ann N Y Acad Sci** 1993;703:135-46; discussion 146-8.
29. **Doll R.** Etiology of lung cancer. **Adv Cancer Res** 1955;3:1-50.
30. **Doll R, Bradford Hill A.** Smoking and carcinoma of the lung. **BMJ** 1950;2:739-748.
31. **Doll R, Bradford Hill A.** The mortality of doctors in relation to their smoking habits. **BMJ** 1954;1:1451-1455.
32. **Doll R, Peto R, Wheatley K, Gray R, Sutherland I.** Mortality in relation to smoking: 40 years' observations on male British doctors. **BMJ** 1994;309:901-11.

33. **Early Breast Cancer Trialists' Collaborative Group.** Effects of adjuvant tamoxifen and of cytotoxic therapy on mortality in early breast cancer. An overview of 61 randomized trials among 28,896 women. **N Engl J Med** 1988;319:1681-92.
34. **Editorial.** Surrogate measures in clinical trials. **Lancet** 1990;335:261-2.
35. **Egger M, Smith GD.** Misleading meta-analysis. **BMJ** 1995;310:752-4.
36. **Ellenberg S, Hamilton JM.** Surrogate endpoints in clinical trials: cancer. **Stat Med** 1989;8:405-13.
37. **Ellenberg SS.** Surrogate end points in clinical trials. **BMJ** 1991;302:63-4.
38. **Enger E.** Kontrollerte kliniske forsøk. Oslo: Universitetsforlaget, 1966.
39. **European Study of Prevention of Infarct with Molsidomine (ESPRIM) Group.** The ESPRIM trial: short-term treatment of acute myocardial infarction with molsidomine. **Lancet** 1994;344:91-7.
40. **Fibrinolytic Therapy Trialists' (FTT) Collaborative Group.** Indications for fibrinolytic therapy in suspected acute myocardial infarction: collaborative overview of early mortality and major morbidity results from all randomised trials of more than 1000 patients. [published erratum appears in *Lancet* 1994 Mar 19;343(8899):742]. **Lancet** 1994;343:311-22.
41. **Fleiss JL.** The statistical basis of meta-analysis. **Stat Methods Med Res** 1993;2:121-45.
42. **Francis T, Kornis RF, Voight RB, Boisen M, al e.** An evaluation of the 1954 poliomyelitis vaccine trials. Summary report. **Am J Pub Health** 1955;45:1-63.
43. **Frazier OH, March RJ, Horvath KA.** Transmyocardial revascularization with a carbon dioxide laser in patients with end-stage coronary artery disease [see comments]. **N Engl J Med** 1999;341:1021-8.
44. **Fugelli P.** Er Helsinkideklarasjonen asosial? **Tidsskr Nor Laegeforen** 1993;113:1327.
45. **Greenland S.** Quantitative methods in the review of epidemiologic literature. **Epidemiol Rev** 1987;9:1-30.
46. **Hennekens CH, Buring JE, Manson JE, Stampfer M, Rosner B, Cook NR, Belanger C, LaMotte F, Gaziano JM, Ridker PM, Willett W, Peto R.** Lack of effect of long-term supplementation with beta carotene on the incidence of malignant neoplasms and cardiovascular disease. **N Engl J Med** 1996;334:1145-9.
47. **Hill AB.** Suspended judgment. Memories of the British Streptomycin Trial in Tuberculosis. The first randomized clinical trial. **Control Clin Trials** 1990;11:77-9.
48. **Hjermann I, Velve Byre K, Holme I, Leren P.** Effect of diet and smoking intervention on the incidence of coronary heart disease. Report from the Oslo Study Group of a randomised trial in healthy men. **Lancet** 1981;2:1303-10.

49. **Hollis S, Campbell F.** What is meant by intention to treat analysis? Survey of published randomised controlled trials. **Bmj** **1999**;319:670-674.
50. **ISIS-4 (Fourth International Study of Infarct Survival) Collaborative Group.** ISIS-4: a randomised factorial trial assessing early oral captopril, oral mononitrate, and intravenous magnesium sulphate in 58,050 patients with suspected acute myocardial infarction. **Lancet** **1995**;345:669-85.
51. **Johnson P, Pearce JM.** Recurrent spontaneous abortion and polycystic ovarian disease: comparison of two regimens to induce ovulation [retracted in *BMJ* 1995 Jul 22;311(6999):231]. **BMJ** **1990**;300:154-6.
52. **Kassirer JP, Angell M.** Redundant publication: a reminder. **N Engl J Med** **1995**;333:449-50.
53. **Kedar RP, Bourne TH, Powles TJ, Collins WP, Ashley SE, Cosgrove DO, Campbell S.** Effects of tamoxifen on uterus and ovaries of postmenopausal women in a randomised breast cancer prevention trial. **Lancet** **1994**;343:1318-21.
54. **Kultur og vitenskapsdepartementet.** Mandat for de regionale komitéer for medisinsk forskningsetikk (19. januar 1989), **1989**.
55. **Kunz R, Oxman AD.** The unpredictability paradox: review of empirical comparisons of randomised and non-randomised clinical trials. **BMJ** **1998**;317:1185-90.
56. **L'Abbe KA, Detsky AS, O'Rourke K.** Meta-analysis in clinical research. **Ann Intern Med** **1987**;107:224-33.
57. **Landmark K, Abdelnoor M.** Magnesiumbehandling ved akutt hjerteinfarkt. **Tidsskr Nor Laegeforen** **1995**;115:2268-70.
58. **Leren P.** The effect of plasma cholesterol lowering diet in male survivors of myocardial infarction. A controlled clinical trial. **Acta Med Scand Suppl** **1966**;466:1-92.
59. **Levine RJ.** The need to revise the Declaration of Helsinki. **N Engl J Med** **1999**;341:531-4.
60. **Lock S.** Lessons from the Pearce affair: handling scientific fraud. **BMJ** **1995**;310:1547-8.
61. **Lowe HJ, Barnett GO.** Understanding and using the medical subject headings (MeSH) vocabulary to perform literature searches. **JAMA** **1994**;271:1103-8.
62. **Lurie P, Wolfe SM.** Unethical trials of interventions to reduce perinatal transmission of the human immunodeficiency virus in developing countries [see comments]. **N Engl J Med** **1997**;337:853-6.
63. **March RJ.** Transmyocardial laser revascularization with the CO2 laser: one year results of a randomized, controlled trial. **Semin Thorac Cardiovasc Surg** **1999**;11:12-8.
64. **Moore TJ.** *Deadly Medicine*. New York: Simon & Schuster, **1995**.
65. **MRC Vitamin Study Research Group.** Prevention of neural tube defects: results of the Medical Research Council Vitamin Study. **Lancet** **1991**;338:131-7.

66. **News.** US Human radiation experiments. **Lancet** 1994;343:106.
67. **Nordrehaug JE.** Hjertelaser - magi eller kur for de sykeste? **Tidsskr Nor Lægeforen** 1999;119:3552.
68. **Oakes M.** The logic and role of meta-analysis in clinical research. **Stat Methods Med Res** 1993;2:147-60.
69. **Ohlsson A, Myhr TL.** Intrapartum chemoprophylaxis of perinatal group B streptococcal infections: a critical review of randomized controlled trials [see comments]. **Am J Obstet Gynecol** 1994;170:910-7.
70. **Perinatal HIV Intervention Research in Developing Countries Workshop participants.** Science, ethics, and the future of research into maternal infant transmission of HIV-1. **Lancet** 1999;353:832-5.
71. **Persen K, Helgesen F, Nordahl A.** Atomforsøk på nordmenn. Oslo: Dagbladet, 1998.
72. **Peto R.** Why do we need systematic overviews of randomized trials? **Stat Med** 1987;6:233-44.
73. **Pocock SJ.** Clinical trials: A practical approach. Chichester: Wiley, 1983.
74. **Psaty BM, Weiss NS, Furberg CD, Koepsell TD, Siscovick DS, Rosendaal FR, Smith NL, Heckbert SR, Kaplan RC, Lin D, Fleming TR, Wagner EH.** Surrogate end points, health outcomes, and the drug-approval process for the treatment of risk factors for cardiovascular disease. **JAMA** 1999;282:786-90.
75. **Reikvam A.** Fra kliniske studier til rutinebehandling. Generelt om kliniske studier innen hjerte- og karsykdommer. **Tidsskr Nor Lægeforen** 1994;114:455-8.
76. **Reikvam A.** Fra kliniske studier til rutinebehandling. Hva avgjør om forskningsresultater implementeres i klinisk rutine? **Tidsskr Nor Lægeforen** 1994;114:459-61.
77. **Retraction.** Recurrent spontaneous abortion and polycystic ovarian disease: comparison of two regimens to induce ovulation [retraction of Johnson P, Pearce JM. In: **BMJ** 1990 Jan 20;300(6718):154-6]. **BMJ** 1995;311:231.
78. **Roberts J.** Focus: Washington "A little of the Buchenwald touch". **BMJ** 1994;308:226.
79. **Robins J, Greenland S, Breslow NE.** A general estimator for the variance of the Mantel-Haenszel odds ratio. **Am J Epidemiol** 1986;124:719-23.
80. **Rochon PA, Gurwitz JH, Simms RW, Fortin PR, Felson DT, Minaker KL, Chalmers TC.** A study of manufacturer-supported trials of nonsteroidal anti-inflammatory drugs in the treatment of arthritis. **Arch Intern Med** 1994;154:157-63.
81. **Rosner B.** Fundamentals of Biostatistics. 3rd ed. Boston: PWS-Kent, 1990.
82. **Rothman KJ, Greenland S.** Modern epidemiology. 2nd ed. Philadelphia: Lippincott-Raven, 1998.

83. **Rothman KJ, Michels KB.** The continuing unethical use of placebo controls. *N Engl J Med* **1994**;331:394-8.
84. **Sacks HS, Berrier J, Reitman D, Ancona-Berk VA, Chalmers TC.** Meta-analyses of randomized controlled trials. *N Engl J Med* **1987**;316:450-5.
85. **Schulz KF.** Randomised trials, human nature, and reporting guidelines. *Lancet* **1996**;348:596-8.
86. **Schulz KF, Chalmers I, Hayes RJ, Altman DG.** Empirical evidence of bias. Dimensions of methodological quality associated with estimates of treatment effects in controlled trials. *JAMA* **1995**;273:408-12.
87. **Smith P, Arnesen H, Holme I.** The effect of warfarin on mortality and reinfarction after myocardial infarction. *N Engl J Med* **1990**;323:147-52.
88. **Sosialdepartementet.** Forskrifter om klinisk utprøving av legemidler (21. august 1981), **1981**.
89. **Stampfer MJ, Goldhaber SZ, Yusuf S, Peto R, Hennekens CH.** Effect of intravenous streptokinase on acute myocardial infarction: pooled results from randomized trials. *N Engl J Med* **1982**;307:1180-2.
90. **Steering Committee of the Physicians' Health Study Research Group.** Final report on the aspirin component of the ongoing Physicians' Health Study. *N Engl J Med* **1989**;321:129-35.
91. **Stephens NG, Parsons A, Schofield PM, Kelly F, Cheeseman K, Mitchinson MJ.** Randomised controlled trial of vitamin E in patients with coronary disease: Cambridge Heart Antioxidant Study (CHAOS) [see comments]. *Lancet* **1996**;347:781-6.
92. **Teo KK, Yusuf S, Collins R, Held PH, Peto R.** Effects of intravenous magnesium in suspected acute myocardial infarction: overview of randomised trials. *BMJ* **1991**;303:1499-503.
93. **Thacker SB.** Meta-analysis. A quantitative approach to research integration. *Jama* **1988**;259:1685-9.
94. **The Alpha-Tocopherol BCCPSG.** The effect of vitamin E and beta carotene on the incidence of lung cancer and other cancers in male smokers. *N Engl J Med* **1994**;330:1029-35.
95. **The Cardiac Arrhythmia Suppression Trial (CAST) Investigators.** Preliminary report: effect of encainide and flecainide on mortality in a randomized trial of arrhythmia suppression after myocardial infarction. *N Engl J Med* **1989**;321:406-12.
96. **The Cardiac Arrhythmia Suppression Trial II Investigators.** Effect of the antiarrhythmic agent moricizine on survival after myocardial infarction. *N Engl J Med* **1992**;327:227-33.
97. **The EUROGAST Study Group.** An international association between *Helicobacter pylori* infection and gastric cancer. [published erratum appears in *Lancet* 1993 Jun 26;341(8861):1668]. *Lancet* **1993**;341:1359-62.

98. **The GUSTO investigators.** An international randomized trial comparing four thrombolytic strategies for acute myocardial infarction. *N Engl J Med* **1993**;329:673-82.
99. **The IFNB Multiple Sclerosis Study Group.** Interferon beta-1b is effective in relapsing-remitting multiple sclerosis. I. Clinical results of a multicenter, randomized, double-blind, placebo-controlled trial. *Neurology* **1993**;43:655-61.
100. **The Norwegian Multicenter Study Group.** Timolol-induced reduction in mortality and reinfarction in patients surviving acute myocardial infarction. *N Engl J Med* **1981**;304:801-7.
101. **The Scandinavian Simvastatin Survival Study (4S).** Randomised trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* **1994**;344:1383-9.
102. **The Standards of Reporting Trials Group.** A proposal for structured reporting of randomized controlled trials. [published erratum appears in *JAMA* 1995 Mar 8;273(10):776]. *JAMA* **1994**;272:1926-31.
103. **Thompson SG.** Controversies in meta-analysis: the case of the trials of serum cholesterol reduction. *Stat Methods Med Res* **1993**;2:173-92.
104. **Thompson SG, Pocock SJ.** Can meta-analyses be trusted? *Lancet* **1991**;338:1127-30.
105. **Tjomsland O, Almdahl SM, Saatvedt K, Nordstrand K, Aaberge L.** Transmyokardial laserbehandling - et nytt tilbud ved uttalt angina pectoris? *Tidsskr Nor Lægeforen* **1999**;119:3597-3601.
106. **Treasure T, MacRae KD.** Minimisation: the platinum standard for trials? Randomisation doesn't guarantee similarity of groups; minimisation does. *BMJ* **1998**;317:362-3.
107. **van Leeuwen FE, Benraadt J, Coebergh JW, Kiemeny LA, Gimbrere CH, Otter R, Schouten LJ, Damhuis RA, Bontenbal M, Diepenhorst FW, et al.** Risk of endometrial cancer after tamoxifen treatment of breast cancer. *Lancet* **1994**;343:448-52.
108. **Warshafsky S, Kamer RS, Sivak SL.** Effect of garlic on total serum cholesterol. A meta-analysis. *Ann Intern Med* **1993**;119:599-605.
109. **Williams JW, Jr., Holleman DR, Jr., Samsa GP, Simel DL.** Randomized controlled trial of 3 vs 10 days of trimethoprim/sulfamethoxazole for acute maxillary sinusitis. *JAMA* **1995**;273:1015-21.
110. **Williams RL, Chalmers TC, Stange KC, Chalmers FT, Bowlin SJ.** Use of antibiotics in preventing recurrent acute otitis media and in treating otitis media with effusion. A meta-analytic attempt to resolve the brouhaha [published erratum appears in *JAMA* 1994 Feb 9;271(6):430]. *JAMA* **1993**;270:1344-51.
111. **Wittes J, Lakatos E, Probstfield J.** Surrogate endpoints in clinical trials: cardiovascular diseases. *Stat Med* **1989**;8:415-25.
112. **Wynder EL, Graham EA.** Tobacco smoking as a possible etiologic factor in bronchiogenic carcinoma. *JAMA* **1950**;143:329-336.

113. **Yusuf S, Collins R, MacMahon S, Peto R.** Effect of intravenous nitrates on mortality in acute myocardial infarction: an overview of the randomised trials. **Lancet 1988**;1:1088-92.
114. **Yusuf S, Peto R, Lewis J, Collins R, Sleight P.** Beta blockade during and after myocardial infarction: an overview of the randomized trials. **Prog Cardiovasc Dis 1985**;27:335-71.