

Bör man randomisera i kliniska prövningar?

Jesper Jerkert
jesper@jerkert.se

Sammanfattning

I samband med kliniska prövningar betyder randomisering en slumpvis uppdelning av försökspersoner på test- och kontrollgrupp. Randomiseringen kan vara enkel eller stratifierad (blockmässig). Med det senare menas att försökspersonerna först indelats i grupper med likartade karakteristika, varefter hälften av personerna i varje grupp lottats att hamna i testgruppen och hälften i kontrollgruppen. Proceduren medger säker balansering av kända faktorer som antas påverka utfallet.

Tre argument brukar framföras för randomisering: (a) randomisering krävs för att statistisk hypotesprövning skall vara tillåten; (b) randomisering ser till att okända, eventuellt påverkande faktorer blir jämnt fördelade mellan försöks- och kontrollgrupp; (c) randomisering bidrar till blindhet. Randomiseringskritiker förnekar giltigheten hos alla dessa argument, men har särskilt inriktat sig på att attackera argument (b). I denna artikel granskas randomiseringskritik som levererats av Urbach (1985) och Worrall (2007), och randomiseringsförsvar från Papineau (1994). Slutsatsen blir att kritiken mot argument (b) är riktig, men att randomisering ändå kan försvaras, både med hänvisning till argument (a) och (c).

1 Inledning

Inom västerländsk sjukvård har man under de senaste decennierna kunnat skönja en viktig strömning. Strömningens bärande tanke är att man, idealt sett, i varje situation skall använda de bästa tillgängliga behandlingsmetoderna. Denna attityd kan verka alldeles självklar, men läkekonst är ett område med många gamla traditioner, och därför har man ibland kunnat ifrågasätta effektiviteten hos vissa behandlingar och ingrepp, även när dessa varit allmänt accepterade och mycket spridda. Behandlingar utan vetenskapligt stöd borde, enligt strömningens företrädare, ge plats för evidensbaserad medicin (eng. *evidence-based medicine*, förkortat *EBM*), med vilket helt enkelt menas behandlingsmetoder och läkemedel för vilka man har vetenskapligt framtagna belägg att tro att de är effektiva.

För att utröna vilka metoder som är effektiva krävs välgjorda undersökningar och systematiska sammanställningar av vad dessa undersökningar visat. Olika behandlingstyper kan utvärderas på olika sätt, och det finns inte

ett gemensamt recept för hur man måste gå tillväga. Icke desto mindre råder tämligen stor enighet om hur man bör gå tillväga i vissa situationer. En sådan situation är när man vill veta huruvida ett nytt läkemedel är bättre eller sämre än ett annat läkemedel mot samma åkomma, alternativt bättre eller sämre än ett placebopreparat. För att ta reda på detta måste man förr eller senare utföra en klinisk prövning (eng. *clinical trial*), det vill säga ett test på verkliga patienter som har åkomman.¹ Det råder omfattande enighet om att sådana undersökningar bör vara *randomiserade* och *kontrollerade* för att man skall kunna fästa stor vikt vid dem. Just förkortningen RCT – *randomized controlled trials* – betecknar de högst ansedda undersökningarna inom medicinen.²

Att en studie är *kontrollerad* betyder bland annat att man har en kontrollgrupp, alltså en grupp personer som inte får det nya läkemedlet, och som jämförs med gruppen som får det nya läkemedlet, kallad försöksgrupp eller interventionsgrupp. Idén är att grupperna skall behandlas exakt lika utom med avseende på någon viss variabel vars inverkan för utfallet man önskar undersöka. En typisk variabel man vill undersöka kan vara vilket preparat som ges, medan ett typiskt utfall man vill mäta är något slags kvantifiering av personernas hälsotillstånd. I en klinisk prövning kan t.ex. försöks- och kontrollgrupp behandlas likadant på alla sätt utom att personerna i försöksgruppen får fastställda doser av ett nytt läkemedel, medan personerna i kontrollgruppen får ett placebopreparat (eller ett annat läkemedel med känd effekt). Den eventuella skillnad man finner mellan grupperna avseende personernas hälsa i slutet av undersökningen blir man då benägen att tillskriva antingen slumpen eller skillnaden i behandling.

Bland filosofer förekommer etiska diskussioner som rör frågan om kontrollgrupper är berättigade då man studerar en allvarlig åkomma och har goda skäl att tro att en viss behandling är klart bättre än den som ges till personerna i kontrollgruppen. Epistemologiskt råder däremot enighet om att användning av kontrollgrupp är ett finurligt sätt att erhålla säkrare kunskap om behandlingsmetoders effektivitet. Situationen är inte densamma rörande randomisering. Randomisering ifrågasätts inte bara av etiska skäl, utan även av epistemologiska. Syftet med denna artikel är att granska några argument för och emot randomisering i aktuell filosofisk debatt, och att försöka ta ställning i den övergripande frågan: bör man randomisera i kliniska prövningar? Nyckelreferenser som kommer att beröras är Urbach (1985), Papineau (1994) och Worrall (2007).

¹Kliniska prövningar brukar utföras i flera faser, framför allt för att försäkra sig om att inga oacceptabla biverkningar uppkommer. Vi fördjupar oss inte här i detta, utan koncentrerar oss på det steg i en fullständig klinisk prövning som är mest intressant för det fortsatta resonemanget, nämligen då behandlingen prövas på personer med åkomman och utfallet hos dessa personer jämförs med utfallet hos personer som inte får behandlingen.

²Ibland förekommer orden i omkastad ordning, *controlled randomized trial*, förkortat CRT.

2 Randomiseringsbegreppet

2.1 Vad betyder randomisering?

Vad betyder randomisering? För att reda ut detta måste vi ta hjälp av begreppet *slumpmässig*, som inte är trivialt. För våra syften räcker dock följande: Om man har ett antal objekt och skall välja något eller några av dem, säger vi att så har skett slumpmässigt om (1) det är lika stor sannolikhet för vart och ett av objekten att bli valt, och (2) valen har inte gjorts i enlighet med någon förutbestämd regel. Villkor (1) innebär att inget objekt är favoriserat framför något annat. Villkor (2) innebär att det inte finns någon systematik i valen, och att det därmed inte går att genomskåda dem genom logiskt tänkande.³

Randomisering, som på engelska heter *randomization* eller *random allocation*, kan i de flesta fall översättas med 'slumpmässig uppdelning i grupper'. Observera alltså, att randomisering *inte* är synonymt med 'slumpmässigt urval' (på engelska *random sampling*). I en psykologisk eller medicinsk undersökning kan både slumpmässigt urval och randomisering förekomma. Man kan göra ett slumpmässigt urval ur den population man önskar uttala sig om. På detta sätt får man alls tag i sina försökspersoner. Därefter kan försökspersonerna randomiseras, det vill säga delas upp slumpmässigt i grupper. Det är två skilda saker, även om de är nära besläktade. Man kan nämligen säga att randomisering tillhandahåller ett slumpmässigt urval ur den population som man redan tidigare skaffat sig på slumpmässig eller icke-slumpmässig väg. (Till detta återkommer jag i avsnitt 5.2.)

Eftersom randomisering är ett vanligt begrepp inom medicinen, och där används om försökspersoner, har ordet ibland kommit att uppfattas som exakt synonymt med 'slumpmässig fördelning av försökspersoner mellan försöksgrupp och kontrollgrupp'. Detta är emellertid en mer inskränkt betydelse än ordets faktiska användning medger. Även inom t.ex. parapsykologin talar man om randomiserade experiment, men där kan man syfta på att bilder som skall användas i telepatiförsök har ordnats i slumpmässig ordning, att en kortlek vars kortmotiv en försöksperson skall gissa har blandats ordentligt, eller dylikt. Försökspersonerna som deltar i parapsykologiska experiment behöver därför inte själva ha fördelats i olika grupper för att begreppet randomisering skall anses motiverat. Randomiseringsbegreppet har en lång historia inom parapsykologin, och filosofen Ian Hacking har hävdat att randomisering som vetenskaplig experimentell praxis först gjorde sin entré just inom parapsykologin (Hacking, 1988).

Ytterligare ett område där man ofta talar om randomisering är lant-

³Jag har formulerat dessa villkor för slumpmässighet själv, men har inte haft som avsikt att avvika från gängse uppfattningar om begreppets innebörd. Villkor (1) torde i någon form ingå i alla definitioner av slumpmässighet, och villkor (2) är en omformulering av den algoritmiska slumpmässighetsdefinition som brukar användas inom informationsteorin.

bruksforskning. Typiska randomiserade lantbruksundersökningar kan gå ut på att vissa grödor sås på vissa jordplättar, och andra grödor sås på andra jordplättar, och sedan kontrolleras vilka som växer bäst. När sådana experiment är randomiserade betyder det att man slumpmässigt har bestämt vilka jordplättar som skall besås med vad. Faktum är att pionjärarbeten inom teorin för vetenskapliga experiment, bl.a. angående randomisering, utförts just med lantbruksforskning i åtanke (Fisher, 1926; Yates, 1939).

2.2 Fullständig randomisering och blockrandomisering

Den enklaste formen av randomisering kan kallas *fullständig randomisering* (eng. *complete randomization*).⁴ Den innebär att randomiseringsproceduren tillämpas på vart och ett av de objekt som skall randomiseras (här: försökspersonerna) utan att dessa ordnats på något annat sätt i förväg. Man tar alltså försökspersonerna en efter en, i valfri ordning, och singlar exempelvis slant för var och en, där krona betyder att personen hamnar i försöksgruppen och klave att personen hamnar i kontrollgruppen (eller tvärtom).

Ganska vanligt är dock att objekten först delas in i ett antal mindre grupper på ett systematiskt sätt, varefter randomisering vidtar inom varje grupp. Ett sådant förfarande kan benämnas *blockrandomisering*. Metoden gör det möjligt att garantera att någon eller några speciella egenskaper hos objekten blir jämnt fördelade mellan de slutliga försöks- och kontrollgrupperna. Låt oss som exempel anta att vi har ett stort antal personer som vi vill dela upp i försöks- och kontrollgrupp för en läkemedelsprövning. Men vi har goda skäl att tro att män och kvinnor reagerar olika på läkemedlet, och vi vill inte riskera att en ojämn könsfördelning mellan grupperna leder oss till att dra fel slutsats om läkemedlets effektivitet. Därför kan vi bestämma att försöks- och kontrollgrupp skall ha samma proportioner mellan män och kvinnor. Det kan åstadkommas genom att vi först delar in försökspersonerna efter kön, och sedan inom varje könsgroup slumpmässigt låter hälften komma till försöksgruppen och resten till kontrollgruppen.

Om vi tror att flera variabler spelar roll för hur man svarar på läkemedlet, kan vi göra en finare gruppindelning innan den slumpmässiga uppdelningen vidtar. Om vi vill att försöks- och kontrollgrupper skall vara balanserade både med avseende på kön och ålder kan vi således först dela alla försökspersoner i två könsgupper, varefter vi delar varje könsgroup i ett antal åldersgrupper (enligt någon lämplig princip, t.ex. i femårsklasser). Slutligen tar vi varje sådan liten grupp, som alltså innehåller personer med identiska kön och likartad ålder, och låter slumpmässigt hälften komma till försöksgruppen och resten till kontrollgruppen. Detta förfarande kan också kallas *stratifierad* (skiktad) randomisering. Olika metoder för att på bästa sätt genomföra stratifierade randomiseringar har föreslagits. En adaptiv metod för

⁴Även kallad *enkel randomisering* (*simple randomization*).

stratifierad randomisering med avseende på ett stort antal faktorer är känd som *minimisering* (Ask, 2004).

Om man driver blockrandomiseringen till sin spets kan man till slut stå med *matchade par*, från vilka randomiseringen skall göras. Ett matchat par innehåller två försökspersoner som är lika varandra i alla avseenden som man tror kan ha betydelse, och som man därför vill se till blir precis jämnt fördelade mellan försöks- och kontrollgrupp. Inom varje matchat par slumpas den ena personen till försöksgruppen och den andra till kontrollgruppen.

Eftersom randomisering således kan vara fullständig eller blockmässig (stratifierad), kan avse försökspersoner, jordplättar eller spelkort, kan syfta till att ge slumpmässig sammansättning eller slumpmässig ordning, med mera, är det svårt att definiera randomiseringsbegreppet på ett entydigt sätt. I fortsättningen skall jag koncentrera mig på betydelsen inom medicinen, alltså slumpmässig uppdelning av försökspersoner i försöksgrupp och kontrollgrupp.

3 Tre argument för randomisering

Eftersom randomisering är ett essentiellt inslag i RCT, vilka i sin tur räknas till de bästa undersökningarna inom evidensbaserad medicin, är det viktigt att göra klart vilka argument som brukar anföras för randomiseringens förträfflighet, och att granska dem. Det går att urskilja tre argument som regelmässigt anföras för randomisering (Worrall, 2002).

3.1 Argument (a): randomisering möjliggör statistisk hypotesprövning

Det första argumentet, låt oss kalla det (a), är att randomisering bidrar till att uppfylla de slumpmässighetsvillkor som krävs för att det skall vara tillåtet att utföra statistisk hypotesprövning. Den moderna statistiska hypotesprövningens förgrundsgestalt, Ronald A. Fisher (1890–1962), uttryckte sig tydligt på denna punkt:

”The purpose of randomization (...) is to guarantee the validity of the test of significance, this test being based on an estimate of error made possible by replication” (Fisher 1951, s. 26, citerat efter Armitage 2003, s. 926).

Argumentet är ingalunda obsolet. Som färskt exempel kan nämnas en svensk biostatistiker som försäkrar att ”den logiska grunden [är] lagd för den efterföljande statistiska analysen” i och med att man randomiserar, och att randomisering därför kan betraktas som ”nyckeln till den statistiska inferensen” (Taube, 2000, s. 4174).

Åtskilliga läroböcker redovisar hur man skall genomföra statistiska tester vid olika typer av blockrandomisering (t.ex. Ostle, 1954, kap. 9–10). Uppenbarligen är det introduktionen av slumpmässighet överhuvudtaget som är viktig enligt argument (a), inte den exakta proceduren. Varför är det principiellt viktigt att introducera slumpmässighet? Jo, osäkerheten i varje mätning kan då göras oberoende av de andra mätningarnas osäkerheter. I en statistisk lärobok uttrycks det på följande sätt:

”As we saw when we discussed the assumptions underlying the techniques used in testing various statistical hypotheses, it was generally assumed that the ‘errors’ were independent from one observation to another. By randomization, we may regard this requirement as satisfied” (Ostle, 1954, s. 424).

Att osäkerheterna är oberoende tillåter i sin tur att den totala datainsamlingen kan beskrivas av väldefinierade statistiska funktioner, vilket är ett villkor för hypotesprövningen. Det heter ju *statistisk* hypotesprövning, och ordet statistisk skall tolkas så att det rör sig om en teori som inkluderar slumpmässighet. Denna slumpmässighet tillhandahålls av randomiseringen.

Som ett enkelt exempel, antag att utfallet i varje mätning kan bli på endera av två sätt, att sannolikheten för det ena utfallet är p under nollhypotesen, att vi gör totalt n mätningar och att varje mätning är oberoende av de andra. Då kan varje utfall modelleras som en Bernoulli-fördelning och det totala antalet utfall av endera sorten blir binomialfördelat med parametrar n och p . Hypotesprövningen fortskrider sedan enligt gängse mönster, vilket innebär att nollhypotesen förkastas (som man uttrycker det) om det totala utfallet hamnar tillräckligt långt ut i endera svansen hos nollhypotesens sannolikhetsfördelningsfunktion.⁵

Det kan synas uppstå problem med oberoendet vid randomisering ur matchade par. Här är ju mätningarna beroende inom varje par. Detta kan dock undvikas genom att man betraktar skillnaden inom varje par som utfallsmått. Då får vi oberoende mellan varje par.

3.2 Argument (b): randomisering kontrollerar för okända obalanser

Det andra argumentet, (b), är att randomiseringen kontrollerar relevanta faktorer som kan tänkas påverka utfallets resultat. Att randomiseringen ”kontrollerar” faktorerna innebär att faktorerna bringas under kontroll, genom att de (åtminstone med hög sannolikhet) förekommer i likartad utsträckning i bägge grupperna. Detta argument är mycket vanligt. Följande formulering är hämtad ur en ansedd bok om RCT:

⁵Se nästan vilken lärobok som helst i statistik för ytterligare information om hur klassisk hypotesprövning går till.

”Genom att fördela deltagarna slumpmässigt mellan de olika behandlingsalternativen har man utsikter att uppnå att deltagargrupperna blir likartat sammansatta vid provningens början (...) Randomiseringens värde ligger bl.a. i att den (...) minskar risken för svåra obalanser mellan provningsgrupperna beträffande viktiga okända faktorer som kan inverka på hur deltagarna reagerar på de interventioner som provas” (Jadad, 2000, s. 21).

Och här är ett liknande yttrande ur ett standardverk om experimentell design:

”[W]e need some means of insuring that a treatment will not be continually favored or handicapped in successive replications by some extraneous source of variation, known or unknown. This is done by the device known as randomization, due to Fisher” (Cochran & Cox, 1957, s. 6).

Observera att det talas om såväl kända som okända faktorer i det sista citatet. Utan tvivel har dock argumentet mest använts om okända faktorer. Kända faktorer kan ju balanseras genom uppdelning i block, och om man väljer att randomisera (det vill säga tillämpa en slumpmekanism) utifrån uppdelade block kan man inte gärna göra det för att man tror att det är viktigt för de kända faktorerna, utan för de okända. Argument (b) är dock förenligt med både fullständig randomisering och med olika typer av blockrandomisering, men vid fullständig randomisering får man förmoda att den som hänvisar till argument (b) tänker sig att randomiseringen balanserar *både* kända och okända faktorer på ett tillfredsställande sätt.

3.3 Argument (c): randomisering bidrar till blindning

Det tredje vanliga argumentet, (c), lyfter fram blindning (maskering) som det egentliga, underliggande syftet med randomisering. Om man slumpmässigt fördelar personer i grupper, och samtidigt ser till att hålla indelningen hemlig för såväl försöksdeltagare som försöksledare, har man uppnått s.k. dubbelblindhet (Jadad, 2000, s. 42f). Om blindningen även i övrigt skötts kompetent betyder det att ingen person som kan ha inflytande över studiens resultat kunna säga vilken försöksdeltagare som hör till vilken grupp.

Man kan naturligtvis gissa vilka personer som hör till vilken grupp, men sådana gissningar kan inte bli bättre än slumpen medger, om blindningen genomförts tillfredsställande. En person som gissar försöksdeltagarnas grupp-tillhörighet kan mycket väl göra detta utifrån uppfattbara skillnader, t.ex. gissningen att alla långa personer hör till den ena gruppen, och alla korta till den andra, eller gissningen att alla kvinnor tillhör den ena gruppen, och alla män den andra. Men poängen är att även om man gissar utifrån sådana upptäckbara kriterier, så kan inte den verkliga gruppindelningen vara relaterad

till sådana kriterier mer än av en ren slump, eftersom indelningen gjordes slumpmässigt, och en av aspekterna på slumpmässighet är ju avsaknaden av systematik och regelbundenhet.

3.4 Olika syn på slumpmässighetens roll

Inget av argumenten, (a), (b) och (c) kan enligt min mening omedelbart avfärdas, utan måste närmare granskas innan man kan avgöra om de är hållbara eller ej. Dock skall medges att argument (c) vid första anblicken inte verkar ha samma principiella tyngd som (a) och (b). Blindning kan ju åstadkommas även utan randomisering.

Att döma av litteraturen framförs (a), (b) och (c) ibland isolerade, ibland i kombination. Argumenten skulle kunna karakteriseras utifrån hur teoretiskt eller praktiskt de betraktar den slumpmässighet som är en essentiell del av randomisering. I argument (a) ses slumpmässigheten som en teoretisk nödvändighet, men man behöver inte hänvisa till några särskilda praktiska konsekvenser för att anse argumentet berättigat. Slumpmässigheten är en teoretisk lösning på ett teoretiskt problem. I argument (b) betraktas slumpmässigheten som en teoretisk tillfredsställande lösning på ett praktiskt problem, nämligen hur man skall se till att okända faktorer som kan ha betydelse för utfallet fördelas jämnt mellan försöks- och kontrollgrupp. I argument (c) är de praktiska konsekvenserna helt i fokus.

Dessa skillnader i hur allvarligt man tar på slumpmässighet återspeglas också i olika syner på huruvida man får randomisera på nytt om en första randomisering givit upphov till grupper som är uppenbart obalanserade i något avseende som man tror är betydelsefullt. Enligt argument (b) är själva poängen med randomisering att se till att grupperna blir likartade och därför jämförbara. Om man efter genomförd randomisering kan konstatera att grupperna är olika med avseende på viktiga kända faktorer, vad vore mer naturligt än att göra om randomiseringen? (De okända faktorerna kan man förstås inte kontrollera, eftersom de är just okända.) Enligt argument (a) är det däremot tvivelaktigt att göra om en randomisering. Randomiseringen genomförs nämligen enligt (a) för att tillhandahålla den slumpmässighet som gör det möjligt att anta en modell där det totala utfallet beskrivs av en väldefinierad slumpvariabel som kan användas för hypotesprövning. Att inte tillåta vissa randomiseringar är liktydigt med att modifiera slumpvariabeln som enligt modellen beskriver det totala utfallet. Därmed borde det inte längre vara tillåtet att genomföra hypotesprövningen (Urbach, 1985, s. 266), åtminstone inte utan att ta reda på hur slumpvariabelns sannolikhetsfördelningsfunktion ändras genom omrandomiseringen.⁶ Detta torde vara en formidabel modelleringsuppgift.

⁶Märkligt nog har R. A. Fisher uttalat sig för att omrandomisering skulle vara tillåten, trots att han ju varit en tydlig förespråkare för argument (a). Fisher citeras i denna fråga i Savage, Barnard & Cox (1962), s. 88, som i sin tur citeras av Urbach (1985), s. 260.

4 Randomisering och bayesianism

Filosofier som är kritiska mot randomisering är ofta bayesianer, eller hämtar åtminstone vissa argument från bayesianismen. Det kan därför vara på sin plats att presentera några drag i bayesianismen som bakgrund till det fortsatta resonemanget.⁷

Bayesianism är en vetenskapsfilosofisk skolbildning som fått sitt namn efter *Bayes' sats*, en matematisk relation inom sannolikhets teorin, känd sedan 1700-talet.⁸ Satsen lyder i sin enklaste form

$$p(B|A) = \frac{p(A \& B)}{p(A)} = \frac{p(A|B)p(B)}{p(A)},$$

där $p(B|A)$ står för sannolikheten för B givet A . Vi kan alltså enligt satsen ta reda på sannolikheten $p(B|A)$ om vi känner till den omvända sannolikheten $p(A|B)$ samt $p(B)$ och $p(A)$.⁹ Om man har behövliga data kan satsen exempelvis användas för att finna sannolikheten att en person är dopad givet att ett dopingtest säger att så är fallet. I sådana sammanhang är användningen av Bayes' sats helt okontroversiell. Men bayesianer menar dessutom att satsen – och överhuvudtaget alla etablerade sannolikhetslagar – kan och bör tillämpas i fråga om vetenskaplig kunskapstillväxt. Om t.ex. B står för en hypotes och A står för någon evidens som man funnit, kan Bayes' sats användas för att avgöra hur sannolikheten för hypotesen skall bedömas med hänsyn tagen till den funna evidensen. Bayes' sats kan då ses som en uppdateringsregel: ur *a priori*-sannolikheten $p(B)$ fås *a posteriori*-sannolikheten $p(B|A)$ i enlighet med satsen.

För att ett sådant resonemang skall kunna tillämpas på vetenskaplig kunskapstillväxt krävs att man sätter sannolikhets siffror på kunskapsmässig tilltro eller osäkerhet. Det verkar svårt att finna någon objektiv grund för sådana sifferbedömningar, och sannolikheter måste inom det bayesianska synsättet därför betraktas som subjektiva, mentala tillstånd. Detta går stick i stäv med den frekventistiska syn på sannolikhet som förefaller naturlig när man diskuterar sannolikhets teoretiska räkneproblem med bollar i urnor m.m.

Bayesianism möter motstånd dels för sin subjektivistiska syn på sannolikhet, dels för att inte alla håller med om att kunskapsmässig osäkerhet kan uttryckas i numerisk form. Det finns också experiment som visar att även om man uttrycker osäkerheter och preferenser i form av sannolikheter, kombineras de inte alltid i enlighet med sannolikhetslagarna. På detta kan

⁷För en fyllig introduktion till bayesiansk vetenskapsfilosofi, se Howson & Urbach (2005). För en kortare och mer matematisk introduktion, se Lindley (2000).

⁸Uppkallad efter den engelske prästen och amatörmatematikern Thomas Bayes (ca 1702–1761).

⁹Kanske känner vi inte till $p(B)$ och $p(A)$ direkt, men kan räkna ut dem med hjälp av lagen om total sannolikhet eller andra tillämpliga räkneregler. Bayes' sats får då på motsvarande sätt en krångligare utformning. Vi går inte här närmare in på sådant.

bayesianer svara att sådana resultat inte visar att bayesianism är fel, utan att människors kalibrering och beslutsfattande borde förbättras så att de överensstämmer med sannolikhetslagarna (Hansson, 2005, s. 37ff).

Bayesianer ber inte om ursäkt för att deras filosofi har ett subjektivt sannolikhetsbegrepp. Tvärtom pekar de gärna på att subjektiva inslag också finns hos dem som påstår sig vara strikt objektiva. Som exempel kan vi ta följande citat ur en lärobok av icke-bayesianer: "The occasions on which randomization is required vary with the type of experiment and must be left to the judgment of the experimenter" (Cochran & Cox, 1957, s. 7). Denna typ av uttalanden tar bayesianer gärna som intäkt för att de själva har rätt: de tycker att de ärligt medger att subjektiva bedömningar spelar en väsentlig roll i vetenskapen (medan icke-bayesianer bara motvilligt erkänner det), och de insisterar att det bästa sättet att få kontroll över dem är att klä dem i sannolikhetsdräkt och sedan tillämpa kända sannolikhetslagar, t.ex. Bayes' sats.

Överhuvudtaget betonar anhängare av bayesianismen att vetenskaplig kunskapstillväxt alltid sker, och *bör* ske, mot bakgrund av vad man redan vet (eller tror sig veta). Bayesianer är skeptiska mot traditionell statistisk hypotesprövning à la Fisher. Klassisk hypotesprövning innebär nämligen inte att kunskapen uppdateras i enlighet med Bayes' sats med vederbörlig hänsyn tagen till tidigare kunskaper, utan i normalfallet att man prövar en hypotes som få tror på eller intresserar sig särskilt mycket för (nollhypotesen). Bayesianer kan kritisera fullständig randomisering på liknande sätt: förfarandet tar inte hänsyn till att man kan ha välgrundade gissningar kring vilka faktorer som kan tänkas spela in. Blockrandomisering är vettigare från bayesiansk synvinkel, men bayesianer ifrågasätter ändå om det slumpmässiga inslaget som finns kvar har något epistemiskt värde.

Låt oss nu göra några nedslag i David Papineaus argumentation för randomisering och i John Worralls och Peter Urbachs argumentation mot randomisering.

5 Nedslag i Papineaus argumentation för randomisering

5.1 Referat

Papineaus (1994) argumentation tar sin utgångspunkt i en tänkt experimentell situation. Han resonerar såsom följer. Antag att en viss behandling, T (för *treatment*), sätts in mot någon viss åkomma, och att sannolikheten för tillfrisknande, R (för *recovery*), befins vara större i befolkningen i stort om man använder T jämfört med om man inte använder T . Med andra ord har vi $p(R|T) > p(R|\sim T)$, där \sim står för negation. Hur kan vi veta om T orsakar R eller ej, frågar Papineau? Det kan vi förstås inte enbart utifrån en sådan

korrelation, eftersom det skulle kunna vara någon annan variabel som skilt sig åt mellan grupperna, och som kan vara den verkliga orsaken till att vi funnit $p(R|T) > p(R|\sim T)$ i just vårt fall. Detta kan vi undersöka. Om vi misstänker att ålder inte blivit likafördelad kan vi inom varje grupp kontrollera det (förutsatt att vi har tillgång till åldersdata). Vi frågar oss då om $p(R|T\&Y) > p(R|\sim T\&Y)$ och om $p(R|T\&\sim Y) > p(R|\sim T\&\sim Y)$, där Y betecknar lägre ålder (*younger*). Men även om bägge dessa olikheter skulle vara sanna, har vi inte bevisat att T orsakar R , ty det kan finnas någon annan variabel X som skiljer sig åt mellan grupperna och för vilken det inte är sant att både $p(R|T\&X) > p(R|\sim T\&X)$ och $p(R|T\&\sim X) > p(R|\sim T\&\sim X)$. Denna faktor X kan vara den egentliga orsaken till att det i förstone såg ut som att R och T var korrelerade, fastän de inte objektivt är det.

Problemet är att det finns många tänkbara X . Det verkar omöjligt att kontrollera dem alla. Men enligt Papineau är det precis detta problem som löses genom randomisering:

”A randomized experiment solves the problem. You take a sample of people with the disease. You divide them into two groups at random. You give one group the treatment, withhold it from the other (...), and judge on this basis whether the probability of recovery in the former group is higher. If it is, then T *must* now cause R , for the randomization will have eliminated the danger of any confounding factors which might be responsible for a spurious correlation” (Papineau, 1994, s. 439, kursiv i original).

Papineau fortsätter i samma säkra stil:

”Since the treatment has been assigned at random – in the sense that all patients, whatever their other characteristics, have exactly the same objective probability of receiving the treatment T – we can now be sure that T *is not* itself objectively correlated with any other characteristic that influences R . So we can rule out the possibility of a spurious correlation (...)” (ibid., s. 440, kursiv i original).

Liknande formuleringar om att randomisering medför att T ”definitivt inte är objektivt korrelerad” med andra faktorer finns även senare i hans text (ibid., s. 444f [citat] och s. 446).

Papineau tar självmant upp invändningen att det av en slump kan bli en ojämn fördelning i något visst avseende mellan grupperna (ibid., s. 447f). Även här kommer han dock tillbaka till ”objektiva korrelationer” som en central del av argumentet, och det verkar inte gå att reda ut vad han menar med mindre än att frasens betydelse klargörs.

Papineau för även ett resonemang kring det som jag har kallat argument (a). Han förnekar att randomisering skulle vara nödvändig för att statistiska

signifikanstest skall vara tillåtna (Papineau, 1994, s. 442). Det är ”slumpmässiga urval” (*random samples*) som rättfärdigar statistiska signifikanstest, skriver Papineau (*ibid.*). Han förklarar att slumpmässigt urval och randomisering inte är samma sak, och verkar med slumpmässigt urval endast syfta på en procedur där försökspersonerna rekryteras.

5.2 Bedömning

Papineaus argumentation kring en behandling (T) som eventuellt leder till tillfrisknande (R) är utan tvivel av typ (b), det vill säga han menar att randomisering bidrar till att göra grupperna likartade och därmed meningsfullt jämförbara. Han uttrycker sig mycket starkt då han säger att randomisering garanterar att ingen missvisande korrelation uppkommer. (Jämför med det försiktigare citatet från Jadad, ovan i avsnitt 3.2.) Ytterligare två saker är noterbara. För det första är Papineaus resonemang inriktat på att fastslå när det är tillåtet att dra kausala slutsatser ur experiment. Vi skall inte behandla den frågan närmare, eftersom det skulle leda för långt att blanda in orsaksbegreppet, och eftersom det går åtminstone delvis att bedöma Papineaus argument även utan att ta ställning till frågan om orsaker.¹⁰

För det andra är det svårt att förstå vad Papineau menar med ”objektiv sannolikhet” och ”objektiv korrelation”. Det är ingen försumbar fråga, eftersom uttrycken förekommer på centrala ställen i Papineaus argumentation. Papineau kan inte gärna med ”objektiv sannolikhet” syfta på den faktiska uppdelning i försöks- och kontrollgrupp som föreligger i ett pågående experiment, även om det hade varit rimligt att kalla uppdelningen ”objektiv” eftersom den är omedelbart observerbar. Den uppdelning som föreligger kan ju mycket väl vara obalanserad i något viktigt avseende, och det verkar därför meningslöst att påstå att en ”objektiv korrelation” inte skulle kunna föreligga i den bemärkelsen.

Så vad kan Papineau mena? Han skriver att randomisering inte har att göra med vår bedömning av vilka objektiva sannolikheter som föreligger. Randomisering kommer in senare, efter att vi redan format vår uppfattning om objektiva sannolikheter, enligt Papineau (1994, s. 439). Jag tror inte att vi behöver ta ställning till om detta är riktigt för att kunna bedöma Papineaus trovärdighet i fråga om ”objektiva sannolikheter”. I samband med att han resonerar kring den självklara invändningen att grupperna kan bli obalanserade av en slump, återkommer han nämligen till ämnet och skriver:

”After all, if T did not cause R , how *could* there be such a correlation (an objective correlation in the underlying probability space, remember, which will show up, not just in this sample, but in the long-run frequencies as the randomized experiment is

¹⁰Papineau har skrivit mer utförligt om hur han ser på möjligheten att dra kausala slutsatser ur statistik i Papineau (1985). Se även Rubin (1978).

done time and again) given that the randomization will ensure that all other causes of R are probabilistically independent of T in the long run?" (Papineau, 1994, s. 447, kursiv i original)

Om T inte orsakade R, hur kan det då finnas en objektiv korrelation, undrar Papineau? Inom parentes förklarar han vad han menar med objektiv korrelation: det är en korrelation som visar sig i långa loppet, därför att den föreligger i den population från vilka försökspersonerna överhuvudtaget rekryteras. (Det är åtminstone den enda tolkning av Papineau som jag finner rimlig.) Om objektiv korrelation skall förstås på detta sätt är det snarast trivialt att konstatera att korrelationen troligen kommer att dyka upp i randomiserade undersökningar som utgår från populationen där korrelationen objektivt föreligger. Men Papineaus språkbruk verkar här strida mot vad han skrev på s. 440 (citerat ovan), där han ger det bestämda intrycket att det är randomiseringen – den slumpmässiga uppdelningen – som leder till att det råder en "objektiv korrelation". Sammantaget har Papineau inte givit några övertygande skäl för att man skall ta hans resonemang om objektiva sannolikheter på allvar.¹¹

Papineaus diskussion kring argument (a), slutligen, övertygar inte heller. Han har naturligtvis helt rätt i att slumpmässigt urval inte är detsamma som randomisering. Slumpmässigt urval brukar man ju tala om när man alls får tag i sina försökspersoner, medan randomisering är en process som tillämpas på de redan funna försökspersonerna. Men Papineau reflekterar inte över att randomisering i praktiken tillhandahåller ett slags slumpmässigt urval, nämligen ur den population som man har rekryterat. Om man har en population P och randomiserar den till försöksgrupp F och kontrollgrupp K, kan man säga att såväl F som K uppkommit genom slumpmässigt urval ur P. Både villkor (1) och villkor (2) nämnda i avsnitt 2.1 är uppfyllda. Villkor (1) är utan tvivel uppfyllt, det vill säga urvalet har gjorts så att varje person har haft samma sannolikhet att hamna i endera gruppen. Vid blockrandomisering kan det i förstounde verka tvivelaktigt om även villkor (2) är uppfyllt. Villkor (2) är att valen inte gjorts i enlighet med någon förutbestämd regel, och att det därför inte går att finna någon systematik genom logiskt tänkande. Tag extremfallet att randomisering sker från matchade par. Finns inte här en genomskådningsbar systematik i gruppindelningen? Jo, men denna systematik gäller endast parens sammansättning och säger inget om vilken person i paret som hamnat i vilken grupp. Om man finner två försökspersoner som är mycket lika i avseenden som man tror spelar roll är det med andra ord förhöjd sannolikhet att de ingår i samma par och därför tillhör olika grupper. Så långt kan man komma med logiskt tänkande. Men det är fortfarande omöjligt att gissa vilken som hör till vilken grupp. Även vid randomisering ur matchade par gäller därför: om man plockar ut en enskild person kan man inte genom logiskt tänkande komma fram till vilken grupp han/hon tillhör,

¹¹Invändningar som liknar mina har även framförts av Worrall (2007), s. 464.

inte ens om man skulle få studera alla andra personer också. Jag vill således hävda att även randomisering ur matchade par uppfyller villkoren (1) och (2) och därför tillhandahåller ett slumpmässigt urval, om än endast ur den redan rekryterade populationen P.

Förekomsten av randomisering säger naturligtvis inget om hur vi fann P från början; det kan ha skett genom ett slumpmässigt urval ur en ännu större population eller på något icke-slumpmässigt sätt. Men det är oväsentligt i sammanhanget. Statistisk hypotesprövning innebär att man jämför det totala utfallet från ett antal oberoende mätningar¹² med en sannolikhetsfördelningsfunktionen för en nollhypotes. För att en sådan jämförelse skall vara meningsfull måste den insamlade datamängden också vara genererad från en tänkt sannolikhetsfördelning. Detta uppfylls genom randomisering av alla typer som vi berört. Papineau har enligt min mening inte förklarat sin avvikande inställning i denna fråga på ett övertygande sätt.

6 Nedslag i Urbachs och Worralls argumentation mot randomisering

Peter Urbach (1985) och John Worrall (2007) har levererat intressant kritik mot randomisering. De är båda skeptiska mot argument (b), och eftersom jag i stort sett håller med dem tänkte jag inte orda om det. Istället skall jag kommentera några saker i deras texter som jag inte finner lika lätta att svälja. Jag blandar referat och kritiska synpunkter i det nedanstående.

Peter Urbachs (1985) kritik mot randomisering går längs bayesianska linjer. Han påtalar att det är ologiskt att rättfärdiga randomisering med argument (a) om man samtidigt tillåter sig att förkasta vissa randomiseringar därför att de resulterade grupperna blivit uppenbart obalanserade i något hänseende som man tror kan ha betydelse (Urbach, 1985, s. 260). Det håller jag med om. Men jag tycker att Urbach gör det lätt för sig när han undlåter att gå till botten med blockrandomiseringens möjligheter. Han ger exempel på hur en matchning troligen skulle ge bättre balans än en fullständig randomisering (ibid., s. 266f), men han reflekterar inte över möjligheten att *både* matcha *och* randomisera.

Urbach förefaller vara mycket ovillig att alls föra in något slumpelement i vetenskapliga experiment. Inte ens argument (c) verkar han ge särskilt mycket för:

”[I]t should be noted that a randomized design is not the only, or even the best, way of instilling confidence that the trial was free from unconscious experimenter bias. The same result could be

¹²Åtminstone är detta en korrekt beskrivning i samband med kliniska prövningar. I andra sammanhang kan en statistisk hypotesprövning ibland göras utifrån en enda mätning eller ett enda utfall.

got by insisting on the allocation being done by a person without any axe to grind or by one who has no ability or knowledge to exercise a prejudice. For example, in a medical trial the selection of patients might be entrusted to an independent, non-medical person (...)" (Urbach, 1985, s. 271).

Finns verkligen sådana personer, som kan anförtros att göra den slutliga gruppbestämningen och som garanterat är lika opåverkbara som en ren slumpprocess? Jag tvivlar.

John Worrall (2007) argumenterar vältaligt mot randomisering, delvis i uttrycklig polemik mot just Papineau. Jag kan hålla med om nästan allt han säger. Men han vill inte godkänna argument (a), eftersom han inte tycker att klassisk statistisk hypotesprövning är en god praxis. Därför kan han tillåta sig att säga till exempel följande:

"Once you have made sure that there is no positive reason to think the two groups are unbalanced (...), then whether or not the division was produced by following some table of random numbers or tossing a fair coin, or just by happenstance, can be of no epistemic account" (Worrall, 2007, s. 466).

Detta resonemang är riktigt om man helt och hållet rör sig inom argument (b), men det är förstås oriktigt om man godtar argument (a).

Worrall är försiktigare än Urbach när det gäller att utesluta att randomisering skulle kunna ha något värde alls. Worrall är inne på att icke-randomisering ofta betyder bristande blindhet i någon form, vilket kan ge obalanser. Men kanske kan man undvika sådana obalanser utan att randomisera, resonerar han:

"If, in some particular test, this bias has been eliminated by dint of randomizing, then that is obviously to the good; but if this bias can, in other particular circumstances, be eliminated by other means, then randomizing seems to be left with no further epistemic virtue" (Worrall, 2007, s. 466).

Jag anser att om randomisering bidrar till blindhet – argument (c) – så bör det räknas randomiseringen tillgodo, även om det skulle finnas andra sätt att ordna blindhet på. Worrall är å sin sida inte benägen att tillerkänna randomisering någon som helst förtjänst i anknytning till argument (c).

7 Diskussion och slutsats

Jag har ovan riktat kritik både mot Papineaus försvar för randomisering och mot enstaka resonemang i Worralls och Urbachs attacker på randomisering.

Nedan skall jag redovisa hur jag sammantaget bedömer argumenten (a), (b) och (c), och jag skall på grundval av dessa bedömningar ta ställning till den övergripande frågan om randomisering är rekommendabel i kliniska prövningar.

7.1 Bayesianism kontra argument (a)

Hos Worrall och Urbach finns en tendens att vilja tillskriva motståndarna bayesianska tankemönster därför att motståndarna går med på att fullständig randomisering kan ge upphov till obalanserade grupper och att detta kan undvikas genom att man subjektivt bedömer vilka faktorer som det vore önskvärt att balansera (Urbach, 1985, s. 271). När motståndarna således avslöjats med att "egentligen" tänka bayesianskt avfärdas även den klassiska statistiska hypotesprövningen, som ju bayesianser inte gillar. Därmed kan Worrall och Urbach avfärda argument (a) med vad som verkar vara ett principiellt argument, men som inte är det. Bara för att man går med på att subjektiva bedömningar kommer in i blockindelningen betyder det inte att man skriver under på hela den bayesianska filosofin. Jag kan inte se annat än att det är fullt möjligt att acceptera vissa subjektiva bedömningar men ändå vilja göra en statistisk hypotesprövning enligt klassiskt mönster (där en nollhypotes prövas).

Argument (a) är därför, såvitt jag kan se, riktigt. Om man verkligen vill genomföra klassisk statistisk hypotesprövning så krävs randomisering i någon form. Om man däremot inte vill genomföra en hypotesprövning så faller argument (a).

7.2 Argument (b) och (c), och frågan om slumpmässighetens värde

Hur är det med argument (b)? Här måste jag ge randomiseringens kritiker rätt i stora stycken. Det är således inte riktigt, som argument (b) utsäger, att randomisering i största allmänhet har en alldeles speciell förmåga att balansera försöks- och kontrollgrupper. Olika former av blockrandomiseringar kan visserligen åstadkomma god balans, men där är det ju inte det slumpmässiga elementet som åstadkommer balansen, utan den blockindelning som görs högst medvetet (ej slumpmässigt) före den slutliga grupptillhörighetsbestämningen. Om man väl har skapat block (eller matchade par) som innehåller individer med likartade karakteristika i alla avseenden man tror kan vara relevanta, är det långt ifrån uppenbart att en slumpmässig bestämning av vilken individ i varje block som hamnar i vilken betingelse (försöks- eller kontrollgrupp) skulle vara bättre än en icke-slumpmässig bestämning. Oavsett bestämningsmetod så blir ju försöks- och kontrollgrupperna balanserade i alla avseenden som blockindelningen tagit hänsyn till. Så långt kan jag instämma i bayesianernas argumentation, och jag håller alltså med bayesia-

nera om att slumpmekanismen i randomisering inte tycks ha något särskilt kunskapsvärde om man enbart håller sig inom ramen för argument (b).

Icke desto mindre ser jag två möjligheter att ej försumbara obalanser uppkommer genom att slumpmässigheten helt slopas. De kan sammanfattas i fraserna ”synlig obalans kan ge bristande blindhet” och ”synlig obalans kan korrelera med okänd påverkande faktor”. Bägge dessa tar hjälp av argument (c) och skall diskuteras nedan.

7.2.1 Synlig obalans kan ge bristande blindhet

Om den slutgiltiga uppdelningen görs av en person kan denna se till att försökspersoner med något slags kännetecken oftare hamnar i den ena gruppen än den andra. Den risken minskas betydligt om uppdelningen sker med en slumpmekanism. Vilka ”kännetecken” skulle det kunna röra sig om? Det kan röra sig om någon faktor som verkligen skulle kunna påverka utfallet, och som personen som gör den slutliga gruppbestämningen känner till eller anar. Om det är möjligt att obalansera grupperna med avseende på en sådan faktor så kan man förstås säga att blockindelningen inte varit tillräckligt omfattande – att den helt enkelt varit dålig – och att en dylik klinisk prövning inte borde ha fått fortgå så långt som till den slutliga gruppbestämningen (oavsett om denna görs med en slumpmekanism eller ej). Men även om blockindelningarna alltid var perfekta och tog hänsyn till alla rimliga faktorer, kan man fortfarande tänka sig att personen som gör gruppbestämningen obalanserar grupperna med avseende på någon faktor som enligt nuvarande kunskap inte i sig självt påverkar utfallet men som likväl kan vara synlig och upptäckbar för andra personer.

Ett enkelt exempel kunde vara ögonfärg. En obalans med avseende på ögonfärg mellan försöks- och kontrollgrupp påverkar inte det strikt fysiologiska utfallet (antar vi för argumentets skull), men utgör ändå en potentiell felkälla genom att personer som under experimentets gång kommer i kontakt med försökspersonerna skulle kunna känna till denna obalans och därmed inte vara blinda för vilken försöksperson som hör till vilken grupp. Om personer som kommer i kontakt med försökspersonerna kan gissa grupptillhörighet så kan det öppna för otillbörlig påverkan, medveten eller omedveten, som snedvrider resultatet. I detta resonemang har jag uppenbarligen tagit hjälp av argument (c), eftersom det riskabla i att inte låta en slumpmekanism bestämna den slutliga grupptillhörigheten ligger i den bristande blindheten.

7.2.2 Synlig obalans kan korrelera med okänd påverkande faktor

Det finns också en möjlighet att en obalans med avseende på t.ex. ögonfärg påverkar det fysiologiska utfallet, nämligen om ögonfärg är korrelerad med en för oss okänd faktor som i sin tur påverkar utfallet. Alla *kända* potentiellt påverkande faktorer kan vi ta hand om genom blockindelningen (eller

skapandet av matchande par), men därutöver kan det förstås finnas okända faktorer som påverkar utfallet, och någon eller några av dessa kan vara korrelerade med t.ex. ögonfärg utan att vi på vår nuvarande kunskapsnivå känner till något skäl att tro att en obalans i ögonfärg skulle spela någon roll.

Detta låter nog som en högst osannolik situation. Förmodligen är det så, men den principiellt avgörande frågan är om risken är större att sådana korrelationer med okända påverkande faktorer skall uppstå vid icke-slumpmässig gruppbestämning än vid slumpmässig gruppbestämning (randomisering). Jag vill hävda att så är fallet, åtminstone om gruppbestämningen görs av en medveten person.

En medveten person kan nämligen låta upptäckbara kännetecken (upptäckbara för synen eller andra sinnen) styra gruppbestämningen. Många sådana kännetecken är biologiska.¹³ Om det således uppstår en biologiskt relaterad gruppskillnad så finns en förhöjd sannolikhet att den är korrelerad med någon för oss okänd biologisk faktor som påverkar utfallet, jämfört med sannolikheten att en slumpmässigt uppkommen gruppskillnad korrelerar med den okända biologiska faktorn.

För att förklara samma sak igen med andra ord: Sannolikheten att det uppstår en korrelation mellan en okänd påverkande faktor X och en upptäckbar faktor Y (t.ex. ögonfärg) som blivit obalanserad vid en ej slumpmässig gruppbestämning är större än sannolikheten att få en korrelation mellan X och en slumpvist uppkommen obalans Z, därför att det är troligare att Y är av biologisk art än att Z är det, vilket i sin tur beror på att åtskilliga biologiska faktorer kan uppfattas med våra vanliga sinnen. Och eftersom det är högst troligt att X är av biologisk art, är det sannolikare att få en önskad korrelation mellan X och Y än mellan X och Z.

Även i detta resonemang faller jag alltså tillbaka på argument (c) i någon mån, eftersom jag tänker mig att en icke-slumpmässig gruppbestämning görs av en medveten person som känner till något om försökspersonerna och därför inte är blindad. Jag har inte upptäckt något sätt att rädda spillrorna av argument (b) utan att ta hjälp av argument (c).

7.3 Slutsats

Bör man randomisera i kliniska provningar? Om man vill göra en statistisk hypotesprövning är svaret ja, i enlighet med argument (a). Även om man inte vill göra en hypotesprövning bör man randomisera, om än inte av så starka skäl som många kanske föreställer sig. Jag måste underkänna argument (b), som säger att randomisering skyddar mot obalanser i största allmänhet. Matchning och blockindelning skyddar mycket bättre mot obalanser. Men en slutlig randomisering kan ändå försvaras, genom att den skyddar mot några

¹³Man inte alla. Att vissa människor har svarta strumpor medan andra inte har det kan kanske inte ges en meningsfull biologisk förklaring.

potentiella obalanser som kan uppstå på grund av bristande blindning, det vill säga en variant av argument (c).

Min slutsats gäller bara kliniska prövningar. Som jag nämnde i avsnitt 2 används randomisering även i andra sammanhang än kliniska prövningar, ibland med lite annorlunda innebörd. Därför kan man inte kopiera min slutsats till sådana sammanhang, utan de kräver separata utredningar.

Litteratur

- Armitage, P. (2003). Fisher, Bradford Hill, and randomization, *International Journal of Epidemiology* 32, 925–928.
- Ask, J. (2004). *The Method of Minimization Proposed by Pocock and Simon – Properties with Regards to Balance and Inferential Validity*. Magisteruppsats i statistik, Uppsala universitet, tillgänglig på www.dis.uu.se/Statistik/essays/d/2004_Jacob_Ask.pdf [2008-05-24].
- Cochran, W. G. & G. M. Cox (1957). *Experimental Design*. New York: John Wiley & Sons (2nd edition).
- Fisher, R. A. (1926). The arrangement of field experiments, *Journal of the Ministry of Agriculture of Great Britain* 33, 503–513.
- Fisher, R. A. (1951). *The Design of Experiments*. Edinburgh: Oliver & Boyd (6th edition).
- Hacking, I. (1988). Telepathy: Origins of randomization in experimental design, *Isis* 79(3), 427–451.
- Hansson, S. O. (2005). *Decision Theory: A Brief Introduction*. Opublicerat manuskript, tillgängligt på www.infra.kth.se/~soh/decisiontheory.pdf [2008-05-24].
- Howson, C. & Urbach, P. M. (2005). *Scientific Reasoning: The Bayesian Approach*. Chicago: Open Court (3rd edition).
- Jadad, A. (2000). *Randomiserade kontrollerade kliniska prövningar*. Lund: Studentlitteratur.
- Lindley, D. V. (2000). The philosophy of statistics [with discussion], *The Statistician* 49(3), 293–337.
- Ostle, B. (1954). *Statistics in Research: Basic Concepts and Techniques for Research Workers*. Ames, Iowa: Iowa State College Press.
- Papineau, D. (1985). Probabilities and causes, *Journal of Philosophy* 82(2), 57–74.
- Papineau, D. (1994). The virtues of randomization, *British Journal for the Philosophy of Science* 45(2), 437–450.
- Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization, *Annals of Statistics* 6(1), 34–58.
- Savage, L. J.; Barnard, G. A. & Cox, D. R. (red.) (1962). *The Foundations of Statistical Inference*. New York: John Wiley & Sons.
- Taube, A. (2000). Om randomiseringens välsignelser, *Läkartidningen* 97(38), 4173–4174.
- Urbach, P. (1985). Randomization and the design of experiments, *Philosophy of Science* 52(2), 256–273.
- Worrall, J. (2002). What evidence in evidence-based medicine? *Philosophy of Science* 69(3), Supplement, S316–S330.

- Worrall, J. (2007). Why there's no cause to randomize, *British Journal for the Philosophy of Science* 58, 451–488.
- Yates, F. (1939). The comparative advantages of systematic and randomized arrangements in the design of agricultural and biological experiments, *Biometrika* 30(3-4), 440–466.