

Upp till bevis – en ESO-rapport om experiment som underlag för politik

Pathric Hägglund och Per Johansson

*Rapport till
Expertgruppen för studier i offentlig ekonomi
2021:4*



STATENS OFFENTLIGA
UTREDNINGAR

Rapportserien finns på eso.expertgrupp.se.

Omslag: Elanders Sverige AB

Tryck: Elanders Sverige AB, Stockholm 2021

ISBN 978-91-525-0213-6 (tryck)

ISBN 978-91-525-0214-3 (pdf)

Förord

Beslutsunderlag av god kvalitet är en förutsättning för välgrundade beslut. Det gäller inom politiken såväl som inom andra områden. Beslutsunderlag av god kvalitet ger kunskap och insikter om vilka åtgärder som är mest ändamålsenliga för att uppnå ett visst politiskt mål. För att minska risken att politiskt beslutsfattande bortser från existerande kunskap eller kunskap som skulle kunna genereras genom forskning framhålls ofta krav på att politiken ska vara evidensbaserad. Det innebär att forskningsrön och expertkunskaper ska vara vägledande för de politiska besluten. Men vilka vetenskapliga metoder finns egentligen som kan sägas ge evidens för att en viss åtgärd eller en viss förd politik ger bättre måluppfyllelse än möjliga alternativ?

Randomiserade experiment har länge framhållits som det bästa sättet att finna vetenskaplig evidens kring effekterna av en viss åtgärd. Det randomiserade experimentet har dock en del begränsningar och metodologiska alternativ är så kallade naturliga experiment, eller kvasiexperiment.

I den här rapporten till ESO diskuterar Pathric Hägglund och Per Johansson experimentets roll som underlag för utformning av politiska beslut. I rapporten redovisas för- och nackdelar med det randomiserade experimentet och kvaliteter och brister i genomförda svenska fältexperiment diskuteras. Författarna resonerar också om vilka förutsättningar och hinder som finns för en evidensbaserad politik.

Johansson och Hägglund konstaterar i rapporten att en evidensbaserad politik förutsätter att det finns forskare och utredare med god metodkompetens. Men det förutsätter också att mottagarsidan, som utgörs av dem som ytterst bereder de politiska besluten, är tillräckligt insatt för att förstå och kunna värdera olika studiers resultat. Jag hoppas att den här rapporten ska kunna bidra i det här

avseendet genom att öka kunskapen om experimentella metoder för att analysera effekter av politiska reformer och därmed också öka förutsättningarna för att beslutsfattande sker utifrån vetenskaplig evidens.

Arbetet med rapporten har följts av en referensgrupp bestående av personer med god insikt i dessa frågor. Gruppen har letts av Peter Thoursie, ledamot i ESO:s styrelse. Som alltid i ESO-sammanhang svarar författarna själva för innehåll, slutsatser och förslag i rapporten.

Stockholm i september 2021

Karolina Ekholm
Ordförande i ESO

Innehåll

Sammanfattning	7
Summary	13
1 Inledning.....	19
2 Vad menar vi med evidensbaserad politik?	21
3 Kausal analys via experiment eller observation	29
3.1 Experiment	30
3.2 Observationsstudier och kvasiexperiment	39
3.3 Incitament och statistiska problem vid empiriska studier	55
4 Förekomsten av experiment inom medicin och ekonomi	59
4.1 Randomiserade experiment	60
4.2 Kvasiexperiment.....	63
4.3 Sökning på experiment inom socialförsäkring och arbetsmarknadspolitik	67
4.4 Summering.....	68
5 Experiment inom den svenska socialförsäkringen och arbetsmarknadspolitiken	71
5.1 Experiment inom socialförsäkringen.....	71

5.2	Experiment inom arbetsmarknadspolitiken.....	82
5.3	Diskussion och summering.....	85
6	Statistik, kausalitet och evidens inom olika utbildningar	93
6.1	Nationalekonomi, statskunskap, sociologi, företagsekonomi och journalistik.....	95
6.2	Medicin.....	98
6.3	Summering	100
7	Hinder för evidensbaserad politik	103
7.1	Kritik av positivism	104
7.2	Forskarens roll.....	113
7.3	Summering	118
8	Slutdiskussion.....	121
	Referenser.....	127
	Bilaga A	139

Sammanfattning

Politik inrymmer målkonflikter om vad som är bra på kort och lång sikt respektive bra och dåligt för olika grupper i samhället. Bedömningar måste göras om kostnader och intäkter för ekonomin som helhet och resultat från olika vetenskapsområden måste vägas samman. Inte sällan åberopar beslutsfattare vetenskapen och vetenskaplig evidens för att motivera ställningstaganden i olika frågor. Någon absolut definition av vetenskaplig evidens finns emellertid inte. I den här rapporten diskuteras begreppet evidens med utgångspunkt i analyser och studier av orsak- och verkanssamband, det vill säga frågeställningar som undersöker effekten av en reform, åtgärd eller en händelse. Vad kännetecknar en studie av hög kvalitet och vilka krav ska man ställa på studier för att kvalificera för underlag till evidens?

En slutsats i rapporten är att en studie som mäter kausalitet på ett framgångsrikt sätt har lyckats isolera betydelsen av den intervention eller behandling som studeras från andra faktorer som är associerade med det aktuella utfallet. Experimentet har länge uppfattats som den bästa designen (*"gold standard"*) vid kausal analys, detta eftersom 1) vi har kännedom om den behandling som ska studeras, det vill säga vi har en tydlig hypotes om vad det är som genererar en eventuell effekt, och 2) tilldelningen av behandling eller inte har bestämts av slumpen, det vill säga inga systematiska skillnader finns mellan de som får respektive inte får behandling med avseende på bakomliggande faktorer som kan störa analysen. Experimentets särställning har emellertid utmanats de senaste decennierna. I takt med den teknologiska utvecklingen, den ökade tillgången på data och inte minst den metodologiska utvecklingen inom framför allt nationalekonomi, har det naturliga experimentet, eller kvasiexperimentet, vuxit fram som ett många gånger bättre alternativ till det klassiska experimentet. Fördelarna med kvasiexperimentet är många, inte

minst blir kostnaden för utvärderingen lägre och ingreppen i eventuellt reguljära verksamheter betydligt mindre, därutöver försvinner problem med etiska aspekter och oönskade beteendeeffekter till följd av att experimentet ibland inte kan blindas, det vill säga genomföras utan deltagarnas vetskap.

Med tiden har också kunskaperna om det randomiserade experimentets begränsningar ökat. I rapporten redovisas genomförda fältexperiment inom social- och arbetslöshetsförsäkringen under senare decennier. Genomgången visar att experimenten visserligen genomgående varit av hög kvalitet, men också haft brister. De har varit framgångsrika i att mäta de totala effekterna av insatserna, det vill säga dem som även beaktar erbjudandet om att delta i insatserna, men sämre på att ge klarhet i vilka delar av insatserna som haft verkan. En slutsats i rapporten är att om randomiserade experiment ska anses motiverade att genomföra vid utvärdering av policy, ska experimenten förmå skapa förståelse för de bakomliggande mekanismerna om framtida liknande åtgärder, och även i andra kontexter.

Förutsättningar för en evidensbaserad politik

Det finns en föreställning bland ekonomer om att randomiserade fältexperiment som forskningsdesign har ökat över tid och att detta gäller speciellt inom utvecklingsekonomi. Denna föreställning förstärktes sannolikt efter Kungliga Vetenskapsakademiens beslut att tilldela 2019 års ekonomipris till Alfred Nobels minne till Abhijit Banerjee, Ester Duflo, och Michael Kremer. Detta för "deras experimentella ansats för att mildra global fattigdom". I den här rapporten redovisar vi i vilken utsträckning randomiserade experiment och kvasiexperiment förekommer bland vetenskapliga publiceringar inom medicin och ekonomi mellan 1989 och 2019. Bilden är tudelad. Antalet publiceringar har ökat tydligt över tid, vilket torde ha en gynnsam inverkan på förutsättningarna för en mer evidensbaserad politik, inte minst eftersom det signalerar en ökad metodkompetens i forskarsamhället. Emellertid är studierna med en experimentell ansats inte bara påfallande lågt citerade, de utgör fortfarande en väldigt liten andel av samtliga studier som publiceras. Detta hänger delvis samman med en kraftig ökning av det totala

antalet publicerade studier. Med en allmän ökning av vetenskapliga alster ökar svårigheten att skilja ”agnarna från vetet”.

För att skilja ut bra studier med goda förutsättningar att besvara frågeställningar av stort policyvärde, från mindre bra studier, ställs krav på metodkompetens på mottagarsidan. Kunskaper i statistik i allmänhet och om kausala strategier i synnerhet utvecklar förmågan till kritiskt tänkande kring frågor om kontrafaktiska utfall och orsaks- och verkanssamband, kort sagt kring frågor kopplade till evidens. I rapporten genomförs en kartläggning av kunskapskraven på olika utbildningar, främst inom de samhällsvetenskapliga ämnena men också inom medicin. Genomgången visar att grundutbildningarna i ämnen som nationalekonomi, statsvetenskap, sociologi, företagsekonomi och journalistik i regel innehåller omkring en halv termin kurser i grundläggande statistik, och att det är svårt att tro att studenter som endast läser dessa kurser kan utveckla ett kritiskt tänkande kring sambandsanalyser med regressionsanalys. Vid genomgång av master- och forskarutbildningar kan konstateras att variationen är stor. I vissa fall baseras de på studenternas egna val när det gäller valbara kurser och val av avhandlingsämne. Detta gäller även inom de medicinska vetenskaperna. Nationalekonomi sticker ut som ett ämne där studenterna inte bara ges teoretiska utan också praktiska kunskaper om de kvasiexperimentella ansatserna, och detta redan på mastersnivån. En sammanfattande slutsats är att det finns en betydande förbättringspotential för experimentell metodik inom de flesta discipliner.

Hinder för en evidensbaserad politik

Avslutningsvis diskuteras i rapporten faktorer som hindrar evidens från att få genomslag i politiken. Vi lyfter särskilt fram det växande intresset för maktordningsteorier med utgångspunkt i kritisk teori (”Critical theory”) som ett sätt att förstå samhället och utforma politiken. Gemensamt för förespråkarna för dessa maktordningsteorier är skepsisen inför den vetenskapliga metoden och att beakta olika förklaringsmodeller bakom ett skeende. Att på förhand låsa sig vid en förklaringsmodell utan att undersöka det empiriska stödet för den valda modellen eller för alternativa förklaringsmodeller är motsatsen till evidensbaserad politik. I rapporten exemplifierar vi

inslaget av kritisk teori med svensk jämställdhetspolitik. Skrivningar på jämställdhetsområdet ger tydliga indikationer om att regeringen i viss mån låst sig vid en teori för att förstå observerade könsskillnader. Detta riskerar paradoxalt nog att få negativa konsekvenser för jämställdheten, definierat som mäns och kvinnors lika villkor, förutsättningar och möjligheter.

Tre förslag för en evidensbaserad politik

Fokus i våra tre förslag är att skapa bättre förutsättningar för en evidensbaserad politik. De handlar om att stärka den metodologiska undervisningen vid högskolor och universitet, framför allt inom de samhällsvetenskapliga ämnena som är vanligt förekommande inom det politiska beslutsfattandet, samt satsningar på analys- och utvärderingsmyndigheter med tydliga uppdrag att genomföra kausala studier med hög kvalitet. Sist men inte minst uppmanas till ifrågasättande av akademisk auktoritet för att nå en evidensbaserad politik.

Stärk den metodologiska undervisningen inom samhällsvetenskapliga ämnen

En evidensbaserad politik ställer krav på kritiskt tänkande och kunskaper inom statistik och vetenskaplig metod. Vår genomgång visar att grundutbildningarna (kandidatexamen) inom de samhällsvetenskapliga ämnena som ofta förekommer inom det politiska beslutsfattandet i regel innehåller en halv termin i ämnet statistik. Mycket lite tyder på att studenterna får en förståelse för skillnaden mellan korrelationer och kausalitet. Endast i undantagsfall nämns begrepp som orsak, korrelation och experiment i kursplanerna. Vidare kan konstateras att de kunskaper som lärs ut om kausala ansatser på master- och forskarnivå, i bästa fall, är att betrakta som yttliga och översiktliga.

Utan att göra anspråk på att ge en komplett bild visar vår genomgång att kraven på studenters förståelse för principerna för hur kausalitet ska förstås och mätas överlag är låga. Vi menar att detta är central kunskap för en evidensbaserad politik och att en översyn av målen med de grundläggande kurserna i statistik är nödvändig. Redan i detta skede bör studenterna introduceras för experiment och

tänkande kring kontrafaktiska tillstånd samt ges möjlighet att förstå de begränsningar som finns med användandet av icke-experimentella data och traditionella regressionsansatser vid kausala utsagor.

Ifrågasätt forskare och forskning

Kartläggningar av politisk hemvist bland forskare har visat att tydliga åsiktskluster kan förekomma inom discipliner, inte sällan till vänster på den politiska högervänster-skalan. Forskares ideologiska hemvist kan dölja sig bakom de frågeställningar som behandlas, de slutsatser som dras och de resultat man framhäver som viktiga – i synnerhet om det finns politiska implikationer av det som studeras. Vi menar att ifrågasättande av akademisk auktoritet är nödvändigt för en evidensbaserad politik.

För att undvika risken för att inomdisciplinära dogmer får genomslag i politiken och, tvärtom, säkerställa att beslutsfattare ges en relevant och rättvis bild av bevisläget inom ett visst område, kan det vara en idé att använda sig av expertis utanför det aktuella forskningsområdet för att granska forskningsläget. Denna expertis ska bestå av oberoende forskare med god metodkompetens.

Fler analysmyndigheter med tydliga uppdrag behövs

En evidensbaserad politik nås inte genom finansiering av enskilda randomiserade experiment utan genom en långsiktig strategi för kunskapsuppbyggnad inom olika ämnesområden och sektorer, exempelvis genom att bygga upp forskningsmiljöer med fokus på empirisk analys. Ett bra sådant exempel är bildandet av Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU) i Uppsala 1997. Nästan alla på IFAU är disputerade forskare och institutets verksamhet är av hög kvalitet.¹

Liksom tidigare konstaterats i en översyn av regeringens analys- och utvärderingsresurser (SOU 2018:79) är vår bedömning att analys- och utvärderingsmyndigheter, genom en kombination av oberoende sak- och metodkompetens, är särskilt väl lämpade för långsiktig kunskapsuppbyggnad. Vi anser därför att fler myndig-

¹ Se Hjalmarsson med flera (2004).

heter med tydliga uppdrag att utvärdera effekterna av, och kostnads-effektiviteten i, olika satsningar, skulle öka förutsättningarna för en väl underbyggd och evidensbaserad politik. Utmaningen ligger i att bemanna verksamheterna med tillräcklig kompetens inom kausal design och analys. Kompetenta forskare kan arbeta med en mängd olika frågeställningar förutsatt att de får samarbeta med personer som besitter nödvändig sakkunskap. Av denna anledning kan det finnas samordningsvinster med att skapa ett antal större utvärderingsmyndigheter. Utvärderingsmyndigheterna bör därutöver placeras i en bredare forskningsmiljö eftersom avgörande för att kunna leverera ett högkvalitativt kunskapsunderlag för politiken är möjligheten att rekrytera personal med rätt kompetens.

Summary

Due to budget restrictions, political decisions often involve conflicts between short and long-term goals, and between different groups in society. Optimally, a decision as to whether or not to adopt a public policy should be preceded by an assessment of all potential revenues from the policy against all potential costs. Decision-makers often refer to scientific evidence as a reason for adopting policies, but these statements often serve as justifications of proposals rather than careful assessments of the actual state of knowledge.

The aim of the report is to evaluate the possibilities of increasing the use of evidence-based policy in Sweden. To this end, we first define evidence as the summary of results of high-quality studies with a causal design. We then define what characterises a high-quality study and what requirements must be met for it to qualify as a study with a causal design.

A causal design states how the counterfactual outcome of being treated (the potential outcome if the individual had not been treated) by a policy is identified in data. A well-conducted experiment in which treatment and control (non-treatment) are randomly allocated allows us to identify the counterfactual outcome of being treated, by using the outcomes of the individuals not treated, or controls.

The randomised experiment has long been perceived as the best design ("gold standard") in causal analysis. However, the special standing of the randomised experiment has been challenged in recent decades. In pace with technological development, the increased availability of data and, not least, methodological development, especially in economics, the natural experiment, or quasi-experiment, has emerged as an alternative to the classical randomised experiment. Quasi-experiments have many advantages; they are not as expensive to perform and the ethical aspects are less of a problem.

In addition, undesirable behavioural effects are not a problem in non-blinded randomised experiments.

The report provides a summary of the results of Swedish randomised field experiments carried out in social insurance in recent decades. The review shows that while all the experiments were of high quality, they did also have shortcomings. They have been successful in measuring the overall effects of the interventions, i.e. those that also take into account the offer to participate in the interventions, but, in general, they do not provide an understanding of the mechanisms behind the results. We conclude that if randomised experiments are to be considered when evaluating policy, they should be able to create an understanding of the underlying mechanisms as this helps in drawing conclusions on similar policies and in other contexts.

Conditions for an evidence-based policy

There is a perception among economists that the use of randomised field experiments has increased over time and that this is especially so in development economics. This perception was probably reinforced after the decision of the Royal Swedish Academy of Sciences to award the 2019 Economics Prize to Abhijit Banerjee, Ester Duflo and Michael Kremer for “their experimental approach to alleviating global poverty”. This report sets out the extent to which randomised experiments and quasi-experiments occur in scientific publications in medicine and economics between 1999 and 2019. The picture is twofold. The number of publications has increased clearly over time, which should have a favourable impact on the conditions for a more evidence-based policy, not least because it signals increased methodological competence in the research community. However, not only are the studies with an experimental approach remarkably poorly cited, they still make up a very small proportion of all published studies. This is partly due to a sharp increase in the total number of published studies. With a general increase in scientific production, it becomes more difficult to sift “the chaff from the wheat”.

To better integrate the consideration of evidence from high-quality studies in political decisions, analysts in the central government administration need a minimum level of methodological competence. Knowledge of statistics in general and of causal strategies in particular develop the ability to think critically. To be critical to empirical studies, you need to have an understanding of how the researcher's design is solving the problem of validly estimating the counterfactual outcomes to the ones treated. This study surveyed study directors at Swedish universities about the requirements for various study programmes, primarily in the social science subjects which are the field of the majority of people working for the central government administration in Sweden. As a comparison, the survey was also sent to a study director in medicine. The review shows that undergraduate programmes in subjects such as economics, political science, sociology, business administration and journalism usually contain about half a term of courses in basic statistics. It is hard to believe that students who only take these courses can develop critical thinking about empirical analysis. We find large variation in requirements in master's and doctoral programmes. In some cases, they are based on the students' own choices regarding elective courses and the dissertation subject chosen. This also applies to the medical sciences. Economics stands out as a subject where students are given not only theoretical but also practical knowledge of quasi-experimental approaches, and this is already done at the master's level. A conclusion is that there is a significant potential for improvements in methodology in most disciplines.

Obstacles to an evidence-based policy

Finally, the report discusses factors that prevent evidence from having an impact on policy. We particularly highlight the growing interest in theories based on Critical Theory as a way of understanding society and shaping politics. Advocates of these theories share scepticism about the scientific method and the consideration of alternative explanatory models behind an event. The report uses Swedish gender equality policy to exemplify the element of Critical Theory. Writings in official governmental

documents on gender equality give clear indications that the Government has been adopting this non-scientific theory in its formation of gender equality policy. Paradoxically, this risks having negative consequences for gender equality, defined as equal conditions and opportunities for men and women.

Three proposals for evidence-based policy

Our three proposals focus on creating better conditions for evidence-based policy. They are about strengthening methodological teaching at universities and other higher education institutions, especially in the social sciences, as well as about investments in analysis and evaluation agencies with responsibility for causal analysis of high quality. Last but not least, we call for the questioning of academic authority.

Strengthen methodological teaching in social science subjects

An evidence-based policy requires critical thinking and knowledge of statistics and scientific methods. Our review shows that the undergraduate programmes (bachelor's degrees) in the social science subjects that often figure in political decision-making usually contain only half a term in the subject of statistics. There is no support that students gain an understanding of the difference between correlation and causality. Only in exceptional cases are concepts such as cause, correlation and experiments mentioned in the syllabus. Furthermore, the material taught on causal approaches at masters and doctoral level is, at best, to be regarded as superficial.

Without claiming to give a complete picture, our review shows that the requirements for the students in understanding the principles of causal thinking and design for credible analysis are low. We believe that this is central to an evidence-based policy, and that a review of the objectives of the basic courses in statistics is necessary. Students should already be introduced at this stage to experiments and thinking about counterfactual outcomes and be given the opportunity to understand the limitations that exist in the use of non-experimental data and traditional regression approaches in causal statements.

Question researchers and research

Surveys of political preferences among researchers have shown that there can be clear clusters of opinion within disciplines, not infrequently on the left of the political right-left scale. The ideological preferences of researchers may underlie the issues addressed, the conclusions drawn and the results highlighted as important, especially if there are political implications of what is being studied. We believe that questioning academic authority is necessary for an evidence-based policy.

In order to avoid the risk that interdisciplinary dogmas will have an impact on policy and to ensure, on the contrary, that decision-makers are given a relevant and fair picture of the state of evidence in a particular area, it may be worth using expertise outside the particular research area to examine the state of research. This expertise must consist of independent researchers with good methodological competence.

More analysis authorities with clear assignments needed

An evidence-based policy is achieved not by funding individual randomised experiments but by having a long-term strategy for knowledge building in different subject areas and sectors, for example by building research environments with a focus on empirical analysis. One good example of this is the establishment of the Institute for Labour Market and Education Policy Evaluation (IFAU) in Uppsala, which was formed in 1997. Almost everyone at IFAU has a PhD and the Institute's activities are of high quality.

Our assessment is that more agencies with clear assignments to evaluate the effects, and cost-effectiveness, of various initiatives would increase the conditions for evidence-based policy. The challenge is to staff the agencies with sufficient competence in causal design and analysis. Competent researchers can work on a variety of issues, provided that they are allowed to collaborate with staff with the necessary subject-based expertise. For this reason, there may be coordination gains in creating a number of major evaluation agencies as also previously suggested in a review of the Government's analysis and evaluation resources (SOU 2018: 79). In addition, the evaluation agencies should be placed in a broader research environment, since

the ability to recruit staff with the right skills is crucial for the ability to deliver a high-quality knowledge base for policy.

1 Inledning

Inte sällan åberopar beslutsfattare vetenskapen och vetenskaplig evidens för att motivera ställningstaganden i olika frågor. Men frågan är vad vetenskaplig evidens är. En enkel definition som de flesta nog skulle ställa sig bakom är att det är en utsaga som bygger på data och empiri snarare än på teori och ideologi. Men vilka krav ska man ställa på studier för att de ska kvalificera sig som underlag till evidens? I denna rapport diskuterar vi begreppet evidens med utgångspunkt i analyser och studier av orsak- och verkanssamband, det vill säga frågeställningar som undersöker effekten av en reform, åtgärd eller en händelse. Exempel på sådana frågeställningar kan vara hur fler lärare och mindre klasser inverkar på elevernas prestationer, om en satsning på rehabiliteringsinsatser ger avkastning i form av ökat arbetslivsdeltagande och om straffskärpningar leder till lägre brottslighet.

Inom den vetenskapliga litteraturen har det randomiserade experimentet länge haft en särställning som metod för att identifiera orsak- och verkanssamband, det vill säga kausala samband. Frågan är om denna särställning är relevant i dag, mot bakgrund av de senaste decenniernas stora framsteg inom metodologi, teknologi och tillgång till data. Man kan också invända att kunskaperna om experimentets inneboende brister har ökat med tiden, samt att många intressanta frågeställningar inte låter sig analyseras med hjälp av experiment – inte minst av etiska skäl.

I rapporten diskuteras för- och nackdelar med experiment som underlag för politiskt beslutsfattande, hur vanliga de är i förhållande till andra metoder att mäta kausalitet och i vilka situationer experiment (trots sina brister) är motiverade att genomföra. I rapporten ges också en introduktion till det som utgör ett alternativ till det randomiserade experimentet, de så kallade kvasiexperimentella ansatserna, i vilka sammanhang de är tillämpliga och vilka

antaganden som måste vara uppfyllda för att resultaten ska kunna tolkas kausalt. Syftet med rapporten är att öka kunskapen kring hur vi bör se på vetenskaplig evidens och lämna förslag på hur politiken i högre utsträckning ska ges förutsättningar att vara evidensbaserad.

Rapporten har följande struktur, i avsnitt 2 redovisas de vetenskapliga utgångspunkter rapporten utgår ifrån och de villkor som måste vara uppfyllda för att experiment och andra statistiska analyser ska kunna utgöra underlag för politiskt beslutfattande. Avsnitt 3 ger en grundläggande diskussion av fördelar och nackdelar med kausala designer, både vid experiment- och observationsdata. I avsnitt 4 redovisas resultatet från litteratursökningar av hur vanligt förekommande randomiserade experiment och kvasiexperiment är i förhållande till andra studier inom den medicinska och samhällsvetenskapliga forskningslitteraturen. I avsnitt 5 ges en kortfattad presentation av genomförda experiment inom arbetsmarknadspolitik och socialförsäkringar i Sverige, och också en diskussion av i vad mån de lyckats undvika de fallgropar som är förknippade med experiment. Avsnitt 6 redovisar förekomsten av ämnet statistik inom utbildningar som är vanligt förekommande i beslutsfattandet, det vill säga i ämnen som exempelvis statskunskap, nationalekonomi och sociologi. I avsnitt 7 diskuteras avslutningsvis potentiella hot mot en evidensbaserad politik. I avsnitt 8 ges en kort summering av rapporten och även våra rekommendationer för en mer evidensbaserad politik.

2 Vad menar vi med evidensbaserad politik?

För gemene man skulle nog verklighetsbeskrivningar baserade på historiska händelser och data betraktas som evidens. Detta till skillnad från teorier, ideologier och resultat från modellsimuleringar.² Denna rapport behandlar evidens av policyåtgärder eller interventioner.

En policyåtgärd eller intervention innebär en förändring av den förda politiken, exempelvis i form av en ny insats inom arbetsmarknadspolitiken eller inom hälso- och sjukvården. Beslutet om en intervention ska helst ha föregåtts av jämförelser av kostnader och effekter mellan olika alternativ. Eftersom effekter av en insats inte exakt kan förutses kan man hämta stöd i befintlig evidens. Evidens i detta sammanhang berör således kontrafaktiska frågor om orsak och verkan eller kausala samband. Eftersom evidensen utgör en grund för aktiva beslut är det mycket viktigt att de studier som utgör evidens till grund för policybeslut är av hög kvalitet.

En högkvalitativ studie av kausala samband ska innehålla en design som tydliggör skillnaden mellan korrelation och kausalitet. Kausalitet handlar om hur den ena variabeln, eller storheten, påverkar den andra. En korrelation är ett statistiskt begrepp som beskriver hur variabler samvarierar. En positiv eller negativ samvariation, eller korrelation, kan antingen vara kausalt betingad eller inte. En samvariation som inte är kausal benämns ibland ”falsk korrelation”. Denna benämning är språkligt märklig då en korrelation ju bara är en beskrivning av data. Möjligen speglar den vår

² Skattade samband baserade på historiska data via regressionsmodeller, eller modeller baserade på maskininlärningsalgoritmer, används ofta till prediktioner. Om prediktionerna överensstämmer med historiska data kan vi tro att modellen säger något om en framtid, men endast under förutsättning att världen eller omgivningen är stabil. I dessa sammanhang (se t.ex. Barda m.fl. (2020) vad gäller covid-19) benämns de verkliga utfallen i form av historiska data för evidens.

undermedvetna vilja att tolka världen i orsaks- och verkansamband. Olika intressegrupper kan också tänkas ha motiv till att övertolka korrelationer i termer av kausalitet.³ Emellertid säger endast kausala samband något om hur det ena påverkar det andra. Det är bara ett kausalt samband som är intressant för att avgöra om en behandling ska anses motiverad eller inte.

En studie som mäter kausalitet har på ett framgångsrikt sätt lyckats isolera betydelsen av den intervention eller behandling som studeras från andra faktorer som är associerade med det aktuella utfallet. Det randomiserade experimentet har länge uppfattats som den bästa designen (*"gold standard"*) vid kausal analys, detta eftersom tilldelningen av behandling eller inte har bestämts av slumpen, det vill säga inga systematiska skillnader finns mellan de som får respektive inte får behandling med avseende på bakomliggande faktorer som kan störa analysen. Experimentets särställning har emellertid utmanats de senaste decennierna av nya designer för att mäta kausalitet. I dessa utnyttjas på naturlig väg uppkommen slumpmässighet i tilldelningen av behandling samt detaljerade data. Dessa kvasiexperimentella designer beskrivs mer ingående i nästa kapitel. Att studier som gör anspråk på att analysera kausalitet har en design för att bryta associationen mellan det utfall som ska analyseras och egenskaper hos behandlade och icke-behandlade, är ett avgörande kriterium för vad som är att betrakta en studie av hög kvalitet.

Noterbart är att nämnda definition av en studie av hög kvalitet är deduktiv, det vill säga vi har på förhand kännedom om vad som potentiellt orsakar skillnader i utfall, med andra ord den behandling som leder till en effekt. Ofta är vi dock intresserade av det omvända, det vill säga att förstå vad det är som orsakar det vi observerar. Exempelvis vill vi gärna veta orsaken till att vi har huvudvärk eller till att det blir varmare på jorden. Att identifiera en orsak till ett utfall är dock ett mycket svårt problem (ett induktivt problem) som utgör grunden i Karl Poppers (en av vetenskapsteoriens mest kända filosofer) falsifieringsprincip. Denna innebär att forskaren ska sträva efter att kritiskt finna motbevis till sin teori.

³ Problemet illustras väl av citatet "Lögn, förbannad lögn och statistik" som används för att beskriva hur statistik används för att stärka svaga argument (Mark Twain anges ofta som källa för uttrycket men han populariserade endast Benjamin Disraelis uttryck).

Ett faktum är emellertid att systematiskt insamlade historiska data, utan deduktiva analyser, har lett fram till mycket starka slutsatser kring orsaks- och verkanssamband. För att ta ett högst aktuellt exempel – att tvätta händerna för att minska risken att bli sjuk – föreslogs av både Ignaz Semmelweis och Florence Nightingale. Ett annat exempel är Charles Darwins empiriska analyser av evolutionsteorin. Teorin förutsäger likartade utvecklingsmönster bland samtliga arter. Om detta mönster inte observerats hade teorin förkastats.⁴ Hypoteser genererade via induktion har oftast senare kunnat validerats deduktivt. På detta sätt råder ett samspel mellan den induktiva och deduktiva ansatsen. Rådet att vara noga med handhygien under rådande covid-19-pandemi har sitt ursprung i induktion men har funnit stöd i den deduktiva ansatsen.

Således, den strikt vetenskapligt deduktiva ansatsen är inte den enda möjliga ansatsen vid studier av orsak- och verkanssamband. För många viktiga kausala frågeställningar är det svårt, för att inte säga omöjligt, att använda den deduktiva ansatsen. I sådana situationer kan det bli nödvändigt att använda sig av teori och analyser och resonemang baserade på analys av historiska data eller dokument för att få en uppfattning om aktuellt kunskapsläge. För att sådana studier ska betraktas som trovärdiga gäller att teorierna ska vara allmängiltiga och de empiriska studierna baserade på kvalitativa eller kvantitativa data ska kunna granskas av utomstående forskare, det vill säga studien ska kunna replikeras. För att ge ett exempel, det finns ingen deduktiv evidens beträffande konsekvenserna av den av människan skapade ökningen av koldioxid på klimatet. Eftersom frågan är viktig används bland annat induktiva modeller för analys av historiska data. Vi har ingen möjlighet att bedöma transparensen i denna forskning. Vi kan dock konstatera att det förekommer en mycket aktiv och kritisk diskussion mellan forskare om val av metod och data, vilket är ett sundhetstecken.

Ett annat förtydligande är att förutsägelser om framtiden baserade på historiska data och teorilösa regressionsanalyser eller analyser via maskinlärningsalgoritmer, exempelvis över klimatförändringar eller antalet döda i covid-19, inte är att betrakta som evidens när det gäller orsak- och verkanssamband. Även om tillförlitligheten i dessa modeller i regel har prövats mot historiska data

⁴ Noterbart är att ny forskning visat att det grundläggande antagandet om slumpmässig förändring av arvsanlag inte kan vara korrekt (Berlinski 2009).

och modellerna kan vara framgångsrika i att förutse utvecklingen under en kortare period, är de inte anpassade för utsagor om konsekvenserna av en genomförd policyåtgärd, som exempelvis krav på munskydd eller nedstängning av ekonomin.

Båda den induktiva och deduktiva metodens vetenskapliga utgångspunkt är att ingen teori går att bevisa. Frågan blir således vad som kan betecknas som evidens. Av samhällsdebatten att döma är det lätt att få uppfattningen att det är detsamma som vetenskaplig konsensus. Ett exempel är klimatområdet där det ibland framhålls att 97-98 procent av alla klimatforskare ställer sig bakom påståendet att mänsklig aktivitet är en viktig förklaring till den globala uppvärmningen. Påståendet har bland annat förts fram av USA:s förre president Obama samt USA:s federala myndighet för luft- och rymdfart, NASA.⁵ Bakgrunden är publicerade studier av bland annat Oreskes (2004), Anderegg m.fl. (2010), Doran och Zimmerman (2011) och Cook m.fl. (2013).

Vetenskaplig konsensus är en oxymoron, det vill säga en själv-motsägelse, eftersom konsensus inte är ett vetenskapligt argument. Evidens är bara en bästa beskrivning av det nuvarande kunskapsläget. Denna beskrivning måste göras genom en sammanställning av litteraturen inom respektive område. För att undvika skevheter i urval av studier ska denna optimalt sätt göras systematiskt, det vill säga att ha tydliga kriterier för inklusion.⁶ Det empiriska underlaget i de ovan nämnda uppsatserna utgörs inte av sammanställningar av resultat från välgjorda empiriska studier som försökt besvara den specifika frågeställningen, utan antingen av svar från forskares uppfattning via enkäter med betydande bortfall eller via granskningar av så kallade "abstracts", det vill säga uppsatssammanfattningar. Slutsatserna i dessa studier har också kritiserats för att vara alltför långtgående utifrån de empiriska underlagen, och i vissa fall

⁵ Barack Obama på Twitter 16 maj 2013: "Ninety-seven percent of scientists agree: climate change is real, man-made and dangerous." (<https://twitter.com/BarackObama/status/335089477296988160?s=20>) På NASA:s hemsida (Scientific Consensus | Facts – Climate Change: Vital Signs of the Planet (nasa.gov), går att läsa: "Multiple studies published in peer-reviewed scientific journals show that 97 percent or more of actively publishing climate scientists agree: Climate-warming trends over the past century are extremely likely due to human activities."

⁶ Statens beredning för medicinsk och social utvärdering (SBU) genomför systematiska sammanställningar av metoder och insatser inom hälso- och sjukvård men även insatser inom socialtjänsten och området funktionstillstånd/-hinder. Campbell collaboration och Cochrane library innehåller systematiska genomgångar inom samhällsvetenskap och hälso- och sjukvård (se <https://www.campbellcollaboration.org/> och <https://www.cochranelibrary.com/about/about-cochrane-library>).

direkt felaktiga.⁷ Andra studier, till exempel Bray och von Storch (2010) samt Strengers m.fl. (2015), ger dessutom en annan bild av forskares inställning i frågan. Sammanfattningsvis kan man utifrån dessa studier inte dra slutsatser om vad klimatforskare anser i frågan.

För att komma närmare en förståelse för vad som kan betraktas som en sanning om världen i termer av orsak och verkan måste det finnas utrymme för ifrågasättande av rådande paradigm, detta via argument och kritik av vetenskaplig metod. Kritiker av ett paradigm är av essentiell betydelse i all vetenskap eftersom det skärper både argument och metoder bland dem som företräder paradigm. Är inte argument och metod av hög kvalitet faller med stor sannolikhet paradigm, ett exempel på detta är Newtons gravitationsteori som ersatts av Einsteins relativitetsteori.⁸ För att återknyta till klimatfrågan kan noteras att de klimatforskare som är skeptiska till det rådande paradigm om koldioxidens skadliga effekter ofta miss-tänkliggörs och skambeläggs, detta i stället för att få sina argument bemötta.⁹

Avslutningsvis, politik handlar om skeenden och förändringar på samhällsnivå samt om helhet och långsiktighet. Ofta finns målkonflikter mellan vad som är bra på kort och lång sikt respektive bra och dåligt för olika grupper. Det betyder att beslutsfattare måste göra bedömningar om kostnader och intäkter för hela ekonomin och väga samman resultat från olika vetenskapsområden. En insats i syfte att förbättra hälsan i befolkningen måste exempelvis vägas mot kostnaden. Ett intressant exempel är antologin med Björn Lomborg som redaktör (Lomborg 2018). Boken är en kritisk analys av FN:s agenda för en hållbar utveckling (även känd som de globala målen). Agendan innehåller sammantaget 169 olika mål och täcker alla områden av internationell utveckling, från hälsa och utbildning till sanitet och konflikt. Att långsiktigt eliminera fattigdom är en komplex uppgift och med en begränsad budget är en befogad fråga om alla mål är lika viktiga, eller om man genom att fokusera på vissa mål kan åstadkomma mer nytta inom ramen för budgeten. Frågan kan besvaras genom en kostnads-effektanalys. Utifrån en sådan identifierar en panel av meriterade ekonomer satsningar inom 19 av de 169 utvecklingsmålen som skulle uppnå fyra gånger så stor global

⁷ Se till exempel Legates m.fl. (2015) kritik av slutsatserna i Cook m.fl. (2013).

⁸ Noterbart är att det finns även kritiker av relativitetsteorin.

⁹ För ett av många exempel i svensk media <https://www.expressen.se/premium/nyheter/de-ar-svenskarna-som-ifragasatter-klimatkrisen/>

hälsoförbättring än vad som skulle uppnås inom den budget som FN tänker sig för att finansiera samtliga 169 mål.

Ett annat exempel är den rådande covid-19-pandemin där europeiska länder och vissa amerikanska delstater vidtagit drastiska åtgärder genom att begränsa individers rörelsefrihet. Med tanke på att dödligheten i covid-19 i det autokratiskt styrda Kina har hållits på en mycket låg nivå efter de drastiska restriktionerna i landet är det inte otänkbart att samma åtgärd skulle ge positiva effekter även i andra länder. Osäkerhet finns emellertid såtillvida att erfarenheter av nedstängningar av demokratiska stater saknas. Vad vi däremot vet utifrån forskning är att en nedstängning skulle innebära lägre inkomster och en ökad arbetslöshet. Vidare finns stark evidens för att ökad arbetslöshet leder till ökad dödlighet (Roelfs m.fl. 2011). På svenska förhållanden konstaterar SOU (2010:102) att arbetslöshet ökar risken för dödsfall med cirka 10 procent. Därmed får evidensen för ökad dödlighet vid nedstängning på grund av arbetslöshet anses starkare än evidensen för minskad dödlighet på grund av minskad smittspridning.¹⁰

Lärdomen från exemplet om de globala målen är att eftersom det finns budgetrestriktioner är en prioritering nödvändig. En sådan prioritering utan politisk hänsyn kan göras baserat på evidens. Lärdomen från båda exemplen är att experter inom ett område varken kan eller ska behöva reflektera över de totala kostnaderna för en insats.

Sammanfattningsvis, att etablera evidens om orsaks- och verkansamband är dels komplicerat, dels en process. Existerande evidens måste utmanas genom användning av nya data och nya metoder. Målkonflikter mellan olika politikområden är oundvikliga, en evidensbaserad politik måste bygga på överväganden mellan kostnader och intäkter på samhällsnivå. I detta arbete ska inte enskilda experter ha sista ordet, besluten måste i stället fattas av beslutsfattare

¹⁰ Med detta vill vi inte säga att en nedstängning av ekonomin var obefogad. I ett initialt skede var osäkerheten om hur hög dödligheten var mycket stor och tidiga estimat från Wuhan om dödlighet bland de som registrerats med smitta var över 20 procent (Roser m.fl. 2020). Simuleringar via prediktionsmodeller antydde också en mycket stor överdödlighet till följd av covid-19 (Banjeree m.fl. 2020). Det har visat sig att de flesta av dessa modellprediktioner var överdrivna (Barda m.fl. 2020). Dödligheten bland dem som registrerats som smittade i hela världen ligger nu mellan 0,25-5 procent (Roser m.fl. 2020). Om vi antar att hälften av dem som smittats registrerats som smittade skulle det innebära en mortalitetsrisk om 0,125-2,5 procent. I detta läge är det relevant att ställa dessa risker (intäkter) i kontrast mot andra risker (kostnader) som en nedstängning av ekonomin innebär.

med förmågan att väga samman konsekvenser på olika politik-
områden.

3 Kausal analys via experiment eller observation

Strävan efter att se eller vilja finna och förstå orsakssamband är mänsklig. Vi ställer oss ofta vardagliga kontrafaktiska frågor eller påståenden som är av kausal natur, exempelvis: "Om jag dricker kaffe nu, kan jag sova senare?". Kausala frågeställningar på samhällsnivå uppkommer på likartat sätt, det vill säga utifrån en hypotes som bygger på en föreställning om en effekt och en vilja att förstå om denna effekt finns. Vi kan till exempel tro att färre elever per lärare gör att eleverna lär sig mer. Tittar vi på data (så kallat observationsdata) så ser vi dock det omvända, elever i större klasser har bättre betyg.

Utvärderings- eller skattningsproblemet med observationsdata består i att de individer som får eller väljer behandling skiljer sig från dem som inte får eller väljer behandling, och att det som skiljer grupperna åt är relaterat till det utfall vi vill mäta. I exemplet med antalet skolelever kan det positiva sambandet antas bero på att vi i Sverige har en kompensatorisk resurstilldelning till skolor, det vill säga mer resurser tilldelas skolor/elever med inlärningssvårigheter vilket möjliggör mindre klasser för dessa elever. En enkel jämförelse kommer därför inte att vara meningsfull om syftet är att analysera den kausala frågeställningen.

I avsnitt 3.1 diskuteras experimentella data för att göra trovärdiga analyser av orsak- och verkanssamband, medan avsnitt 3.2 diskuterar samma möjlighet med observationsdata. Avsnittet avslutas med en diskussion i avsnitt 3.3 om problem kopplade till vilka studier som når vetenskaplig publikation.

3.1 Experiment

Experiment är ett av de mest centrala begreppen inom vetenskaplig metodik. Exempelvis skrev Albert Einstein i ett brev till J.S Switzer 1953 att *“Development of Western science is based on two great achievements: the invention of the formal logical system (in Euclidean geometry) by the Greek philosophers, and the discovery of the possibility to find out causal relationships by systematic experiment (during the Renaissance).”*¹¹

Den viktigaste ingrediensen i ett experiment är möjligheten att kunna förkasta en teori. Om förutsägelsen förkastas måste teorin uppdateras och om experimentet kan replikeras kan teorin iterativt korrigeras. Om de teoretiska förutsägelserna är precisa går det att konstruera experiment som är kraftfulla. Idén att använda experiment för att empiriskt testa teorier blev (som citatet av Einstein illustrerar) tidigt viktiga inom fysiken. Ett exempel är när Galileo Galilei 1609 fann bevis för den heliocentriska modellen, det vill säga uppfattningen att jorden kretsar runt solen. Det är ingen tillfällighet att experiment först började användas inom fysiken och först långt därefter bland forskare inom andra discipliner. Inom fysiken kan precisa teoretiska prediktioner göras och förutsättningarna för att konstanthålla (eller kontrollera) störande (externa) element är goda. Av denna anledning var de tidiga studierna i fysik inte i behov av statistisk analys för att utgöra framsteg. De teoretiska prediktionerna gav lite spelrum för tolkning och experimenten kunde upprepas.

Inom områden som medicin och samhällsvetenskap är det generellt sett svårare att genomföra experiment med samma trovärdighet som inom fysiken. Detta eftersom experimenten som regel involverar människor. Ett första problem med experiment på människor är den stora naturliga variation som finns mellan människor i både förutsättningar (exempelvis med avseende på humankapital och gener) och beteenden (som kan formas av den sociala miljön), som innebär att det som ger verkan för en individ inte behöver göra det för en annan. Är urvalspopulationen inte representativ för den population som experimentet vill uttala sig om (policypopulationen) är resultaten kanske därför inte särskilt användbara. Dessa problem i kombination med att vi människor har en fallenhet att se mönster

¹¹ Einstein Archive 61-381.

där inga mönster finns (så kallad *apophania*) gör det svårt att separera slumpmässigt förekommande atypiska resultat från de systematiska. Om man ska göra framsteg med experiment som involverar människor måste man därför ta hänsyn till denna naturliga variation och kunna bedöma när en effekt finns som skiljer sig från den naturliga variationen i utfall. Det är här som det klassiska randomiserade experimentet kommer in i bilden (Fisher 1926, 1935).

Intressant att notera är att James Lind år 1747 genomförde det första kända experimentet inom medicin. Experimentet genomfördes på tolv sjömän med likartade symptom på skörbjugg. Alla sjömän erbjöds något av de sex olika kosttillskott som Lind ville undersöka som botemedel mot skörbjugg, varav ett kosttillskott var citrusfrukter. Han fann att citrusfrukter var överlägset bäst – de patienterna var helt arbetsföra efter sex dagar. Näst bäst var cider – de männen var på bättringsvägen då försöket avbröts efter fjorton dagar. I likhet med de fysikaliska experimenten var inte tilldelningen av de sex behandlingarna randomiserade bland de tolv sjömännen. Ändå trodde, i alla fall Lind och sedermera även James Cook, på resultaten från experimentet. Detta sannolikt för att man hade en homogen population av sjömän och att skillnaden i männens tillfriskande var så påtaglig. I och med att James Cook inte förlorade en enda man i skörbjugg efter att ha tillhandhållit citrusfrukter under sin världsomsegling på 1770-talet kan man säga att teorin om citrusfrukters goda verkan bekräftades.

I avsnitt 3.1.1 diskuteras de grundläggande idéerna kring hur man kan göra sannolikhetsbedömningar av huruvida en effekt existerar i en population av intresse (det som benämns statistisk inferens i den vetenskapliga litteraturen). I avsnitt 3.1.2 diskuteras det ideala randomiserade experimentet när enheterna för analys är individer. I 3.1.3 summeras avsnittet.

3.1.1 Det randomiserade experimentet

Principen kring the randomiserade experimentet för kausal inferens kan härledas till Fisher (1926, 1935). Andra tidiga viktiga bidrag är Yates (1937) kring faktoriella experiment och Box och Wilson (1951) kring responsytor vid randomiserade kliniska prövningar. Den metodologiska forskningen kring experimentell design har fått

ett förnyat intresse på senare tid. Det primära skälet till detta är ökad tillgänglighet till data via digitalisering och intresset från affärsvärlden (t.ex. Google, Amazon, Spotify etc.) att utnyttja sina data för att öka sina vinster. De nya experimentella designerna använder datorer för att skapa mer effektiva allokeringar än de som ges från renodlade randomiserade experiment som beskrivs nedan.¹²

En behandlingseffekt definieras som skillnaden i utfall för en individ som genomgått behandling jämfört med den som hade blivit fallet om samma individ inte genomgått behandling. Eftersom endast ett av dessa utfall går att observera i verkligheten (den andra är kontrafaktisk) kan en behandlingseffekt inte identifieras på individnivå utan bara på grupp nivå. Den sanna genomsnittliga effekten utgörs av medelvärdet av alla individuella effekter. Eftersom det kontrafaktiska utfallet inte observeras kan vi aldrig veta vad den sanna effekten är, varken i själva experimentet eller, självfallet, i populationen.

En effekt kan skattas genom att bilda skillnader i medelvärde mellan de som får behandling och de som inte får behandling. Om dessa skillnader i medelvärderna ska vara en bra skattning av den sanna effekten bör det inte finnas en systematik i vem som får behandling eller inte. Om vi låter slumpen avgöra vem som tilldelas behandling finns ingen systematisk sortering till behandling, de som får behandling kan i genomsnitt förväntas ha samma egenskaper som de som inte får behandling. Det gäller egenskaper som både kan och inte kan observeras i data. När behandlade och icke-behandlade är lika är de genomsnittliga skillnaderna i utfall en bra skattning på den sanna effekten av behandlingen för experiment- eller studiepopulation.

En randomiserad tilldelning av behandling innebär emellertid inte att balans mellan de jämförande grupperna uppstår per automatik. I synnerhet vid småskaliga experiment (det vill säga få deltagare) är det sannolikt att betydande skillnader finns mellan grupperna ändå. Med andra ord kan resultatet i ett enskilt experiment vara långt från den sanna effekten för den aktuella studiepopulationen. Fördelen med det randomiserade experimentet i jämförelse med andra ansatser är att vi har kännedom om sannolikheten att få behandling. Det betyder att vi utan modellantaganden kan skapa ett konfidensintervall. Konfidensintervallet säger oss hur troligt det är att den

¹² För en jämförelse med de så kallade optimala designerna och det randomiserade experimentet se Kapelner m.fl. (2020).

sanna effekten ligger inom ett intervall. Tolkningen är sådan att alla värden inom detta konfidensintervall är lika troliga att vara det sanna värdet, det vill säga den specifika punktskattningen är inte mer sannolik än konfidensintervallets ändpunkter.

Om studiepopulationen är representativ för hela populationen av intresse (det vill säga om studiepopulationen dragits slumpmässigt från populationen av intresse) är det även troligt att den sanna medelvärdeseffekten för hela populationen ligger i detta konfidensintervall. Är effekten olika för olika individer, det vill säga heterogen, och studiepopulationen inte är representativ för hela populationen, är konfidensintervallet att betrakta som internt valid, men ej externt valid. Med andra ord kan den sanna effekten i populationen vara långt ifrån det som täcks av det i experimentet beräknade konfidensintervallet.

Det randomiserade experimentet har sitt ursprung i experiment med grödor och gödning, det vill säga i biologiska material med genetisk variation men där enheterna inte har något medvetande. I dessa sammanhang passar det randomiserade experimentet bra eftersom heterogeniteten inom till exempel en och samma gröda ofta är låg. Framförallt kan experimentet enkelt upprepas under likartade förhållanden vilket möjliggör slutsatser om en sann medelvärdeseffekt. Experiment på människor är svårare att upprepa under likartade förhållanden och heterogeniteten är betydande varför det är svårare att tro vi kan identifiera en sann effekt för populationen i fråga.

Experiment där tilldelning av behandling bestäms deterministiskt baserad på information om enheter som ska ingå i experimentet är sannolikt mer effektiva än randomiserade experiment för att förstå orsakssamband. Till exempel förordade William Gosset (primärt känd för "Student t-fördelning") icke-randomiserade experiment vid utveckling av Guinness öl.¹³ Optimala designer som numera ofta används av it-företag är inte heller randomiserade i traditionell mening. Icke-randomiserade experiment är sannolikt användbara för att förbättra en produkt inom ett företag. Detta eftersom det inte finns några målkonflikter inom företaget, det vill säga det finns ingen anledning för företaget att fuska med tilldelningen för att visa

¹³ Konflikten mellan effektivitet och väntevärdesriktighet illustreras väl i Student (1938) (Uppsatsen publicerades med hjälp av Egon Pearson och Jerzy Neyman, efter William Sealy Gossets (dvs. "Student") död 1937. Se också Ziliak (2013) för en mycket intressant kritik av Levitt och List (2009) rörande experiment inom industriell ekonomi.

på ett specifikt resultat. Fördelen med det randomiserade experimentet är att resultaten blir trovärdiga gentemot utomstående avnämare, exempelvis forskningssamhället eller myndigheter som godkänner eller subventionerar behandlingar baserade på resultaten. Vad som är viktigt att förstå är emellertid att trovärdighet kan skapas på andra sätt, exempelvis genom att publicera en empirisk design, eller försöksplan, för en analys innan analysen genomförs, något vi diskuterar mer utförligt i avsnitt 4.

Sammanfattningsvis, det finns två distinkta fördelar med det ideala (definieras i avsnitt 3.1.2) randomiserade experimentet: (1) det finns en enkel beslutregel för att avgöra om det finns en effekt eller ej (till exempel genom att bilda ett konfidensintervall), och (2) tilldelning av behandling är trovärdig mot utomstående avnämare, exempelvis forskningssamhället eller myndigheter som godkänner eller subventionerar behandlingar baserade på resultaten.

Intern och extern validitet

Forskning på människor kan innebära betydande risker för de som deltar i ett experiment. Innan 1964 fanns det inga riktlinjer för hur denna forskning skulle bedrivas (European Commission 2013). Numera är regelverket strikt världen över (ibid). Ett krav vid genomförandet av ett medicinskt experiment är informerat samtycke om att delta i experimentet. Dessutom får vissa grupper, exempelvis fertila kvinnor, inte delta i vissa typer av experiment. Eftersom det kan vara riskfyllt att delta i experiment kommer sannolikt mer riskbenägna grupper av individer utgöra experimentpopulationer. Utöver detta vill forskare, som illustrerats ovan, gärna göra experiment på en population som är så homogen som möjlig, till exempel friska och arbetsföra. Eftersom yngre män är riskbenägna, i regel friska och utan hormonsvängningar kommer de relativt ofta att utgöra experimentpopulation, i alla fall när det finns risker med experimentet. Därmed är inte experimentpopulationen nödvändigtvis representativ för den population som främst kan förväntas använda medicinen, det vill säga experimentet kan ha hög intern validitet men låg extern validitet. Intressant att notera är därför det

starkt ökande intresset för observationsstudier med syfte att uppskatta effekter i klinisk praxis inom medicin.¹⁴

Policyexperiment

Policyexperiment har som syfte att få kunskap om hur olika satsningar, exempelvis arbetsmarknadspolitiska program för arbetslösa, rehabilitering av sjukskrivna eller utbildningsinsatser för barn och unga, påverkar deltagarna. Inom samhällsvetenskapen är informerat samtycke ibland inte nödvändigt.¹⁵ Det innebär att dessa experiment, i motsats till experiment inom medicin, inte behöver brista i den externa validiteten. I stället består svårigheten många gånger i (1) att kunna bedöma den samhällsekonomiska effekten om åtgärden införs i full skala och, (2) att förstå om effekten beror av kontexten. Vad beträffar (1) består svårigheten i att en åtgärds långsiktiga effekt oftast beror av både utbud och efterfrågan på en marknad. En positiv effekt från ett experiment kan bero på att de som fick åtgärden fick fördelar gentemot de som inte fick åtgärden, något som inte uppstår när åtgärden implementeras i full skala. En åtgärd kan också leda fram till beteendeförändringar som på lång sikt

¹⁴ I situationer där randomiserade experiment (RCT) inte har betraktats som etiskt och/eller där de inte har varit möjliga att utföra (godkännande under exceptionella omständigheter infördes 1993) eller när patienter med allvarligt försvagande tillstånd eller livshotande sjukdom behöver snabbare tillgång till läkemedel (villkorligt godkännande antaget 2004) godkänner numer europeiska läkemedelsmyndigheten (EMA) observationsstudier både som substitut och som komplement till RCT. En ny förordning som antogs 2010 och infördes 2012 möjliggör också villkorade godkännanden av läkemedel efter utvärdering av säkerhet och/eller effektivitet i klinisk praxis (dvs observationsstudier). Detta tillvägagångssätt är en del av EMA:s ansträngningar för att förbättra patienters tillgång till nya läkemedel i tid genom att initialt godkänna läkemedlet baserat på bevis från en begränsad population eller med intermediära utfall och sedan komplettera med bevis från observationsstudier på en större population eller med mer slutgiltiga utfall, t.ex. dödlighet, för godkännande för en större befolkning eller för fortsatt godkännande.

¹⁵ Enligt den svenska etikprövningslagen krävs informerat samtycke för en forskningsstudie om den syftar till att påverka forskningspersonen fysiskt eller psykiskt, eller löper en uppenbar risk att skada forskningspersonen fysiskt eller psykiskt. Av lagens förarbeten och av praxis framgår att man med denna formulering framför allt avser medicinsk forskning och forskning inom klinisk psykologi. År 2005 prövade centrala etikprövningsnämnden ett ”informations-experiment” och slog fast att lagen inte var tillämplig trots att samtycke inte kunde inhämtas, då risk för skada inte förelåg. Utöver detta kan man använda filosofiska argument kring vad som är bäst när det gäller beslut under osäkerhet, dvs. om staten inte vet vad som är det bästa för en individ.

oftast minskar eventuella positiva eller negativa kortsiktiga effekter.¹⁶ Det betyder att en effekt konstaterad i ett mindre policy-experiment som inte kunnat ge svar på varför effekten uppstått ofta har ett begränsat policyvärde. Vad beträffar (2) är problemet att ett randomiserat experiment genomfört i Sverige kan ge helt annorlunda effekter än ett experiment genomfört i andra delar av världen. För att exemplifiera, en rehabiliteringsinsats avsedd för sjukskrivna i Sverige kan ha en helt annan effekt än den har för sjukskrivna i USA eftersom våra socialförsäkringssystem är så olika. Även när det gäller läkemedel kan kontexten ha betydelse (till exempel på grund av olikheter i kost eller genetiska variationer i en homogen befolkning), men problemet är sannolikt mycket större vad gäller sociala experiment.

En slutsats är således att det blir svårare att bygga upp en sammanvägd evidens inom områden där kontexten kan ha betydelse för den skattade effekten än för exempelvis läkemedel. Ett sätt att försöka beakta problem (1) är att genomföra experimentet i två steg. I ett första steg tilldelas omfattningen av den åtgärd som studeras på "marknadsnivå" (till exempel mellan samhällen eller regioner) slumpmässigt. I det andra steget fördelar man sedan slumpmässigt individer (inom samhället eller regionen) till den del av gruppen som får del av åtgärden. På så sätt kan man skatta effekten på både individnivå och marknadsnivå och därmed också storleken på eventuella spridningseffekter. Se Crépon m.fl. (2013) och Cheung m.fl. (2019) för bra exempel för att studera effekten av aktiv arbetsmarknadspolitik på både individnivå och samhällsnivå.

3.1.2 Det ideala randomiserade experimentet

Experiment på människor är förknippade med tydliga problem i jämförelse med experiment på grödor och gödning. Skälet är att människor har ett medvetande. Bara vetenskapen om att vi deltar i ett experiment kan påverka vårt beteende. Detta går i utvärderings-

¹⁶ En åtgärds långsiktiga effekt beror oftast på både utbud och efterfrågan på en marknad, vilket inte är fallet med medicin som åtgärd. Således kan en omedelbar kortsiktig positiv effekt leda till en omvänd långsiktigt negativ effekt på grund av beteendeförändringar. Detta problem har benämnts som Lucaskritiken (Lucas 1976) men skulle snarare betecknas som Marschak-kritiken. Detta eftersom Jacob Marschak berörde detta redan 1953 i Marschak (1953), dvs. 23 år innan Robert Lucas.

litteraturen under benämningen för ”*Hawthorne-effekt*” eller ”*placeboeffekt*” och skapar problem i tolkningen av resultaten: Är det behandlingen i sig eller vetskapen om att vi är med i experimentet som orsakar resultaten? För att undvika denna beteendeeffekt måste man dölja, eller ”blinda”, vem som får en aktiv åtgärd. Helst ska även den som ger åtgärden eller behandlingen vara omedveten om vilken behandling denna ger (detta benämns ”dubbelblindning”).¹⁷ Dubbelblindning är tämligen lätt att åstadkomma när åtgärden avser ett läkemedel men kan vara betydligt svårare, eller rentav omöjligt, i andra sammanhang. På grund av vår sociala natur finns också en stor risk att de individer som deltar i ett experiment påverkar varandra.¹⁸ Om man inte kan blinda behandlingen och individerna informeras om att de deltar i ett experiment är det sannolikt så att den empiriska studien har sämre validitet än många observationsstudier.¹⁹

Ett exempel

För att exemplifiera betydelsen av heterogena effekter och olika potentiella bias inom ramen för ett randomiserat experiment kan vi tänka oss utvärderingen av ett fiktivt arbetsmarknadspolitiskt program där målet är att deltagaren kommer i arbete. Vi kan tänka oss två scenarion. I scenario I görs en etikprövning eftersom utvärderingen betraktas som forskning och det dessutom finns en risk för att det erbjudna programmet är skadligt för deltagarens hälsa. Då kommer det också att krävas informerat samtycke om att delta i experimentet. I scenario II görs ingen etikprövning eftersom experimentet inte betraktas som forskning och inte heller som skadligt för

¹⁷ Forskaren som genomför experimentet kan potentiellt påverka (sannolikt omedvetet) resultat eftersom han/hon ofta är drivande för att en specifik behandling ska vara verkningsfull. Det har visat sig att i studier där randomiseringen (dvs. vem som fått behandling eller inte) varit otillräckligt dold får ca 30–40 procent större effekter än när randomisering varit dold för forskarna (Schulz 1995).

¹⁸ Ett illustrativt exempel är experiment med aidsmediciner. Schuklenk (2010) visar att försökspersonerna (dvs. både behandlade och kontroller) i vissa studier delade piller (dvs. piller med aktiv ingrediens och sockerpiller) för att på så sätt försäkra sig om att få den aktiva ingrediensen. I andra studier gick försöksindivider till apotekare för att ta reda på om de fick en aktiv ingrediens eller sockerpiller. Om de fick reda på att de fick sockerpiller betedde de sig kanske annorlunda än om de fick den aktiva ingrediensen. Båda handlingarna, baserade på individers incitament att överleva, minskade således studierna validitet. Ett annat exempel är Hartman m.fl. (2013) som vi kommer att diskutera utförligt i rapporten.

¹⁹ Notera att typ av behandling och utfall har betydelse. För att exemplifiera. Om behandlingen är ett hjärtkirurgiskt ingrepp och utfall är dödlighet är nog avsaknad av blindning inget problem, men om behandlingen berör beteendeterapi och utfall berör enkätsvar på upplevd hälsa finns stora problem med avsaknad av blindning.

deltagarnas hälsa. Något informerat samtycke behövs då inte. Arbetsförmedlingen, som är den som administrerar de arbetsmarknadspolitiska programmen, skulle exempelvis kunna hänvisa till sin instruktion som anger att de insatser som erbjuds ska vara effektiva för att motivera genomförande av experimentet.²⁰ Detta har i praktiken gjorts vid upprepade tillfällen, vilket redovisas närmare i avsnitt 5.

Om scenario II genomförs och deltagarna i experimentet dras slumpmässigt ur huvudpopulationen kan vi med en given risk (ofta väljs 5 procent) om felaktig slutsats avgöra om åtgärden har verkan eller inte. Eftersom varken arbetslösa eller handläggare är informerade om experimentet kommer resultaten vara externt valida, med andra ord giltiga för samtliga arbetslösa.

Om scenario I genomförs och om informationen om experimentet inte påverkar deltagare eller handläggare kommer resultaten även här vara externt valida, men för en annan population. De effekter som mäts i experimentet är inte representativa för samtliga arbetslösa, eftersom bara de som är positivt inställda till programmet kommer att vilja delta i experimentet, men eftersom det finns ett inslag av frivillighet också vid reguljärt utförande av programmet blir resultaten externt valida för en population av primärt intresse.

Emellertid finns potentiella problem förknippade med att informera om experimentet. De som då anmäler intresse för att delta men lottas in i kontrollgrupp skulle eventuellt kunna försöka få likartade insatser via reguljär verksamhet, vilket skulle göra att den beräknade effekten i experimentet underskattas. Detta brukar kallas *substitutionsbias*. Även handläggarnas beteende skulle kunna påverkas av vetenskapen om experimentet. Exempelvis skulle de kunna lägga extra mycket resurser på deltagarna vilket skulle riskera att de sanna effekterna överskattas. Även det omvända är tänkbart, det vill säga att handläggarna blir extra benägna att kompensera icke-deltagarna för de uteblivna insatserna, vilket alltså skulle kunna leda till att effekterna underskattas i förhållande till vad insatserna skulle generera vid reguljärt utförande.

²⁰ Se Förordning (2007:1030) med instruktion för Arbetsförmedlingen, SFS 2007:1030 t.o.m. SFS 2020:94 - Riksdagen. Experimentet som beskrivs av Cheung m.fl., (2019), och som genomfördes av Arbetsförmedlingen i samarbete med IFAU, var blindat. Arbetsförmedlingen hävdade att det rörde sig om utveckling medan kritiker menade felaktigt att krav om samtycke behövts enär detta var forskning eftersom det rörde sig om känsliga uppgifter (se <https://www.dagenssamhalle.se/debatt/oacceptabelt-att-moerka-forskning-pa-arbetsloesa-16584>).

3.1.3 Summering

Tack vare de stora framgångarna med evidensbaserad medicin med sitt fokus på det randomiserade experimentet finns i dag en övertro på det randomiserade experimentet som en lösning för alla möjliga frågeställningar.

Diskussionen ovan visar att det krävs mycket mer än en randomisering för att resultat från ett randomiserat experiment ska kunna ses som ett stöd för förekomsten av en kausal effekt. Även ett idealt dubbelblindat experiment utgör inte någon evidens. Resultatet från ideala experiment med ett stort antal individer²¹ innebär, förvisso, att man med stor säkerhet kan säga att en effekt existerar, men detta endast i den givna kontexten. Replikering av resultat från ideala experiment utgör evidens eller vad vi skulle kunna säga en sanning.

Enligt vår mening kan randomiserade experiment för policyutvärdering motiveras förutsatt att experimentet kan leda till en fördjupad förståelse om varför effekten uppstår och att det även kan ge en förståelse om framtida liknande åtgärder, och även då i något andra kontexter. Denna syn kan jämföras med läkemedelsexperiment där man innan experiment på individer har (1) en tydlig teori om de aktiva ingrediensernas betydelse och (2) genomförda djurförsök som styrker teorin. Om experimentet ger ett statistiskt säkerställt resultat har vi en teori/kunskap om varför effekten uppstår. Vid policyexperiment är det oftast svårt att ha en lika stark teori om varför en effekt kommer att uppstå, många gånger kan effekten dessutom skapas via förändringar i incitament/preferenser hos individer. Av denna anledning är det av stor vikt att inte bara samla information om utfall för individerna efter det att en behandling givits, utan även information om varför en eventuell effekt uppstår, så kallade mediatorvariabler i utvärderingslitteraturen.

3.2 Observationsstudier och kvasiexperiment

I tidiga observationsstudier försökte man uppnå jämförbarhet mellan behandlade och icke-behandlade genom att i regressionsmodeller justera för "confounders", det vill säga bakomliggande

²¹ Vad som är stort experiment beror på den sanna effektens storlek och heterogenitet i den population som studeras. Experiment med mer än 1 000 individer kan i regel betraktas som stora nog att kunna utesluta en nolleffekt.

skillnader i kovariater mellan grupperna. Dessa ibland slentrianmässiga kovariatjusteringar är inte lämpliga för analyser av kausalitet eftersom resultaten blir helt beroende av vilka kovariater som inkluderas eller exkluderas i analysen (Pearson 1899). Insikten att relationer kan försvinna eller till och med byta tecken är känd som Simpsons paradox.²² I stället för att utveckla metoder för trovärdig analys av observationsdata var rekommendationen från dåtidens statistiska forskare att man endast skulle uttala sig om kausalitet vid randomiserade experiment. Vid observationsstudier skulle forskarna använda sig av begreppen "association" eller "korrelation". Kruket är att även om association ofta nämns i forskningsartiklar så tolkas resultat från regressionsmodeller oftast i slutändan som kausala av forskare och medier. Anledning till detta är att de intressanta frågeställningarna är kausala vilket gör att associationer oftast blir ointressanta.

Inte minst inom nationalekonomi växte frustrationen med att regressionsansatserna inte kunde ge trovärdiga utsagor om kausalitet. Detta illustreras i ett flertal kritiska artiklar och böcker under det sena 1970-talet och tidiga 1980-talet (se Hendry 1980; Sims 1980; Leamer 1978, 1983). Kritiken markerar början på en empirisk revolution inom nationalekonomi som innebar framväxten av analyser med syftet att efterlikna det randomiserade experimentet, detta genom att använda extern variation i behandlingsprevalens, så kallade naturliga experiment.

Under samma tidsperiod hände något annat inom statistikämnet. Grunden för denna förändring i perspektiv var en serie uppsatser av Donald Rubin (Rubin, 1974, är den första) som visade att även analyser med observationsdata bör hålla isär design och analys. I traditionella observationsstudier med regressionsanalys ligger fokus på att modellera utfallet, det vill säga design och analys sker i ett enda steg. Detta förfarande lockar forskaren till att manipulera regressionsmodellen genom att exkludera och/eller inkludera kovariater för att bekräfta sin teori. Rubins inställning var i stället att fokusera på tilldelningen av behandlingen på samma sätt som vid det randomiserade experimentet. Designen kan, som vid det randomiserade experimentet, genomföras utan att det efterföljande utfallet kan observeras. När väl designen är fastställd görs analys på samma sätt

²² Edward H. Simpson beskrev först detta fenomen 1951, men Karl Pearson (1899) och Udny Yule (1903) hade nämnt liknande effekter tidigare.

som vid ett randomiserat experiment. Holland (1986) benämner detta ramverk ”The Rubin Causal Modell” (RCM).

Både naturliga experiment och RCM går i dag under benämningen kvasiexperimentella design. Trots en positiv utveckling är majoriteten av de observationsstudier som genomförs fortfarande av låg kvalitet. En anledning kan vara att rekommendationen att endast diskutera association vid analys av observationsdata bidrar till att ansträngningar för att korrigera för eventuell selektion till behandling inte görs. Man skattar trots allt bara en association. En annan orsak kan vara att förbättrad mjukvara och ökad tillgänglighet till data har gjort det lätt för forskare att skatta regressionsmodeller (eller maskinlärningsalgoritmer) med utfall på vänster sida och kovariater på höger sida om likhetstecknet. Det är lätt att tolka dessa modeller som att högerledet ger vänsterledet. I många läroböcker beskrivs dessutom fortsatt parametrar i regressionsmodeller i kausala termer (”den marginella effekten av X på Y”). I kapitel 7 illustreras problemet med feltolkade regressionsanalyser med ett exempel kring brottslighet och socioekonomiska faktorer.

Det är med andra ord inte konstigt att observationsstudier fortsatt har ett dåligt rykte, och då speciellt inom medicin. För att få till en förändring måste den metodologiska kunskapen förbättras inom alla discipliner. Professionen, likväl som allmänheten, måste kunna skilja mellan bra och dåliga observationsstudier.

Observationsstudier med en kvasiexperimentell design är ofta av mycket hög kvalitet. Det är värt att notera att Benson och Hartz (2000), som jämför resultat från observationsstudier med resultat från randomiserade experiment, inte finner några belegg för att de genomsnittliga effekterna från observationsstudier (fem studier var för två olika typer av behandlingar) efter 1985, kvantitativt skiljer sig från de genomsnittliga effekterna från randomiserade experiment (nio och tio studier för de två behandlingarna).²³ Orsaken, enligt författarna, är utvecklingen med allt bättre icke-randomiserade metoder, det vill säga kvasiexperimentella ansatser, efter 1985.

Inom nationalekonomi är begreppet kvasiexperiment snävare definierat än inom statistik och epidemiologi. Inom nationalekonomi är oftast kvasiexperiment synonymt med ett naturligt

²³ I studien ingick tio RCT-studier och fem observationsstudier av effekten av hormonterapi för kvinnor på bendensitet. Vidare hade man nio RCT-studier och fem observationsstudier av effekten av kalciumkanalblockerare på dödlighet för njurtransplanterade patienter.

experiment. Dessa har, som namnet antyder, sin utgångspunkt i att använda variationer som finns i naturen, till exempel nederbörd, vind, tvillingfödelse, syskonkön med mera. Men även förändringar i institutioner och politik klassificeras som naturliga experiment av ekonomerna.²⁴ En kritik som kan riktas mot det naturliga experimentramverket är att det reducerar vilka analyser som kan göras. Innebörden är att frågorna styrs av vilken variation som kan utnyttjas snarare vilka frågor som är intressanta att studera.

I avsnitt 3.2.1–3.2.3 exemplifierar vi kvasiexperimentella designers inom nationalekonomi med dem som är mest förekommande. Dessa är, *”Instrumentvariabelansatsen”*, *”Skillnad i skillnad”-ansatsen* och *”Regressionsdiskontinuitsansatsen”*. I avsnitt 3.2.4 presenteras Rubins kausala modell och avsnittet avslutas av med en summering i avsnitt 3.2.5.

3.2.1 Instrumentvariabelansatsen (IV-ansatsen)

IV-ansatsen utvecklades först av Philip och Sewal Wright för cirka hundra år sedan men upptäcktes på nytt under 1940-talet (Haavelmo 1943). Ursprungligen var intresset att skatta parametrar i ekvations-system där variabler är simultant bestämda, exempelvis priser och kvantitet på en marknad.²⁵ Metoden används numera ofta vid mer traditionell policyutvärdering vilket beskrivs nedan.

För att få en uppfattning om hur IV-ansatsen fungerar i praktiken använder vi ett exempel från Permut och Hebel (1989). I denna studie var man intresserad av att analysera effekten av mödrars rökning på spädbarns hälsa, där hälsa mäts via barnens födelsevikt. Det är sannolikt så att barn till rökande mödrar har en lägre födelsevikt än barn till icke-rökande mödrar. Denna relation behöver dock inte bero på rökningen i sig utan kan bero på skillnader i socio-ekonomisk status eller andra riskbeteenden (till exempel alkohol- och näringsintag). För att isolera betydelsen av rökning skickades ett brev ut till en grupp slumpvis utvalda rökande mödrar, vald från

²⁴ För att skilja ”slumpmässig” variation från naturen från situationen med ”extern” variation från en politik definierar Rozenzweig och Wolpin (2000) ”naturliga experiment” som studier som använder ”slumpmässig” variation från naturen.

²⁵ Enligt Stock och Trebbi (2003) är Philip Wright (1928) den som först använder ordet instrument som medel för att identifiera effekter av endogena variabler. Bilaga B till Wright (1928) bevisar, sannolikt av hans son Sewal Wright, både identifikation av en av ekvationerna (parametern) samt identifikation av hela systemet.

en större population av konstaterade rökare. I brevet informerades om riskerna med rökning. Syftet med brevet, som utgör det så kallade *instrumentet* här, var att skapa en systematisk skillnad mellan grupperna med avseende på sannolikheten att sluta röka. Eftersom brevet inte kan förväntas ha något direkt samband med barnets hälsa och födelsevikt, utan bara indirekt via moderns rökning, kan den uppkomna skillnaden i rökning sedan användas för att beräkna effekterna av rökningen på barnens hälsa.

Forskarna konstaterade att utskicket fått en betydande inverkan på mottagarnas rökbeteende. Bland dem som mottog brevet slutade 43 procent att röka. Motsvarande procentandel i kontrollgruppen var 20 procent, således en skillnad på 23 procentenheter. Vidare konstateras att barnen till mödrarna som fick brevet i genomsnitt var 98 gram tyngre vid födseln än barnen till de som inte fick brevet. Detta är alltså en underskattning av den faktiska effekten av rökning på födelsevikten eftersom några i gruppen som mottog brevet fortsatte att röka och några i gruppen som inte mottog brevet slutade röka ändå (idealt ska skillnaden vara 100 procentenheter). För att beräkna en genomsnittlig effekt av att sluta röka behöver vi därför multiplicera skillnaden i födelsevikt mellan grupperna med 4,34 ($= 100/23$). Slutsatsen är att om de rökande mödrarna slutar röka så kommer deras barn i förväntan vara 426 gram tyngre än om de fortsatt att röka. Noterbart är att effekten inte avser alla rökande mammor utan bara mammor i den studerade populationen och, i strikt mening, för de som skulle ändrat sitt beteende till följd av brevet. Effekten för alla mammor kan vara både mindre och större.

Utmaningen med IV-ansatsen är att troliggöra att det instrument som används inte har någon inverkan på det utfall som studeras annat än den som går via behandling. Eftersom de som mottog brevet utsågs slumpmässigt och inte själva kunde välja är detta antagande sannolikt uppfyllt i ovan exempel. I de fall där instrumentet inte är randomiserat kan antagandet dock alltid ifrågasättas. Ett exempel är studier av effekterna av högre utbildning. Eftersom de som har bättre förutsättningar att klara högre studier också är mer benägna att läsa vidare behövs ett instrument för att bryta associationen mellan det som ska analyseras (inkomst/lön) och utbildning, annars riskerar effekten att överskattas. Card (1995) använde avstånd till närmaste studieort som instrument. Analyser visar att en längre distans minskar sannolikheten att börja på högre utbildning.

Emellertid kan även kostnaderna för högre utbildning förväntas öka med avståndet till studieorten vilket innebär att bara de med förväntat hög avkastning av utbildningen kan förväntas läsa vidare bland dem som bor längst bort från studieorten. Med andra ord kan det inte uteslutas att det finns en direkt koppling mellan avståndet från studieort och den framtida inkomsten eller lönen, vilket innebär att den beräknade effekten snedvrids.

3.2.2 "Skillnad i skillnad"-ansatsen (DiD)²⁶

DiD-strategin kan dateras tillbaka till studien av John Snow som testade sin hypotes om smutsigt vatten som orsak till koleraepidemin i London 1854. Han använde det faktum att två vattenföretag, Southwark & Vauxhall Company och Lambeth Company, båda tog vatten från (den smutsiga) floden Themsen. År 1852 flyttade Lambeth Company sitt vattenverk uppåt floden, till ett ställe som var relativt fritt från avloppsvatten. För att testa sin hypotes jämförde Snow dödligheten före och efter 1852 bland dem som betjänades av vatten från respektive företag. Han konstaterade att medan dödligheten i populationen som fick vatten från Southwark & Vauxhall Company var oförändrad, minskade dödligheten i populationen som fick vatten från Lambeth Company efter flytten. Snow kunde därför dra slutsatsen att hans hypotes om smutsigt vatten som orsak till dödligheten var korrekt.²⁷

Noterbart i exemplet är att individer alltså inte tilldelades smutsigt respektive rent vatten av slumpen, som i ett randomiserat experiment, men eftersom konsumenterna inte fritt valde sin vattenförsörjning kunde effekten beräknas ändå. Detta genom tillgången till historiska data, där skillnaden i dödlighet för respektive population kunde beräknas innan Lambeth Company flyttade sitt vattenverk. Denna skillnad kunde sedan användas som utgångspunkt (benchmark) för den skillnad som skulle ha realiserats efter flytten om flytten inte blivit av. Den eventuella förändringen av denna ursprungsskillnad som uppstod efter flytten kunde sedan härledas till flytten.

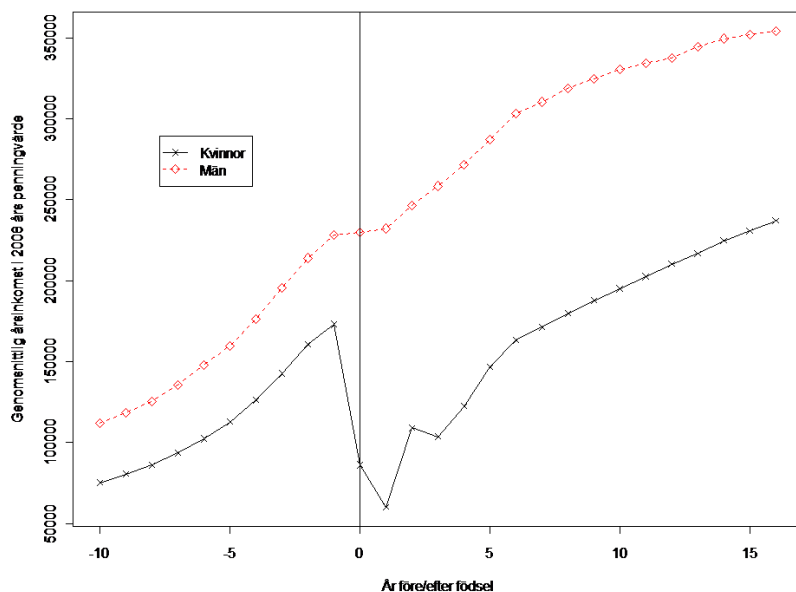
²⁶ DiD kommer av *difference in difference*.

²⁷ I formell mening kan man inte bekräfta en hypotes utan endast förkasta en nollhypotes, det vill säga att inga skillnader finns. Snow gjorde inte heller något formellt statistiskt test (t.ex. genom att bilda konfidensintervall).

I en DiD-ansats härleds effekten således som en skillnad i skillnaden (difference in difference), det vill säga som förändringen mellan två skillnader i det utfall som studeras. Först, skillnaden efter gentemot före en behandling och sedan skillnaden mellan, åtminstone, två grupper. Skillnaden för den obehandlade gruppen är ett mått på den kontrafaktiska förändring för gruppen som påverkas av en förändring (i fallet ovan flytt av vattenverk).

För att effekten ska mätas korrekt krävs att vissa antaganden är uppfyllda. Det viktigaste är att vi måste anta att den skillnad som fanns redan innan förändringen hade bestått även efter förändringen om förändringen inte skett. Detta benämns "parallella trender-antagande" i litteraturen.

Figur 3.1 Inkomstutveckling för mödrar och fäder före och efter föräldraskap. Förvärsarbetande två år innan föräldraskap åren 1990-2002 med inkomst



Källa Angelov m.fl. (2013).

För att illustrera parallella trender-antagandet presenteras figur 3.1 hämtad från Angelov m.fl. (2013). Figuren visar inkomsten för män och kvinnor tio år före och femton år efter att de fått (gemensamt)

barn. Som framgår har grupperna en likartad inkomstutveckling före barnafödandet, men en kraftigt avvikande utveckling efter. Analyserna i Angelov m.fl. (2013) fick stor uppmärksamhet när de presenterades, sannolikt eftersom de är så transparanta. Eftersom inkomstutvecklingen innan barnafödande är likartade för könen kan man anta att den utvecklingen hade fortsatt i frånvaro av föräldraskapet. På samma sätt kan validiteten i traditionella DiD-analyser illustreras om man har data på gruppernas utfall före en förändring.²⁸ Om trenderna divergerar redan innan den händelse man vill studera är analyserna mindre valida.

Vid traditionell DiD med data ordnad på gruppnivå innebär det parallella trend-antagandet att populationssammansättningarna inte får ändras över tid, inte heller får samhällsförändringar påverka de två grupperna olika. I föreliggande exempel med män och kvinnor är frågan om sammansättning inget problem eftersom individdata användes i analysen. Ett potentiellt problem är emellertid att tidpunkten för ”behandling”, det vill säga skaffa barn, bestäms av dem själva (förändringen är inte att beteckna som exogen). Om valet att skaffa barn vid den aktuella tidpunkten är baserad på förväntningar på framtida inkomstutveckling för till exempel kvinnan, men inte för mannen, kan detta medföra problem att tolka fertilitet som en orsak till mödrars lägre lön än fäders.

Ett ytterligare problem vid DiD-analyser är att resultaten kan påverkas av vilken statistisk modell som används. Om de två gruppernas medelvärden i historiska utfall är snarlika är detta inget problem, dock kan problem uppstå om gruppernas medelvärden skiljer sig åt i föreperioden. I analyser av mäns och kvinnors inkomster och löner så ser vi att männens inkomster är högre än kvinnornas innan föräldraskap. Eftersom detta till stor del beror på att männen är två år äldre i genomsnitt kan detta tas hänsyn till i en regressionsanalys.²⁹

²⁸ Angelov m.fl. (2013) är ingen traditionell DiD-analys eftersom båda grupperna är behandlade (dvs. har fått barn) och att analyserna använder individdata. Analyserna bygger dock på samma antaganden som i en traditionell DiD.

²⁹ I detta sammanhang kan nämnas att modellskattning med logariterade inkomster eller inkomster är väldigt små (Angelov m.fl. 2013). Således är det inte troligt att slutsatserna drivs av vilken modell som används vid regressionsjustering.

3.2.3 Regressionsdiskontinuitetsansatsen (RD)

RD-ansatsen presenterades för första gången av Thistlethwaite och Campbell (1960) i en studie av skolstipendiets betydelse för karriären. Ansatsen utnyttjar avbrott, eller diskontinuiteter, i en kontinuerlig variabel som bestämmer behandling. I exemplet med stipendierna utnyttjades det faktum att dessa tilldelas på basis av testresultat, där värden över en förutbestämd gräns (tröskelvärde) kvalificerar för ett stipendium. Med andra ord bestäms "behandlingen" till fullo av det testvärde som är observerbart för forskarna. Det är också just detta som är det fundamentala antagandet för metodens validitet. Om individerna involverade i beslut om behandling kan justera den kontinuerliga variabeln baserat på kunskap om tröskelvärdet är metoden inte valid. Det som är bra med metoden är att om detta förekommer kommer detta att synas i data på så sätt att en större andel av individerna än förväntad kommer att finnas just på den sida av tröskelvärdet som kan anses vara fördelaktig, t.ex. på stipendiesidan av tröskelvärdet.

RD-ansatsen drar nytta av det faktum att den variabel som avgör behandling är kontinuerlig samt antagandet att individer med testresultat precis i närheten av tröskelvärdet är jämförbara förutsatt att det fundamentala antagandet är uppfyllt. I det aktuella exemplet skulle det innebära att effekten av stipendiemottagande går att beräkna genom att jämföra karriärmässiga utfall för individer precis under (icke-behandlade) och precis över (behandlade) tröskelvärdet.

För att beskriva metoden tar vi hjälp av ett exempel på svenska förhållanden som handlar om effekterna av klasstorlek i skolan (Fredriksson m.fl. 2012; Fredriksson m.fl. 2013). Författarna ville studera effekterna av klasstorlek för elever i årskurs 4–6 på dels kognitiv och icke-kognitiv förmåga vid 13, 16 och 18 års ålder, dels sannolikheten för högre utbildning samt löner och inkomster i åldern 27–42 år. Den studerade populationen omfattade elever födda 1967, 1972, 1977 och 1982. Problemet med att identifiera effekter består i att elever som går i små klasser skiljer sig från dem som går i stora klasser. Ett skäl till detta är den kompensatoriska resurstilldelningen som innebär att extra resurser utgår till skolor med elever som har inlärningssvårigheter, som i sin tur möjliggör mindre klasser för dessa elever. En enkel jämförelse mellan elever i stora och små

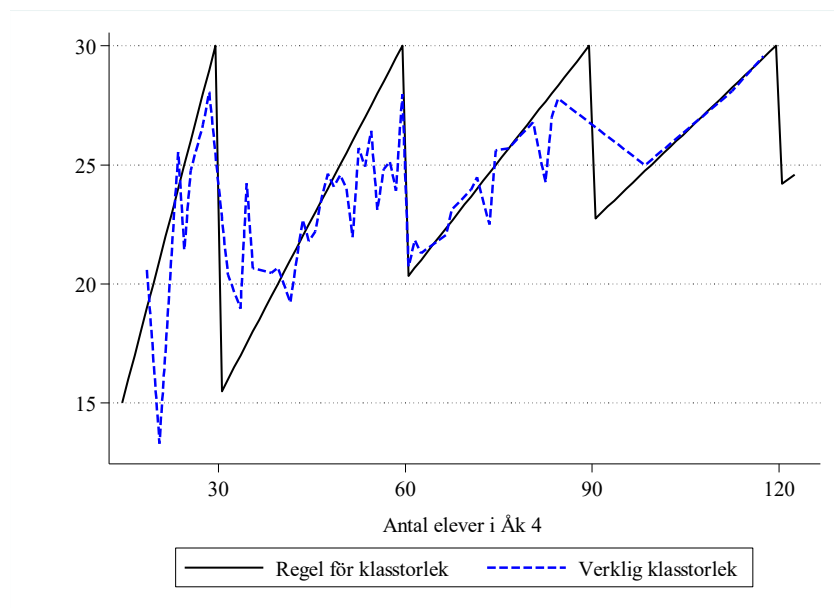
klasser skulle därför inte att vara meningsfull om syftet är att skatta effekter.

För att skatta effekterna använder sig forskarna av en regel från 1962 som innebar att den maximala klasstorleken för elever i årskurs 4–6 var 30 elever. I praktiken innebar regeln att om 30 elever skulle börja i en viss skola så blev det i en enda klass, om 31 elever skulle börja blev det i stället i två klasser, en med 15 elever och en med 16 (se den heldragna linjen i figur 3.2 för en beskrivning av hur klasstorlek bestäms teoretiskt). Det som styrde antalet elever i en klass var antalet barn i en viss ålder i ett visst skolområde. Ett skolområde har ofta bara en skola men kan också ha flera. Således berodde storleken på en elevs klass på det totala antalet elever som var inskrivna i samma skolområde för klass 4. Elever som var inskrivna i ett skolområde med 20–30 elever hamnade ofta i en stor klass medan elever i skolområde med 31–50 elever ofta hamnade i mindre klasser. Således var antalet elever i ett skolområde den variabel som avgjorde om en elev ”behandlades” eller inte. Tilldelning var dock inte deterministisk vilket tydligt framgår av streckade linjen i figur 3.2, som beskriver den faktiska klasstorleken som en funktion av det verkliga antalet elever i skolområdet.

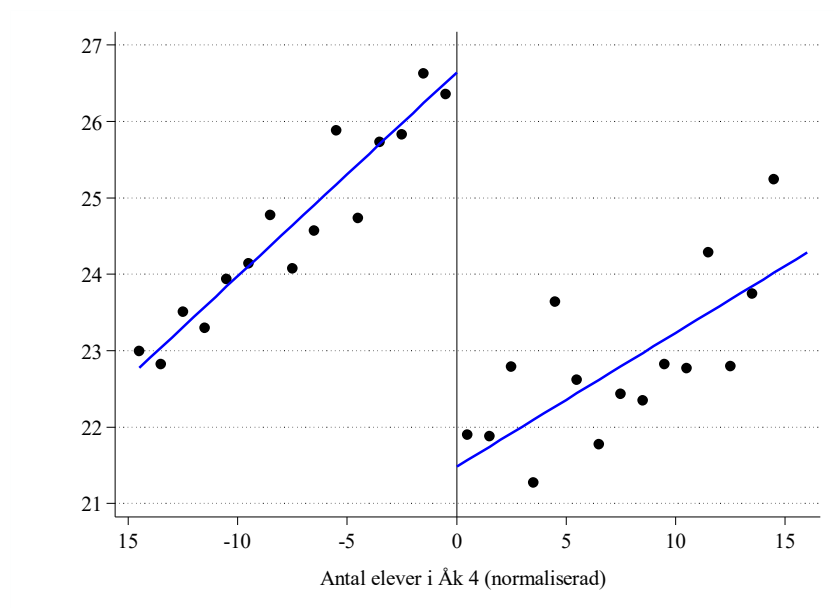
För att underlätta förståelsen för hur regeln fungerar i praktiken redovisas i figur 3.3 antalet elever i en klass när hänsyn tagits till ålderskohorten (det vill säga födelseår 1967, 1972, 1977 och 1982) och område, samt när tröskelvärdet (brytpunkten) normaliserats till noll. I figuren framgår att precis till vänster om brytpunkten (30, 60, 90, 120) har man i genomsnitt 26,5 klasskamrater medan man just till höger om brytpunkten har i genomsnitt 21,5 klasskamrater, det vill säga en minskning med cirka 5 klasskamrater om antalet elever precis överstiger 30, 60, 90 eller 120 barn i samma kohort.

Figur 3.4 presenterar samma grafiska analys för två utfallsvariabler: kognitiv förmåga vid 13 års ålder (vänster panel) och lön vid 27–42 års ålder (höger). Figuren visar att den kognitiva förmågan ökar från i genomsnitt 46,5 till 53 just vid brytpunkten, en ökning med 6,5 punkter, och att månadslönen ökar med 1 000 kronor, från 26 000 till 27 000.

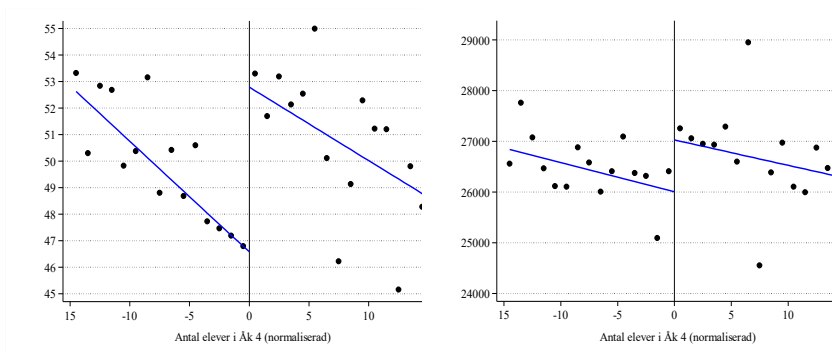
I tabell 3.1 redovisas effekterna av klasstorlek på samtliga studerade utfall. Resultaten ges i form av ökning (i procent eller procentenheter) av att minska klasstorleken med 5 elever och visar på positiva förändringar för samtliga utfall.

Figur 3.2 Förväntad och faktisk klasstorlek i åk 4–6

Källa: Fredriksson m.fl. (2012).

Figur 3.3 Regel för klasstorlek – samma brytpunkt

Källa: Fredriksson m.fl. (2012).

Figur 3.4 Kognitiv förmåga vid 13 års ålder (vänster) och lön vid 27–42 års ålder (höger)

Källa: Fredriksson m.fl. (2012).

Tabell 3.1 Genomsnittseffekter av att minska klasstorleken med fem elever

Utfall (enheter) (antalet observationer)	Effektskattning (standardfel)
1 Kognitiv förmåga vid 13 års ålder (procent av en standardavvikelse) [30 818]	+ 10,0 (4,8)
2 Icke-kognitiv förmåga vid 13 års ålder (procent av en standardavvikelse) [26 240]	+8,3 (4,8)
3 Standardprovresultat vid 16 års ålder (procent av en standardavvikelse) [30 784]	+9,9 (4,7)
4 Sannolikhet att ha åtminstone högskoleexamen vid 27–42 års ålder (procentenheter) [31 590]	+2,8 (1,8)
5 Relativ löneförändring vid 27–42 års ålder (procent) [17 499]	+3,4 (1,5)
6 Sannolikhet att ha årsinkomst vid 27–42 års ålder (procentenheter) [31 590]	+0,2 (1,1)

Anm.: Skattningarna avser IV-regressioner. Standardfel inom parentes. Skattningar som är signifikanta på 5%-nivån visas i fet stil; skattningar som är signifikanta på 10%-nivån är kursiverade; icke-signifikanta skattningar visas i normal stil. Standardfelen tillåter godtycklig korrelation mellan individer i samma kohort som bor i samma rektorsområde. Effektskattningarna är baserade på regressioner där klasstorlek i åk 4–6 instrumenteras med förväntad klasstorlek enligt klasstorleksregeln. Regressionerna innehåller också variabler som konstanthåller antalet elever i rektorsområdet i åk 4 (3:e gradens polynom), kommun (som skattas separat per kohort), elevens kön och födelsemånad, föräldrars utbildning och inkomst (3:e gradens polynom), moderns ålder vid barnets födelse, invandringsbakgrund, om föräldrarna separerat innan 10 års ålder, samt antalet syskon.

Källa: Fredriksson m.fl. (2012).

3.2.4 Rubins kausala modell (RCM)

Rubins kausala modell (hädanefter RCM) bygger på idén att forskaren med hjälp av data kan beskriva den process som ligger bakom selektionen av individer in i behandling. Genom att skapa en balans i kovariater mellan individer i respektive grupp i perioden före behandlingen kan det slumpmässiga experimentet återskapas i efterhand.

En nackdel med RCM är att vi, i motsats till det ideala slumpmässiga experimentet, inte kan vara säkra på att det inte finns icke-observerbara skillnader. Med andra ord, vi är inte säkra att vi har balans bland kovariater som inte observeras och som är betydelsefulla för att förklara skillnader i utfall. Problemet är därmed detsamma som vid traditionell regressionsanalys om syftet med analysen är att göra kausala utsagor. Det finns emellertid några viktiga skillnader mellan en RCM-ansats och en traditionell regressionsansats. En viktig sådan är att man i en RCM-ansats gör åtskillnad mellan design och analys. Detta åstadkoms genom att designen görs i frånvaro av utfallsdata. Poängen är att forskaren har en på förhand given plan, en så kallad föranalysplan, om hur en effekt ska skattas, det vill säga på vilka kovariater balans ska uppnås och hur den slutgiltiga analysen ska göras. Denna föranalysplan kan sedan med fördel göras allmänt tillgänglig innan forskaren har tillgång till utfallsdata. På samma sätt som i det randomiserade experimentet förhindrar det forskaren från att söka efter statistiskt säkerställda resultat, så kallat p-hackande, som är ett betydande problem inom empirisk forskning (se t.ex. Amrhein m.fl. 2019, och Wasserstein m.fl. 2019). I en regressionsansats görs design och analys i samma steg, dessutom tas i regel inte hänsyn till effektheterogenitet på samma sätt som i en RCM-ansats. RCM-ansatsen har sammantaget bättre förutsättningar att korrekt mäta det som avses.

Svagheten med RCM-ansatsen är att vi inte vet om designen också lyckas balansera de jämförda grupperna med avseende på faktorer som vi inte kan observera. En stor litteratur diskuterar olika analyser som kan genomföras för att pröva detta antagande (se t.ex. Rosenbaum 1999; Pearl 2000; Rosenbaum och Rubin 1984; Imbens och Rubin 2015; Imbens 2003; Ding och VanderWeele 2016).

Vi illustrerar RCM-metoden med utvärderingen av den så kallade rehabiliteringsgarantin (RG, 2012-16), som publicerades av

Hägglund m.fl. (2020). RG var en överenskommelse mellan regeringen och SKL om att landstingen ska erbjuda evidensbaserade rehabiliteringsinsatser i syfte att stärka den enskildes ställning i sjukskrivningsprocessen och öka återgången i arbete (Prop. 2007/08:1). De två insatserna var (1) Kognitiv beteendeterapi (KBT); För patienter med lindrig eller medelsvår psykisk ohälsa och (2) Multimodal rehabilitering (MMR); För patienter med långvarig smärta i nacke, skuldror, ländrygg eller generaliserad smärta som fibromyalgi. Dessa diagnoser utgör tillsammans ungefär 70 procent av alla pågående sjukfall. Insatserna omfattade såväl sjukskrivna som ej sjukskrivna.

I utvärderingen, som avgränsades till deltagare i Region Skåne, matchades patienter som fick KBT eller MMR inom RG med jämförelsepatienter som hade samma observerade hälsa och egenskaper men som inte fick samma behandlingar. Effekten av att få KBT eller MMR beräknas genom att jämföra utfallen (hälsa och sjukfrånvaro) mellan behandlade och icke-behandlade. Eftersom många olika faktorer kan tänkas påverka framtida hälsa och risk för sjukskrivning är metodens lämplighet beroende av i vilken utsträckning de data som finns tillgängliga fångar de eventuellt systematiska skillnader som finns mellan dem som får och inte får behandling, och att det, betingat av observerbar karaktäristik, finns en slumpmässighet i tilldelning av behandling. Att metoden är framgångsrik i det avseendet kan delvis, men inte till fullo, prövas med hjälp av de olika känslighetsanalyser som vi refererade till ovan.

I analysens första steg skattas sannolikheten att vid en viss tidpunkt påbörja behandling inom RG baserat på en mängd uppgifter om individens aktuella hälsotillstånd (uttryckt bland annat i tidigare vård- och läkemedelskonsumtion, diagnos med mera), ålder, utbildningsnivå, socioekonomisk situation (inkomst, gift/ogift), arbetslöshets- och sjukskrivningsstatus, arbetsmarknadsstatus, bostadskommun etc. Analysen ger en förväntad sannolikhet för varje individ att vid tidpunkten påbörja en behandling, en så kallad propensity score. Baserat på denna förväntade sannolikhet paras sedan en behandlad ihop med den icke-behandlade som är mest lik den behandlade. Det som gör att vi tror att föreliggande design är framgångsrik i att balansera behandlade och icke-behandlade även på information som inte finns tillgänglig är, 1) behandlingarna byggdes ut successivt. Eftersom kompetensen att utföra behandlingarna inte

fanns tillgängliga överallt i Skåne när RG initierades, skapades en variation i tillgängligheten till behandlingarna bland patienter med samma hälsa och i samma behov av behandling beroende på var patienten var bosatt, och 2) mycket detaljerade uppgifter om varje patient, inte minst beträffande hälsan.

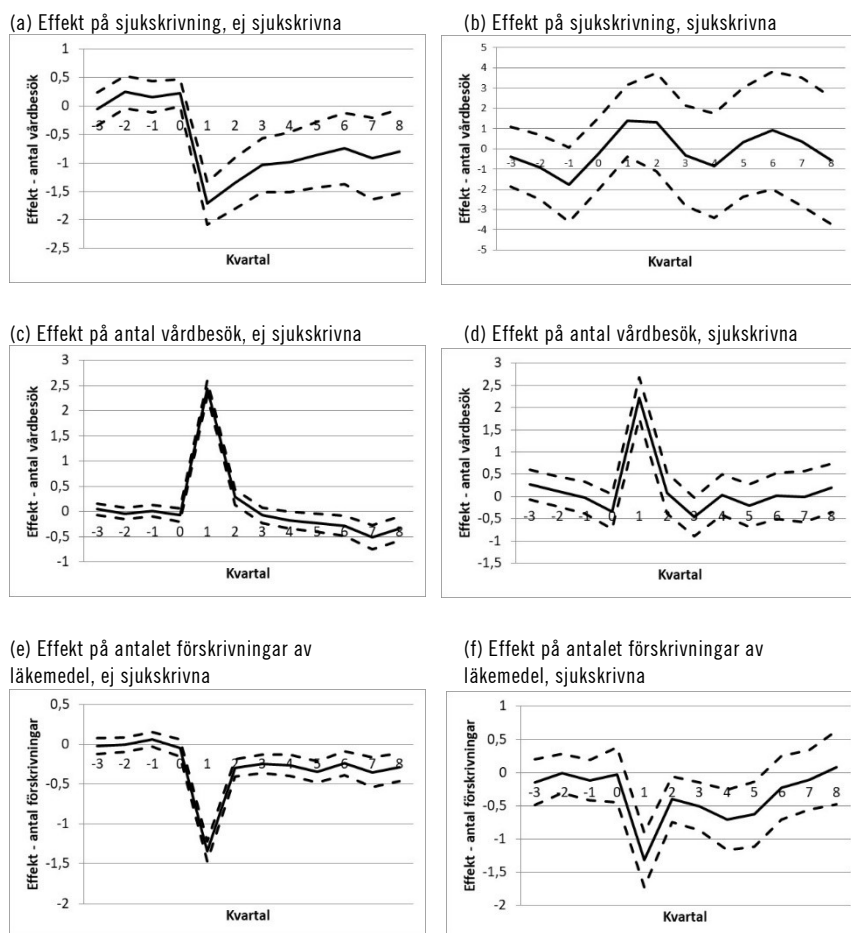
Figur 3.5 visar resultaten av KBT-behandlingen för sjukskrivna respektive ej sjukskrivna med avseende på i) sjukfrånvaro, ii) öppenvårdskonsumtion, och iii) läkemedelsförskrivning. Resultaten presenteras från fyra kvartal före behandlingsstart till åtta kvartal efter. Centralt är att inga skillnader mellan de jämförande grupperna framträder innan behandlingsstart. Situationen är således densamma som vid en väl designad DiD-analys. Noterbart är att grupperna inte alltid matchats med avseende på historiska utfall som studeras. Exempelvis har inte tidigare sjukfrånvaro använts i matchningen för gruppen ej sjukskrivna. Detta stärker trovärdigheten i design och analys. I rapporten presenteras också andra känslighetsanalyser som samtliga tyder på att dolda skillnader till följd av icke observerbara variabler inte är ett problem.

För individer som inte var sjukskrivna när behandlingen inleddes (figur 3.5 (a)) ledde KBT-behandlingen till en omedelbar minskning av antalet sjukskrivningsdagar på cirka 1,5 dagar. Trots en med tiden avtagande effekt återfanns en statistiskt säkerställd negativ skillnad kvar även efter åtta kvartal. För samma grupp visas i figur 3.5 (c) att antalet vårdbesök ökade under det första och andra kvartalet, detta är väntat eftersom en KBT-behandling i typfallet består av ungefär tio behandlingstillfällen. Från tredje kvartalet och framåt återfinns dock en statistiskt säkerställd minskning med ungefär 0,5 vårdbesök per kvartal. Från figur 3.5 (e) kan utläsas att antalet läkemedelsförskrivningar tydligt minskade som en konsekvens av KBT-behandlingen. Trots att nedgången är som störst under själva behandlingen kvarstår en statistiskt säkerställd minskning även efter behandlingens avslut och upp till åtta kvartal efter behandlingsstart. Sammantaget tyder resultaten på att KBT i preventivt syfte är en framgångsrik insats i så måtto att den minskar risken för framtida sjukfrånvaro och även i övrigt förefaller förbättra hälsan.

För KBT som insats till redan sjukskrivna är resultaten emellertid inte lika positiva. I figur 3.5 (b), 3.5(d) samt 3.5 (f) framgår att

behandlingen inte hade några långsiktiga effekter vare sig på sjukfrånvaro, antalet vårdbesök eller på antalet förskrivningar av läkemedel.

Figur 3.5 Effekter av KBT med avseende på antal sjukskrivningsdagar, vårdbesök och förskrivningar av läkemedel, streckad linje anger konfidensintervall på 5-procentnivån



Källa: Hägglund m.fl. (2020).

3.2.5 Summering

För att man med hjälp av en observationsstudie, det vill säga med utgångspunkt i befintliga data, ska kunna tro att en analys är kausal

krävs en design. En kvasiexperimentell design som formas från RCM (Rubin 1974) efterliknar det randomiserade experimentet. En kvasiexperimentell design ska: (1) visa att behandlade och obehandlade individer har samma fördelning av observerbara karaktäristika i analysdata, (2) ha en beskrivning av institutioner och regelverk för att troligöra att det finns en slumpmässighet i tilldelning av behandling, (3) ha en plan för hur analyser ska göras och (4) innehålla känslighetsanalyser för essentiella antaganden. Denna beskrivning ska helst skrivas ned före analysens genomförande i en så kallad föranalysplan, för att minska möjligheten till manipulation. Med de goda förutsättningar som finns att länka samman olika register är denna strategi sannolikt ofta bättre än att göra randomiserade experiment. Mängden information gör det i regel möjligt att på ett tillförlitligt sätt kontrollera för selektion till behandling. Strategin är också jämförelsevis billig.

En annan typ av kvasiexperimentell design använder sig av skillnader och förändringar i institutionella regler, väder, geografi etc., där tid och rum skapar skillnader i sannolikheten att bli behandlad på ett likartat sätt som vid det randomiserade experimentet. Dessa designer, som ofta benämns naturliga experiment, är inte avhängiga tillgången på detaljerade data och balans i kovariater i samma utsträckning som den kvasiexperimentella design som bygger på RCM.

3.3 Incitament och statistiska problem vid empiriska studier

Även om forskning inom samhällsvetenskap till stor del är offentlig-finansierad drivs den av egenintresse. Egenintresset består i att publikationer i tidskrifter leder till meritering och avancemang inom den akademiska hierarkin. Tidskrifterna vill ha läsare och prenumeranter och vill därför publicera artiklar med intresseväckande resultat. För de bättre tidskrifterna innebär detta väl genomförda studier (för att bibehålla trovärdighet) men också studier som visar på statistiskt säkerställda resultat. För forskaren kan det således, allt annat lika, vara enklare att få en uppsats med statistiskt säkerställda resultat publicerad än en som inte har det. Detta ger i sin tur drivkrafter för forskaren att försöka finna statistiskt signifikanta

resultat som oftast innebär att det skattade p-värdet är mindre än 0.05.³⁰ Detta förfarande att söka efter statistiskt säkerställda resultat benämns ”p-hackande” i litteraturen och har, på goda grunder, gjort skada för forskningens trovärdighet. Ioannidis (2005) hävdade att ”den mesta empiriska forskningen är felaktig när man drar slutsatser baserade på hypotestest”. Den så kallade ”Replikationskrisen” inom psykologi (Open Science Collaboration 2015), där man har visat att det är svårt att replikera resultat har lyft frågan högst upp på den vetenskapliga dagordningen. Replikationsgraden för laboratorieexperiment inom nationalekonomi är högre enligt Camerer m.fl. (2016).

Problemet består sannolikt inte till största delen av renodlat fusk utan beror snarare på en bristande metodologisk kompetens. Experimenten är ofta småskaliga och används för att hitta eller identifiera verkningsfulla åtgärder, snarare än att testa om väl genomtänkta och teoretiska motiverade åtgärder har en effekt. Som diskuterades i avsnitt 3.1.1 har småskaliga experiment små möjligheter att generera statistiskt säkerställda effekter, men på grund av en ökad risk för obalanser mellan behandlings- och kontrollgrupp kan de uppstå ändå. Om ett resultat från ett småskaligt experiment är statistiskt säkerställt måste den skattade effekten per definition vara stor, vilket gör resultatet mer intressant för tidskrifter att publicera. Emellertid är resultaten väldigt sällan möjliga att upprepa.

Ytterligare ett problem, som inte diskuterats i Ioannidis (2005), uppstår när det randomiserade experimentet inte kan blindas och när utfall mäts via enkäter, så som ofta är fallet inom psykologi men även inom socialmedicin, kriminologi och socialt arbete.³¹ Eftersom

³⁰ En statistik signifikant effekt, tolkad via konfidensintervall, innebär att konfidensintervallet inte täcker noll. Konfidensintervallet är emellertid godtyckligt. Vi måste bilda detta genom att välja en risknivå. Risken är att vi gör en felaktig tolkning, dvs. tolkar det som att det finns en effekt när ingen effekt finns. Om ett konfidensintervall som inte täcker noll kan bildas med en risk om r procent har vi ett p-värde som är r . I den ekonomiska litteraturen används oftast en risknivå på 5 procent. Innebörden är att om p-värdet är mindre än 0,05 så har vi ett statistiskt signifikant resultat. En tydlig indikation på att p-hackande förekommer inom nationalekonomi ges i Brodeur m.fl. (2016). Författarna visar att det finns en lägre andel vetenskapliga resultat publicerade i tidskrifterna *American Economic Review*, *Journal of Political Economy* och *Quarterly Journal of Economics* med p-värden mellan 0,25 och 0,10 än vad man borde förvänta sig, medan det finns fler resultat med p-värden precis under 0,05.

³¹ Ett exempel är en genomgång av ett antal experiment där multimodal rehabilitering (MMR), som innebär en kombination av psykologiska insatser och fysisk aktivitet, har studerats med avseende på sjukfrånvaro och återgång i arbete. Genomgången avslöjar allvarliga metodologiska brister. Förutom att experimenten i flera fall var så småskaliga att jämförbarheten mellan grupperna kan ifrågasättas, baserades de i regel på självrapporterade uppgifter vilket skapade

svaren i enkäten kan påverkas av vetenskapen om behandlingen är resultaten från dessa studier svårtolkade

När det gäller observationsstudier är problemet med "p-hackande" sannolik större än vid både randomiserade experiment och kvasiexperiment. Eftersom en observationsstudie inte går att replikera på samma sätt som ett randomiserat experiment är det emellertid svårare att analysera förekomsten av "p-hackande" i denna typ av studier (Ioannidis och Doucouliagos 2013).³²

Resultaten från Ioannidis (2005) och andra är nedslående men ska inte tolkas som att den empiriska forskningen med nödvändighet är sämre i dag än vad den har varit historiskt, snarare tvärtom. En ökad medvetenhet och noggrannhet i det empiriska arbetet har lett fram till denna kritik. Ett förslag för att minska problemet är att inte längre diskutera statistisk signifikans (Amrhein m.fl. 2019; Wasserstein m.fl. 2019). Detta synes vara som att kasta ut barnet med badvattnet eftersom beslutsfattare behöver en beslutsregel som kan ge vägledning om när det finns anledning att tro att en effekt existerar. En bättre lösning är att vid effektstudier publicera en analysplan som i detalj beskriver vilka analyser som ska göras. Dessutom bör större försiktighet vidtas vid publicering av resultaten från småskaliga experiment. Genom att använda bättre experimentella designer (exempelvis stratifierade randomiserade experiment³³) än den renodlade randomiseringen kan problemet reduceras ytterligare.

bortfall som i sin tur hanterades på ett metodologiskt felaktigt sätt. Därutöver tillkom sedan problemet med att såväl forskare som individer var informerade om experimentets genomförande (Johansson och Hägglund 2013).

³² Se Ioannidis m.fl. (2017) för en diskussion om trovärdigheten inom nationalekonomi.

³³ Om man tror sig veta att olika faktorer (exempelvis ålder och kön) är associerade med utfallet av intresse kan man göra det randomiserade experimentet inom varje strata. Detta kallas för ett stratifierat randomiserat experiment. I det enklaste fallet med en binär faktor (kovariat) har vi två strata. Säg att lön är ett utfall, i detta fall är det motiverat att randomisera behandling inom gruppen män och kvinnor separat.

4 Förekomsten av experiment inom medicin och ekonomi

Det finns en föreställning, åtminstone bland oss nationalekonomer, om att randomiserade fältexperiment som forskningsdesign har ökat över tid och att detta gäller inom framför allt utvecklingsekonomi. Denna föreställning förstärktes sannolikt efter Kungliga Vetenskapsakademiens beslut att dela ut 2019 års ekonomipris till Abhijit Banerjee, Ester Duflo, och Michael Kremer, detta för ”deras experimentella ansats för att mildra global fattigdom”. Många nationalekonomer skulle nog också hävda att skälet till den kraftiga ökningen av randomiserade fältexperiment är den ”revolution” inom framför allt arbetsmarknadsekonomi som pågått sedan slutet på 1970-talet, detta med avseende på identifikation via naturliga experiment eller kvasiexperiment.

Helt sannolikt har förekomsten av både randomiserade experiment och kvasiexperiment ökat över tid, men det betyder inte nödvändigtvis att de har ökat i betydelse för det politiska beslutsfattandet. Som diskuterats ovan är det svårt för en beslutfattare att bedöma kvaliteten i forskning. Med en allmän ökning av vetenskapliga alster ökar svårigheten att sälla ”agnarna från vetet”. En relevant frågeställning är således om andelen randomiserade respektive naturliga experiment har ökat som andel av den totala produktionen av forskningen. För att få en uppfattning om detta presenteras i detta avsnitt en strukturerad sökning av genomförda randomiserade experiment via ”Web of science”³⁴ åren 1989-2019, inom såväl ekonomi som medicin. I avsnitt 4.1 redovisas resultat när

³⁴ Innehåller alla tidskrifter som ingår i Science Citation Index Expanded (SCI-EXPANDED), Social Sciences Citation Index (SSCI), Arts & Humanities Citation Index (A&HCI), Conference Proceedings Citation Index- Science (CPCI-S), Conference Proceedings Citation Index- Social Science & Humanities (CPCI-SSH), Emerging Sources Citation Index (ESCI).

sökningar avgränsas till randomiserade experiment, i 4.2 redovisas resultaten när sökningen avser naturliga experiment. I avsnitt 4.3 avgränsas redovisningen till utvecklingen inom den svenska socialförsäkringen och arbetsmarknadspolitiken. Experimenten inom dessa områden presenteras i mer detalj i kapitel 5.

4.1 Randomiserade experiment

I det här avsnittet har sökningar gjorts på begreppen "Randomized/Randomised controlled trial", "Randomized/Randomised experiment", "Social experiment" och "Field experiment".³⁵

Inom ämnesområdet ekonomi hittades totalt 3 677 publicerade artiklar med ovan sökkriterier perioden 1989-2019. En uppdelning på de sex femårsperioderna (1989-93, 1994-98, 1999-2003, 2004-08, 2009-13 och 2014-19) visar en successiv ökning: 48, 90, 129, 328, 829 och 2 232. Ökningen motsvarar en ökning med 4 600 procent om man jämför den sista perioden med den första. Under perioden 1989-93 publicerades totalt 117 137 artiklar inom ämnesområdet ekonomi och under perioden 2014-2019 439 957 artiklar. Således utgjorde de experimentella studierna 0,04 procent under perioden 1989-93 och 0,51 procent under perioden 2014-19, en relativ ökning med 1 175 procent.

Inom ämnesområdet medicin hittade vi totalt 40 550 publicerade artiklar med något av sökorden under perioden 1989-2019. Fördelningen över de sex femårsperioderna är 398, 1 033, 2 884, 6 132, 9 621 och 20 482. Detta innebär en ökning med 5 100 procent över den studerade perioden. Sammanlagt publicerades nästan 3,6 miljoner artiklar under 30-årsperioden. Av dessa publicerades 369 685 under den första perioden och 1 174 386 under den sista. Således utgjordes 0,11 procent av artiklarna av experiment under den första perioden, och 1,74 procent under den sista perioden. En relativ ökning om 1 485 procent.

³⁵ Sökord: SU=(Economics OR Medicine) AND (TS=("Randomized Experiment") OR TS=("Randomized Controlled Trial") OR TS=("Randomised Experiment") OR TS=("Randomised Controlled Trial") OR TS=("Social Experiment") OR TS=("Field Experiment")) AND SU=(Medicine) AND (TS=("Randomized Experiment") OR TS=("Randomized Controlled Trial") OR TS=("Randomised Experiment") OR TS=("Randomised Controlled Trial") OR TS=("Social Experiment") OR TS=("Field Experiment")).

Av genomgången kan tre slutsatser dras, 1) de experimentella studierna utgör en väldigt liten andel av all vetenskaplig produktion inom ekonomi och medicin, 2) stora relativa ökningsar av förekomsten av randomiserade experiment har skett mellan 1989 och 2019 inom båda dessa ämnen, och 3) ökningen har, något överraskande för oss, varit större inom medicin än inom ekonomi.

De randomiserade experimenten kan klassificeras inom åtminstone en underkategori. I figur 4.1 fördelas experimenten inom ekonomi på de sex största underkategorierna: nationalekonomi/statistik (Economics), företagsekonomi (Business), organisation (Management), tillämpad psykologi (Psychology applied), hälso- och sjukvård (Health Care Science Service) och hälsopolicy (Health Policy Service). Totalt finner vi 6 589 träffar i 39 olika kategorier. Eftersom ämnesområdet ekonomi endast hade 3 677 artiklar innebär det att en enskild artikel i regel klassificeras i två olika kategorier. Från figuren kan vi se att omkring 75 procent av dessa 6 589 träffar under varje femårsperiod publiceras inom dessa sex kategorier.

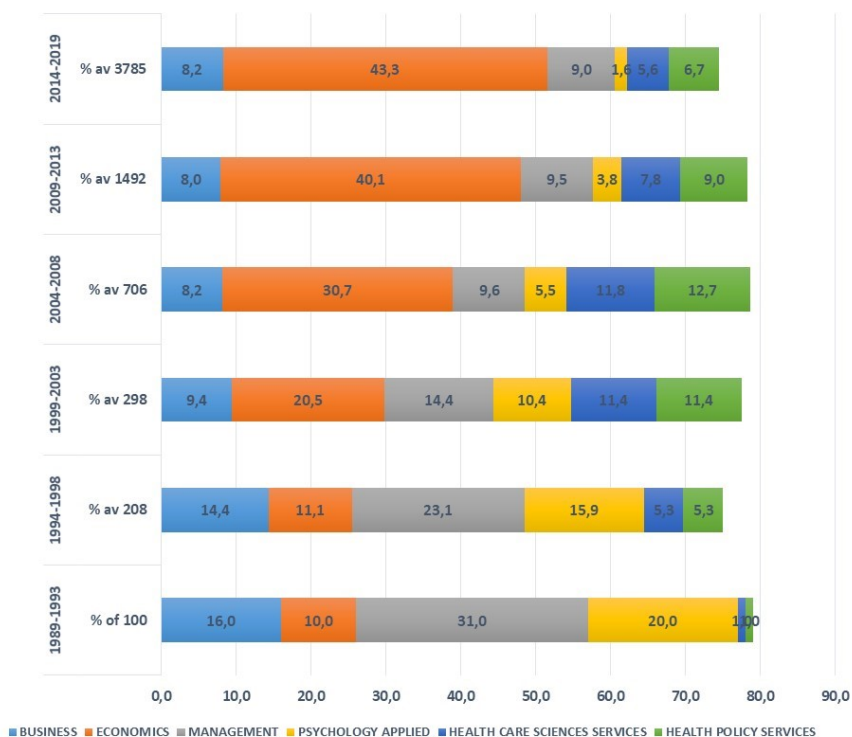
Från figuren kan vi se att andelen experiment inom området nationalekonomi/statistik vuxit tydligt över tid, från 10 procent under den första femårsperioden (av totalt omkring 100 artiklar) till 43 procent (av totalt 3 785 artiklar) under den sista femårsperioden. Intressant att notera är den över tid krympande andelen experiment inom tillämpad psykologi. En möjlig förklaring är den kritik som riktats mot svårigheterna i att upprepa experimentresultaten inom, primärt, psykologi (Open Science Collaboration 2015).

I figur 4.2 fördelas experimenten på fyra underkategorier till området nationalekonomi/statistik: utvecklingsekonomi (Development Studies), utbildningsekonomi (Education Educational Research), miljö- och resursekonomi (Environmental Studies) samt arbetsmarknadsekonomi (Industrial Relations Labor). De tre förstnämnda kategorierna har rönt ett växande intresse över tid såväl bland beslutsfattare som bland aktiva forskare. Skälet för att inkludera arbetsmarknad är att detta berör socialförsäkringsområdet (inklusive arbetslöshetsförsäkring).

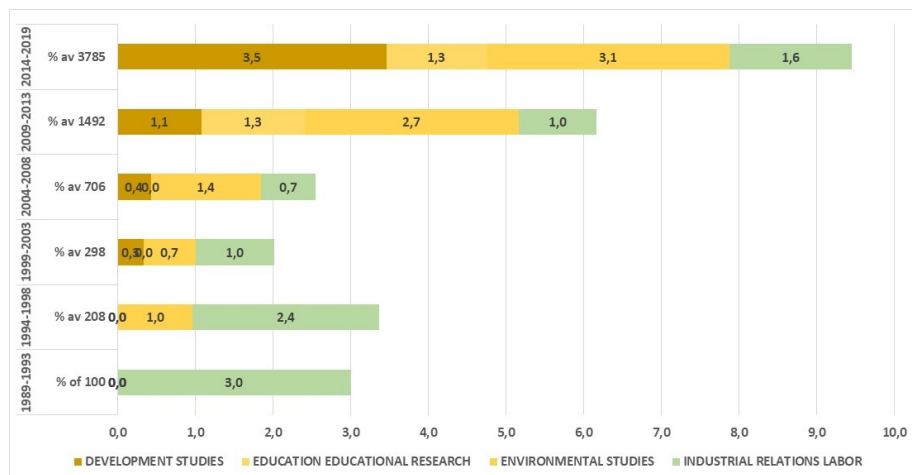
Under den första femårsperioden utgjorde dessa fyra kategorier endast 3 procent av alla publicerade experiment (eller 3 studier varav alla inom arbetsmarknad) medan i den sista mer än 9 procent av alla träffar. Antalet publicerade experiment inom arbetsmarknad har ökat över tid och uppgick till sex stycken under den sista perioden.

I relation till det totala antalet studier har dock antalet randomiserade studier inom arbetsmarknad minskat över tid. Intressant att notera är den stora ökningen av experiment inom utvecklings-ekonomi. Den första artikeln publicerades under perioden 1999-2003 och i den sista perioden publicerades 133 artiklar (3,5 procent av 3 785). Även inom miljöområdet kan vi se en kraftig ökning med 2 artiklar (1 procent av 208) publicerade perioden 1994-98 och 120 (3,1 procent av 3 785) artiklar under den sista perioden.

Figur 4.1 De randomiserade experimentens fördelning på delområdena företagsekonomi (Business), nationalekonomi/statistik (Economics), organisation (Management), tillämpad psykologi (Psychology Applied), hälso- och sjukvård (Health Care Science Service) och hälsopolicy (Health Policy Service) över tid



Figur 4.2 De randomiserade experimentens fördelning på delområdena utvecklingsekonomi (Development Studies), utbildningsekonomi (Education Educational Research), miljö- och resursekonomi (Environmental Studies) samt arbetsmarknadsekonomi (Industrial Relations Labor) över tid



4.2 Kvasiexperiment

I avsnittet redovisas litteratursökningar på orden "Natural Experiment", "Quasi Experiment" eller "Quasi-Experiment". Inom ämnesområdet ekonomi finner vi totalt 2 406 artiklar som uppfyller dessa villkor mellan 1989 och 2019. Fördelning över de sex femårsperioderna var 16, 70, 110, 234, 514 och 1 462, vilket motsvarar en ökning med 9 037 procent mellan sista och första perioden. Som andel av samtliga artiklar inom ämnesområdet har de kvasiexperimentella studierna ökat från 0,04 procent till 0,33 över tid.

Inom ämnesområdet medicin finner vi 277 artiklar som uppfyller sökkriterierna mellan 1989 och 2019. Fördelning över de sex femårsperioderna är 9, 8, 13, 34, 57 och 156, vilket motsvarar en ökning med 1 633 procent mellan sista och första perioden. Som andel av samtliga artiklar inom ämnesområdet utgjorde kvasiexperimenten 0,002 procent under den första perioden och 0,01 procent under den sista perioden.

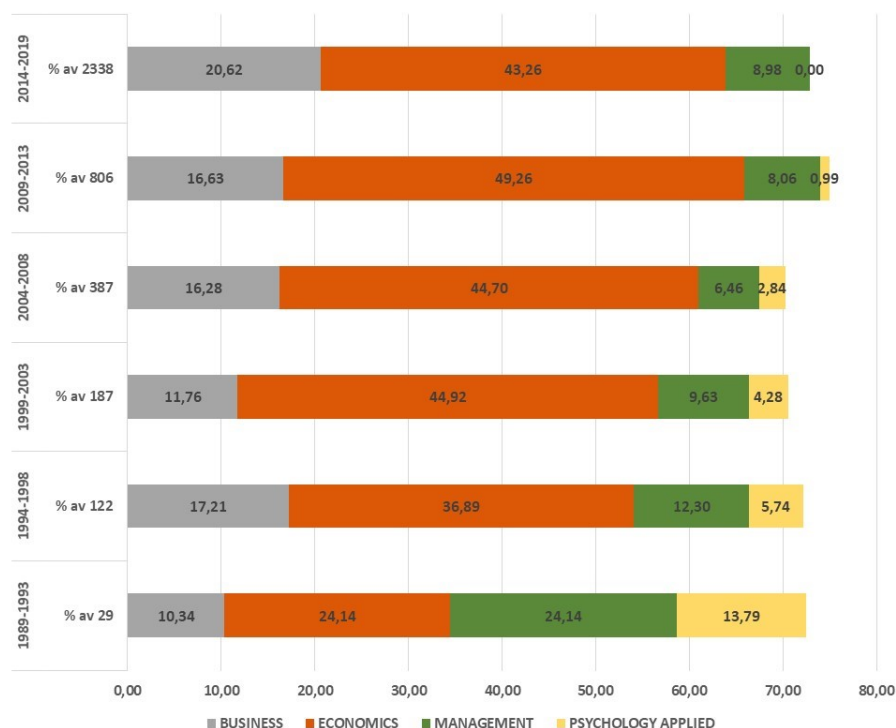
Det har därmed skett en dramatisk ökning av antalet kvasiexperimentella studier inom både ekonomi och medicin över tid, dock

utgör de fortfarande en mycket liten andel av samtliga artiklar inom ämnesområdena.

I figur 4.3 fördelas de publicerade kvasiexperimenten över tid och mellan de fyra största kategorierna inom ämnet ekonomi: nationalekonomi/statistik (Economics), företagsekonomi (Business), organisation (Management) och tillämpad psykologi (Psychology applied). Totalt återfinner vi 3 859 träffar inom 38 olika kategorier. Eftersom ämnesområdet endast hade 2 406 artiklar totalt innebär detta att en artikel i genomsnitt ingår i lite drygt 1,5 kategorier. Från figuren kan vi se att cirka 70–75 procent av alla träffar under varje femårsperiod publiceras inom dessa fyra kategorier.

Från figur 4.3 kan vi se att andelen publicerade kvasiexperiment i ämnet nationalekonomi/statistik ökar över tid. Även inom företagsekonomi och organisation är andelen växande. Vi kan också konstatera att andelen publicerade kvasiexperiment inom tillämpad psykologi, i likhet med de randomiserade experimenten, är sjunkande över tid.

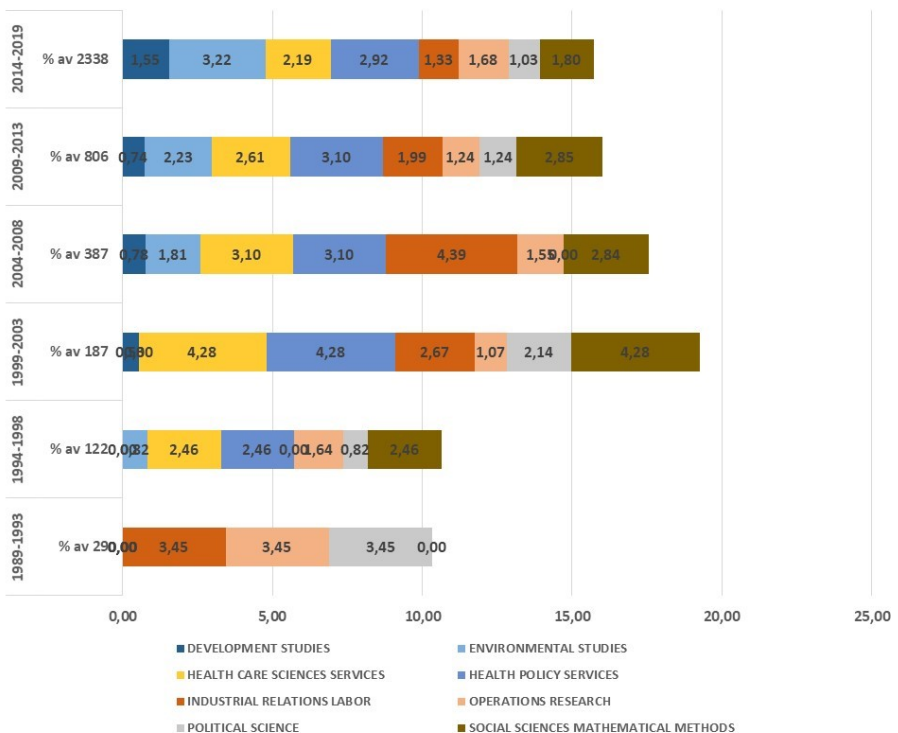
Figur 4.3 Kvasiexperimentens fördelning på delområdena företagsekonomi (Business), nationalekonomi/statistik (Economics), organisation (Management) och tillämpad psykologi (Psychology applied) över tid



I figur 4.4 fördelas de publicerade kvasiexperimenten mellan åtta mindre kategorier: Utvecklingsekonomi (Development Studies), miljö- och resursekonomi (Environmental Studies), arbetsmarknadsekonomi (Industrial relations labor”) hälso- och sjukvård (Health Care Science Service och hälsopolicy (Health Policy Service), statsvetenskap (Political Science), metoder inom sociala vetenskaper (Social Science Mathematical Methods) och operationsanalys (Operations Research). Skälet för att inkludera arbetsmarknad är att detta berör socialförsäkringsområdet (inklusive arbetslöshetsförsäkring). Skälet till att inkludera de övriga kategorierna är helt enkelt att de förutom de fyra kategorierna i figur 4.3 är de största kategorierna. Dessa åtta kategorier utgör från drygt 10 procent till knappt 20 procent av alla kvasiexperimentella studierna inom ämnesområdet ekonomi.

Intressant att notera är att andelen kvasiexperimentella studier inom arbetsmarknad minskat från 4,49 procent perioden 2004-08 till endast 1,33 under perioden 2014-19. Även här ser vi en dramatisk ökning av studier inom utvecklingsekonomi och miljö- och resursökonomi.

Figur 4.4 Kvasiexperimentens fördelning på delområdena Utvecklingsekonomi (Development Studies), miljö- och resursekonomi (Environmental Studies), hälso- och sjukvård (Health Care Science Service) och hälsopolitik (Health Policy Service), arbetsmarknadsekonomi (Industrial relations labor), operationsanalys (Operations research), statsvetenskap (Political Science) och metoder inom sociala vetenskaper (Social Science Mathematical Methods) över tid



Resultatet från kartläggningen är något förvånande, vi som författare hade bilden av att kvasiexperiment företrädesvis använts inom arbetsmarknadsforskning och att de metodologiska landvinningarna spritts till andra områden. En sannolik förklaring till resultaten är att

definition av kvasiexperiment i ”Web of science” är betydligt bredare än den definition som beskrivs i kapitel 3. Oavsett detta kan vi konstatera att den empiriska litteraturen där kvasiexperiment använts har ökat över tid men att den fortfarande utgör en väldigt liten andel.

4.3 Sökning på experiment inom socialförsäkring och arbetsmarknadspolitik

I nästa kapitel ges en kort beskrivning av randomiserade experiment som genomförts inom arbetsmarknadspolitik och socialförsäkringen i Sverige. I det här avsnittet görs en litteratursökning på såväl randomiserade experiment som kvasiexperiment inom dessa områden.³⁶

Totalt återfinns 20 publicerade artiklar kring randomiserade experiment inom arbetsmarknadspolitik och socialförsäkring mellan åren 1989 och 2019, dessa presenteras i tabell A.2 i Bilaga A. Den första artikeln publicerades 1995 men majoriteten av artiklarna är publicerade efter 2013. Noterbart är den låga citeringsfrekvensen dessa artiklar har, den artikel som har flest citeringar (25) publicerades 2004 (Google scholar ger 50 citeringar). Det är också intressant att notera att hela 25 procent av artiklarna är gjorda av forskare i Sverige och med svenska data.

När sökningen avgränsas till kvasiexperiment finner vi 37 publicerade artiklar, dessa presenteras i tabell A.3 i Bilaga A. Den första artikeln publicerades år 2000 och majoriteten är publicerade efter 2012. Även här är den låga citeringsfrekvensen slående. Artikeln som har flest citeringar publicerades 2000 och har 190 citeringar (Google scholar ger 775 citeringar). Nio av artiklarna berör kvasiexperiment genomförda i Sverige, det vill säga även här ungefär 25 procent av artiklarna.

Den generellt sett låga citeringsfrekvensen är nedslående, samtidigt är det kanske inte så förvånande. Resultaten är ofta kontextuella och kan av den anledningen vara svåra att tolka.

³⁶ Sökord: ”Social Insurance” eller ”Unemployment Insurance” eller ”Active labor market training” eller ”Active labour market training” eller ”labour market program” eller ”labor market program”.

Träffsäkerheten i dessa sökningar är inte perfekt, vi saknar artiklar som borde ha varit med bland de randomiserade experimenten och åtminstone två av de svenska artiklarna borde ha klassificerats som kvasiexperiment. Även vid sökningen på kvasiexperiment saknas några svenska artiklar. Under antagandet att svensk forskning inte är överrepresenterad bland dem som hittats i sökningarna är det intressant att notera att de svenska forskarna är väl representerade inom båda områdena. Sannolikt är tillgången till registerdata en stor fördel för att genomföra kvasiexperimentella studier. Det som kanske förvånar är att andelen randomiserade experiment som genomförts i Sverige är så hög. Att genomföra randomiserade experiment är i regel en komplicerad process, resultaten antyder att förutsättningarna för genomförande är relativt goda i Sverige, åtminstone inom dessa områden.

4.4 Summering

I kapitlet har redovisats i vilken utsträckning randomiserade experiment och kvasiexperiment förekommer bland vetenskapliga publiceringar inom medicin och ekonomi mellan 1999 och 2019. En slutsats är att även om publiceringar med experimentella ansatser överlag ökat till antalet över tid, utgör de fortfarande en mycket liten andel av det totala antalet studier inom dessa områden. Inom medicin var andelen randomiserade experiment och kvasiexperiment cirka 2 respektive 0,01 procent under perioden 2014-19. I ekonomi var motsvarande andelar 0,5 respektive 0,33 procent. En nedbrytning inom ämnet ekonomi visar att både randomiserade experiment och kvasiexperiment är relativt vanligast inom området nationalekonomi/statistik, där de utgör omkring 40 procent av de publicerade experimenten. Inom detta område ser vi att randomiserade experiment ökat framför allt i utvecklingsekonomi och miljöekonomi.

Sammantaget är bilden tudelad. Förekomsten av randomiserade experiment och kvasiexperiment har ökat tydligt över tid, vilket är positivt. Det visar inte bara på ett växande antal studier av hög kvalitet, det tyder också på en ökad metodkompetens i forskarsamhället vilket i sin tur torde ha en positiv inverkan på förutsättningarna för en mer evidensbaserad politik. Det som talar emot

detta är att studier med en experimentell ansats inte bara är påfallande lågt citerade, de utgör fortfarande en väldigt liten andel av samtliga studier som publiceras. Detta hänger delvis samman med en kraftig ökning av det totala antalet publicerade studier. Sammantaget kan det förväntas vara fortsatt svårt att skilja ut bra studier med goda förutsättningar att besvara frågeställningar av stort policyvärde, från mindre bra studier. Det ställer också krav på metodkompetens på mottagarsidan, vilket vi diskuterar närmare i kapitel 6.

5 Experiment inom den svenska socialförsäkringen och arbetsmarknadspolitiken

De senaste 15–20 åren har flera fältexperiment genomförts inom den svenska socialförsäkringen och arbetsmarknadspolitiken. I avsnittet ges en kort beskrivning av dessa experiment med avseende på bakgrund, utformning och resultat. Därefter görs en bedömning av experimenten med utgångspunkt i de för- och nackdelar som diskuterats tidigare i rapporten. För att underlätta läsningen har respektive experiment tilldelats ett namn och årtal, där året inom parentes avser året eller åren för experimentets genomförande. Studierna är antingen den svenska rapport där experimentet presenterades för första gången eller den vetenskapliga publikation som eventuellt följde.

5.1 Experiment inom socialförsäkringen

Göteborgs- och Jämtlandsexperimentet (1988)

Det enligt vår kännedom största experiment som genomförts inom socialförsäkringen genomfördes i Sverige 1988. Det var Försäkringskassan i Jämtlands län och Göteborgs kommun som tog initiativ till att pröva i vad mån längden på den sjukintygsbefriade perioden i början på ett sjukfall inverkade på hur länge individen var sjukskrivnen och på så vis kostnaderna för sjukförsäkringen. Bakgrunden till experimentet var att man ville påvisa att minskade krav i sjukskrivningsprocessen skulle leda till besparingar för samhället genom att minska antalet onödiga läkarbesök vilket skulle reducera kostnaderna för individen, sjukvården och staten. Visserligen förväntade

man sig att en del skulle stanna hemma längre som en konsekvens av förändringen, men detta skulle i sin tur reducera risken för framtida sjukskrivning.

En intressant detalj är att resultatet från experimentet endast publicerades i interna rapporter hos de lokala försäkringskassorna, och inte som en offentlig myndighetsrapport. Det sistnämnda hade sannolikt genererat mer uppmärksamhet. En möjlig orsak kan ha varit att resultaten inte var de som förväntades. Orsaken till att resultaten så småningom nådde allmänheten var att forskare vid Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU) i Uppsala fick kunskap om experimentet och kunde, utifrån Försäkringskassans register, återskapa experimentets behandlings- och kontrollgrupper.

Experimentet pågick under andra halvåret 1988. Liksom i dag var det upp till individen själv att avgöra behovet av sjukskrivning de första sju dagarna. Det räckte med att anmäla sjukdom till Försäkringskassan. Från och med den åttonde dagen krävdes ett intyg från läkare för att ha rätt till ersättning.

I experimentet ville man undersöka hur en förändring av tidpunkten för läkarintygskravet, från 7 till 14 dagar, skulle påverka sjukfrånvaron.³⁷ Vid tidpunkten fanns det ungefär 70 000 försäkrade, det vill säga ersättningsberättigade, i Jämtland och omkring 240 000 i Göteborg. Indelningen i behandlingsgrupp (som omfattades av en fjortondagarsperiod), och kontrollgrupp (sjudagarsperiod), gjordes på basis av den försäkrades födelsedatum. För de som var födda ett jämnt datum gällde 14 dagar utan läkarintyg medan de som var födda ett ojämnt datum omfattades av den mer strikta sjudagarsperioden.

I Hesselius m.fl. (2005) och Hartman m.fl. (2013) konstateras att förlängningen av den läkarintygsbefriade perioden från 7 till 14 dagar förlängde sjukskrivningarna med i genomsnitt en dag. Effekten är på intet sätt obetydlig, utifrån evidens kring de ekonomiska incitamentens betydelse (se Johansson och Palme 2005) motsvarar den en ökning av ersättningsnivån med 6 procent. Författarna konstaterar också att så mycket som 25 procent av den högre sjukfrånvaron i Sverige jämfört med Finland skulle kunna förklaras av den svagare kontrollen av den sjukskrivne i Sverige.

³⁷ Som enda län i landet tillämpades 14 dagars läkarintygsbefriad period redan i Jämtland. Där förkortades perioden i stället till sju dagar under experimentperioden.

En viktig aspekt i Göteborgs- och Jämtlandsexperimentet var att information om experimentet spreds till alla parter – den sjukskrivne, läkarna, arbetsgivaren och allmänheten – i förväg. Information spreds via planscher och broschyrer och även via massmedia. Kortare information om experimentet fanns också på den blankett som alla sjukskrivna fyllde i för att få ersättning från Försäkringskassan.

Att experimentet inte var blindat gav en unik möjlighet att analysera betydelsen av normer i sjukskrivningsbeteendet, det vill säga hur människor via sociala interaktioner påverkar och påverkas av omgivningens sjukfrånvaro. Att förstå betydelsen av interaktioner är potentiellt viktigt ur policysynpunkt, men också svårt att analysera eftersom det kräver information om vilka individen umgås med. Hesselius m.fl. (2008, 2013) använder sig av likheter i observerbara karaktäristika som indikator för interaktion, närmare bestämt födelseland. Strategin gick i ett första steg ut på att analysera hur korttidssjukfrånvaron inom respektive födelselandsgrupp påverkades av förlängningen av den läkarintygsbefriade perioden, för att i nästa steg använda denna information för att analysera relativa förändringar av sjukfrånvaron bland personer inom respektive födelselandsgrupp som inte fick läkarintygsbefriade period om 14 dagar, det vill säga kontrollerna.

Resultaten visade att ökningen av sjukfrånvaron bland dem som påverkades direkt av experimentet, det vill säga de som fick den läkarintygsbefriade perioden förlängd, spillde över till dem som inte påverkades av experimentet, det vill säga till kontrollgruppen. Eftersom även kontrollerna påverkades innebär detta att effekten av kontroll (i form av förändringar av den läkarintygsbefriade perioden) underskattats i Hesselius m.fl. (2005) och Hartman m.fl. (2013).

I Hesselius m.fl. (2009) analyseras hur den ökade sjukfrånvaron till följd av det senarelagda kravet på läkarintyg, påverkade arbetskollegornas sjukfrånvaro. I studien jämförs sjukfrånvaron före och efter experimentet för personer i kontrollgruppen, som alltså inte påverkades direkt av de temporärt förändrade reglerna. I studien utnyttjas det faktum att olika arbetsplatser hade olika fördelning av anställda i behandlings- och kontrollgrupp. Resultaten visar ett tydligt samband såtillvida att sjukfrånvaron bland kontrollpersoner anställda på arbetsplatser med en hög andel behandlade, ökade sin

sjukfrånvaro mer än kontrollpersoner anställda på arbetsplatser med en lägre andel behandlade. Man prövar också olika hypoteser om vilka bakomliggande mekanismer som påverkar resultaten och konstaterar att det mesta tyder på att resultaten uppkommer som en konsekvens av känslor kopplade till reciprocitet och rättvisa, eventuellt på grund av en ökad arbetsbörda då kollegor är frånvarande.

Resultaten från båda dessa studier har varit viktiga för att förstå Sveriges, i förhållande till andra länder, stora variationer i sjukfrånvaron, både över tid och mellan grupper. Med en social och dynamisk multiplikatoreffekt kan även mindre förändringar i exempelvis sjukförsäkringens regelverk få omfattande konsekvenser, inte främst via den direkta effekten av reformen utan via sociala interaktioner och förändrade normer i synen på användningen av försäkringen.

Informationsexperimentet (2005-06)

Flera studier inom den svenska sjukförsäkringen har påvisat förekomsten av överutnyttjande, eller *moralisk risk*, till följd av att den försäkrade inte bär hela sjukfrånvarokostnaden själv och dessutom kan förväntas ha ett informationsövertag beträffande möjligheterna att arbeta. Detta skapar incitament att utnyttja försäkringen mer än vad fallet skulle vara om den försäkrade bar hela kostnaden själv. I Johansson och Lindahl (2013) prövas ett förhållandevis billigt sätt att öka incitamenten i sjukförsäkringen, nämligen att kalla den försäkrade till ett obligatoriskt informationsmöte med syftet att tydliggöra skyldigheter och rättigheter i försäkringen.

Experimentet genomfördes i Uppsalas län 2005-06, i ett samarbete mellan Försäkringskassan och IFAU, och riktades mot 275 försäkrade med en svag förankring på arbetsmarknaden. En behandlingsgrupp av försäkrade kallades till informationsmötet medan försäkrade i en kontrollgrupp fick vänta en månad på en kallelse till samma möte. Resultaten visar att kallelsen till mötet förkortade tiden som sjukskriven med 65 dagar, eller drygt 20 procent. Eftersom själva deltagandet på mötet var lågt (ungefär 30 procent) tolkar författarna resultatet som att kallelsen i sig hade en drivande inverkan på benägenheten att avsluta sjukskrivningen. Mötet uppfattades således som en kontrollåtgärd av vissa försäkrade. Samtidigt

minskade deltagande i mötet sannolikheten att på kort sikt återgå i sjukskrivning.

VAB-experimentet (2006)

Ett experiment som fick stort medialt genomslag vid publicering var ett experiment inom den tillfälliga föräldrapenningen, närmare bestämt inom förmånen ”vård av sjukt barn” (VAB). Experimentet genomfördes under 2006 i ett samarbete mellan forskare på IFAU och Försäkringskassan, där syftet var att undersöka förekomsten av överutnyttjande. Bakgrunden till experimentet var ett uppdrag från den av regeringen tillsatta Delegationen för åtgärder mot felaktiga utbetalningar från trygghetssystemen (FUT-delegationen), vars uppdrag var att mäta omfattningen av felaktig användning av de offentliga försäkrings- och bidragssystemen och lämna förslag på att åtgärda den.

I experimentet utnyttjades det faktum att försäkrade som utnyttjar förmånen på ett felaktigt sätt kan förväntas förändra sitt uttagsmönster om risken för att bli påkommen ökar. Ett slumpmässigt urval om drygt 42 000 försäkrade mottog därför ett brev med information om att deras användning av förmånen skulle komma att kontrolleras under en period, samtidigt som uttaget bland de resterande cirka 1,3 miljoner försäkrade utgjorde jämförelse. Det totala överutnyttjandet beräknades sedan som skillnaden mellan grupperna med avseende på dels uttaget av förmånen, dels på andelen som via stickprovskontroller kunde konstateras använda förmånen felaktigt. Brevet varierades i utformning på så vis att vissa bara fick information om reglerna i förmånen medan andra fick information om att uttaget skulle kontrolleras, en tredje grupp fick information om både och.

Resultaten från experimentet presenterades i Engström m.fl. (2006), och visade att överutnyttjandet av VAB var betydligt större än det som tidigare, via stickprovskontroller, hade kunnat konstateras. I dessa beräknades överutnyttjandet uppgå till mellan sex och tio procent. Den aktuella studien uppskattar i stället överutnyttjandet till 22,5 procent, motsvarande drygt 650 miljoner kronor. Författarna menar att experimentet, genom att mäta de försäkrades egen

respons på informationen om kontroll, ger en mer korrekt bild av överutnyttjandet än vanliga stickprovskontroller.

Som en följd av dessa resultat infördes i juli 2008 ett krav på ett intyg om barns frånvaro från förskolan eller skolan innan VAB-ersättning kunde betalas ut. Intyget innebar dock omfattande administration för Försäkringskassan, föräldrar, förskolor och skolor samt bedömdes innebära för långtgående kontroll i förhållande till risken för missbruk. Intygskravet slopades därför från och med den 1 januari 2013, och i stället skulle Försäkringskassan införa slumpmässiga kontroller av barns frånvaro riktade mot förskolor och skolor. Försäkringskassan avsåg också att justera upp nivån på de befintliga kontrollerna riktade mot arbetsgivarna. ISF (2013) och Häkkinen-Skans och Johansson (2018) använder den slumpmässiga kontrollen för att skatta effekten inom den reguljära verksamheten. Resultaten ger stöd åt slutsatserna i Engström m.fl. (2006) om att det är lönsamt att öka kontrollinsatserna inom VAB till en betydligt högre nivå än vad den var 2013. Häkkinen-Skans och Johansson (2018) finner vidare att individer anpassar sin riskbedömning efter den nya kontrollnivån, vilket ska tolkas som att förändringar i graden av kontroll får långsiktiga konsekvenser för användningen av försäkringen.

Tidiga insatser-experimentet (2007)

Ett annat omfattande experiment i sjukförsäkringen genomfördes 2007, ännu en gång i ett samarbete mellan Försäkringskassan och IFAU. Syftet var att undersöka i vad mån ett tidigare agerande av Försäkringskassan i sjukskrivningsprocessen, i form av utredningar av den försäkrades arbetsförmåga, skulle generera en snabbare återgång i arbete. Experimentet omfattade 13 547 sjukskrivna och gick till på så vis att sjukskrivna födda ett jämnt datum, efter en inledande bedömning, skulle prioriteras till en utredning av behovet av rehabilitering medan de födda ett ojämnt datum fick vänta i minst sex veckor på samma utredning. Vidare delades behandlingsgruppen in i en slumpmässigt utvald grupp som skulle prioriteras till ett så kallat avstämningssmöte medan övriga fick vänta.

Utredningen skulle bara genomföras i de fall där det ansågs nödvändigt. På så vis erbjöds en stor flexibilitet vilken gjorde det enkelt

att implementera experimentet i den reguljära arbetsprocessen. Detta bidrog i sin tur till att experimentet kunde bli så omfattande som det blev. Experimentets unika design skapade en systematik i när under sjukfallet en viss åtgärd utfördes, vilken kunde utnyttjas för att analysera effekterna av ett tidigare agerande.³⁸

Resultaten i experimentet, som presenteras i ISF (2011) och Engström m.fl. (2015), visar något oväntat att ett tidigare agerande av Försäkringskassan förlänger sjukfrånvaron och ökar den framtida risken för mottagande av sjuk- och aktivitetsersättning (det som tidigare kallades förtidspension). Orsaken till detta är handläggarnas svårigheter att tidigt i sjukfallet avgöra för vilka en insats är nödvändig för att kunna återgå i arbete, och för vilka en insats snarare hindrar en återgång. För vissa skulle en utredning och en rehabiliteringsinsats kunna förstärka en sjukdomsidentitet som i sin tur försvårar en arbetsåtergång.³⁹ För andra, med en låg arbetsmotivation, kan en utredning och en insats fungera som en möjlighet att signalera dålig hälsa till Försäkringskassans handläggare och läkare.

Läkarintygsexperimentet (2007)

I Engström m.fl. (2009) undersöks huruvida en förstärkt kontroll från Försäkringskassans sida av de medicinska underlagen (läkarintygen), påverkar den sjukfrånvaro som utfärdas och den faktiska sjukfrånvaron som följer. Experimentet genomfördes 2007 och syftet var att undersöka om information om ökad kontroll av läkarnas roll i sjukskrivningsprocessen skulle ha en gynnsam inverkan på läkarintygens kvalitet och på sjukfrånvaron. I experimentet valdes två kommuner ut i varje län (förutom Gotland) varav den ena slumpmässigt valdes ut som behandlingskommun och den andra som kontrollkommun. Vårdcentralerna i behandlingskommunerna mottog ett brev med information om att Försäkringskassan under en tid skulle höja ambitionsnivån vad gäller läkarintygens kvalitet. Försäkringskassans handläggare ökade antalet kontroller i både behandlings- och kontrollkommuner men hade inte information om vilka kommuner som var behandlade.

³⁸ En närmare beskrivning av experimentet och de nämnda utredningarna återfinns i ISF (2011).

³⁹ Se Johansson och Lindahl (2012) för referenser kring denna forskning.

Resultaten i experimentet var mycket överraskande. De visar att den förstärkta kontrollen via en ökad förekomst av längre deltidssjukskrivningar (25 procents deltidsnivå), ökade den faktiska sjukfrånvaron. Vidare konstateras i en kompletterande studie (Goine m.fl. 2009), att kvaliteten i läkarintygen sjönk (!) som en konsekvens av informationen om kontroll. I signifikant fler ärenden begärde Försäkringskassan en komplettering av det medicinska underlaget, vilket är en möjlig delförklaring till de längre sjukfallen. Som möjlig förklaring till den försämrade kvaliteten i intygen framför författarna möjligheten att hälso- och sjukvården ser sin grindvaktsfunktion i sjukskrivningsprocessen som ett komplement till Försäkringskassans grindvaktsfunktion. När Försäkringskassan meddelar ökat fokus på kontroll skulle således de egna ansträngningarna att agera portvakt minska. I sammanhanget är det också viktigt att känna till att Försäkringskassan inte har några sanktionsmöjligheter gentemot läkare som utfärdar felaktiga eller bristfälliga medicinska underlag.

Bostadslotteriet (2003–13)

Bakgrunden till detta experiment är det svenska bruksvärdesystemet, det omdiskuterade systemet för hyressättning på den svenska bostadsmarknaden, vars syfte är att motverka segregering till följd av att boende med begränsade ekonomiska resurser inte har råd att bo i attraktiva bostadsområden. Syftet med Riksrevisionens granskning 2017 (Riksrevisionen 2017a) var att undersöka vilken effekt tillgången till en hyresrätt med lägre hyra än marknadshyra har för inkomstutvecklingen och arbetsutbudet. Till hjälp för att besvara frågeställningen utnyttjade man ett bostadslotteri som genomfördes i ett kommunalt bostadsbolag i Solna mellan 2003 och 2013. I lotteriet annonserades nya hyresobjekt ut på hemsidan varefter intressenter tilldelades ett slumptal med hjälp av ett datorprogram. Den person som fick slumptalet 1 erbjöds lägenheten, om den personen avböjde gick frågan vidare till nummer 2 och så vidare tills någon tackade ja. Alla i hela landet hade möjlighet att delta i lotteriet, för att bli godkänd som hyresgäst gällde emellertid att hushållet hade en stadigvarande inkomst. I underlaget för analysen ingick 638 lägenheter.

Resultaten visade att tilldelningen av en lägenhet hade en negativ effekt på arbetsinkomsten motsvarande 50 000 om året år 3, 4 och 5 efter lotteriet. Samtidigt konstaterades en positiv effekt på utbildningsnivån, som höjdes åren efter lotteriet. Eftersom den negativa arbetsinkomsten kvarstod även efter kontroll för att individen påbörjat studier drar Riksrevisionen slutsatsen att den förda bostadspolitiken inte ligger i linje med intentionerna om att politiken ska gynna arbete och tillväxt.

Rehabiliteringsexperimentet (2008-10)

En ständigt aktuell fråga i debatten är huruvida privat produktion av offentliga tjänster kan öka effektiviteten. Denna fråga undersöktes 2008 då Försäkringskassan och Arbetsförmedlingen fick i uppdrag att bedriva en försöksverksamhet med såväl privata som offentliga utförare av arbetslivsinriktad rehabilitering. Försöksverksamheten genomfördes i fyra regioner (Stockholm, Göteborg, Dalarna och Västerbotten) mellan 2008 och 2010, och avsåg långtidssjukskrivna och mottagare av sjuk- och aktivitetsersättning. Sammanlagt ingick drygt 4 000 försäkrade i experimentet och målet var att få deltagarna i någon form av arbete, subventionerat eller icke-subventionerat, eller utbildning. Försöket föregicks av ett upphandlingsförfarande där privata aktörer fick konkurrera om kontrakten. Ersättningsmodellen innebar att en privat aktör i bästa fall kunde erhålla 60 000 kronor per deltagare. Ersättning betalades ut vid insatsens start samt när deltagaren fick arbete respektive kunde konstateras vara kvar i arbete efter sex månader.

Resultaten från försöksverksamheten presenterades i Laun och Skogman Thoursie (2014) och visar att deltagarnas jobb- och utbildningsmöjligheter efter 1–2 år inte påverkades av huruvida aktören bakom rehabiliteringen var privat eller offentlig. Den enda skillnad som kunde konstateras var vilken typ av subventionerad anställning deltagarna fick, vilket dock inte hade några finansiella implikationer. Författarna drar därmed slutsatsen att det inte finns något som tyder på att införandet av privat tillhandahållen arbetslivsinriktad rehabilitering skulle ge betydande effektivitetsvinster.

”Supported employment”-experimentet (2014-15)

I takt med att antalet unga (19–29 år) med funktionsnedsättning, inte minst på grund av psykisk sjukdom, har vuxit kraftigt både i Sverige och inom OECD, har intresset för att hitta arbetsmetoder för att hjälpa denna grupp på arbetsmarknaden ökat. ”Supported employment” är en rehabiliteringsmetod där individen tidigt i rehabiliteringsprocessen, och med hjälp av en handläggare, ska komma i arbete och där handläggaren upprätthåller en nära kontakt med både individen och arbetsgivaren för att öka möjligheterna för en varaktig anställning. Idén är att en närmare kontakt ska minska den risk arbetsgivaren tar vid rekrytering av en person med osäker produktivitet.

Experimentet var ett samarbetsprojekt mellan Försäkringskassan och Arbetsförmedlingen och pågick under ett år 2014/2015 i 25 kommuner runt om i landet. Sammanlagt deltog 1 062 individer. Deltagande var frivilligt och föregicks av bedömningar av individens motivation och möjligheter att tillgodogöra sig rehabiliteringsinsatser. Lämpliga kandidater informerades i generella drag om experimentet och de som tackade ja fick underteckna en blankett om samtycke. Två kontrollgrupper utsågs till dem som slumpades till att bli föremål för Supported employment: den ena gruppen mottog vanliga rehabiliteringsinsatser med samma fokus på arbetsutfall som Supported employment men mindre personalintensivt, den andra gruppen blev föremål för ”Case management” vars främsta syfte var att öka individens välbefinnande och minska tiden som inlagd på sjukhus. Denna insats var lika personellt krävande om Supported employment men hade inte samma fokus på arbete som utfall.

Resultaten, som presenteras i bland annat Fogelgren m.fl. (2019), visar på tydligt positiva effekter av Supported employment på sannolikheten att vara i arbete efter 18 månader, och främst då subventionerat arbete. Emellertid fordras en längre uppföljningshorisont för att slutsatser om insatsens kostnadseffektivitet i förhållande till de reguljära rehabiliteringsinsatserna ska kunna dras.

Tandvårdsexperimentet (2015)

Riksrevisionen presenterade 2017 en rapport (Riksrevisionen 2017b) baserad på ett experiment inom tandvårdsstödet. Experimentet genomfördes i ett samarbete mellan nationalekonomiska institutionen i Uppsala och Försäkringskassan och syftade till att mäta överutnyttjande till följd av överbehandling. Det hade länge funnits en misstanke om att överutnyttjande förekom i tandvårdsstödet och att det kunde vara omfattande, emellertid kan det vara svårt att upptäcka eftersom gränsen mellan vad som är att betrakta som nödvändig behandling och vad som är vård utöver det, kan vara svår att definiera.

Experimentet omfattade 735 tandvårdsmottagningar (15 procent av alla mottagningar i landet) och sammanlagt drygt 800 000 unika patienter. 366 slumpmässigt utvalda mottagningar mottog ett brev med information om att deras tandvård med avseende på en viss åtgärd, åtgärd 801 (permanent tandstödd krona, flera i samma käke), skulle komma att kontrolleras i efterhand. Resterande 369 mottagningar mottog inte brevet men skulle likväl kontrolleras. Experimentets design byggde på att vårdgivare med kännedom om kontrollen i högre grad än övriga skulle agera i enlighet med tandvårdsstödet syfte. I så måtto att det fanns ett billigare och mindre omfattande alternativ till den kontrollerade åtgärden, skulle de informerade vårdgivarna således vara mer benägna att övergå till det. Det förändrade mönstret i den utförda tandvården skulle således ge en uppfattning om överbehandlingens omfattning i den specifika åtgärden.

I Riksrevisionen (2017b) konstateras att informationen om kontroll minskade användningen av den kontrollerade åtgärden med omkring 30 procent bland de 10 procent största vårdgivarna. Effekten motsvarade en kostnad för överbehandling på omkring 145 000 kronor per mottagning och år, en kostnad som fördelas lika mellan patienterna och staten. Samtidigt konstateras i rapporten en ökning i användandet bland samma vårdgivare av den mindre omfattande och mindre lönsamma åtgärden 800 (permanent tandstödd krona, en per käke). Författarna drar slutsatsen att resultaten tydligt visar att de privata vårdgivarna utför mer omfattande och mer lönsam tandvård än nödvändigt, samtidigt som billigare och mindre lönsamma alternativ finns tillgängliga.

5.2 Experiment inom arbetsmarknadspolitiken

Eskilstuna-experimentet (1975)

Det första kända experimentet inom arbetsmarknadspolitiken genomfördes på arbetsförmedlingen i Eskilstuna 1975. Ansvarig för genomförandet var Lennart Delander, licentiat vid högskolan i Växjö, och resultaten presenterades i SOU 1978:60. Syftet med experimentet var att undersöka effekterna av en utökad förmedlingsservice, i detta fall bestående av insatser riktade till såväl företag – i syfte att informera om de arbetssökande, påverka deras kravspecifiering av arbetskraften och möjliggöra subventionerade anställningar, som arbetssökande – i syfte att effektivisera och bredda deras arbetssökande och agera mellanhand i relationen till arbetsgivarna. Sammanlagt deltog 400 långtidsarbetslösa i experimentet.

Resultaten av de förstärkta insatserna var genomgående positiva, och betydande. Försöksgruppen uppvisade inte bara tydligt högre sysselsättning såväl i slutet på försöksperioden som vid uppföljningstillfället nio månader senare, kvaliteten i de förvärvade anställningarna, mätt som lön, anställningstrygghet och arbetsmiljö, var också högre. De kostnads- och intäktsanalyser som genomfördes uppvisar också genomgående mycket positiva utfall.

AMS-experimenten (2004)

Under 2004 sjösatte Arbetsmarknadsstyrelsen (AMS) en serie pilotverksamheter runt om i landet med syftet att bedriva intensifierad förmedlingsverksamhet för utsatta grupper av arbetslösa. Bakgrunden var ett regeringsuppdrag från 2002 där sex olika myndigheter fick i uppdrag att samverka på lokal nivå på orter och i regioner som drabbats av uppsägningar. AMS ambition var att identifiera behovet av nya insatser för särskilt utsatta på den regionala arbetsmarknaden. Insatserna skulle dokumenteras och utvärderas i syfte att finna goda exempel som kunde implementeras i den reguljära verksamheten. AMS tilldelades 20 miljoner extra inom ramen för förvaltningsanslaget som efter ansökningsförfarande skulle allokeras vidare till regioner som inkommit med förslag på hur deras försöksverksamhet skulle utformas och till vilka grupper av

arbetslösa den skulle riktas. I en handfull av projekten beslutades även att deltagarna skulle utses via slumpen och att en kontrollgrupp skulle utgöra jämförelse för att möjliggöra utvärdering av hög kvalitet.

Eftersom det vid denna tidpunkt var stort fokus på arbetslösheten bland unga valde flera projekt att rikta insatserna till denna grupp. I några fall gjordes ytterligare avgränsningar till långtidsarbetslösa unga (Östergötland) eller till unga akademiker (Uppsala). Valet av arbetsmetoder skiljde sig också åt men fokus låg på en kombination av jobbsökaraktiviteter, personlig vägledning, kontroll av sökbeteende och uppsökande arbete gentemot arbetsgivare. Deltagarna kunde även ha tillgång till program inom AMS reguljära serviceutbud, det vill säga dem som de arbetssökande i kontrollgrupperna var hänvisade till. Även om det uttalade syftet var att finna nya arbetsmetoder kan man i efterhand konstatera att aktiviteterna i typfallet var väl beprövade, dock genomförda med en större konsekvens och högre kvalitet. Insatserna var obligatoriska och i de flesta fall tidsbegränsade.⁴⁰

Experimentens design var rättfram, från den på förhand definierade målgruppen slumpades arbetssökande in i försöks- respektive kontrollgrupp. I takt med att arbetssökande fick arbete eller lämnade arbetslöshet av annan anledning upprepades randomiseringen, på så vis fylldes försöks- och kontrollgrupper successivt på. Randomiseringen skedde maskinellt och centralt från huvudkontoret.

Resultaten från fem av projekten publicerades i svenska rapporter vid IFAU (Hägglund 2006, 2009) och i vetenskapliga publikationer (Hägglund 2011, 2014). Effekterna av de intensifierade förmedlingsinsatserna analyserades med avseende på bland annat antal arbetslöshetsveckor, anställningsstatus och arbetsinkomst. En generell slutsats var att insatserna hade stärkt deltagarnas ställning på arbetsmarknaden. Fyra av fem projekt kunde uppvisa statistiskt säkerställda ökningar i utflödet till arbete och i tre av projekten återfanns positiva effekter på antingen sysselsättningsstatus eller arbetsinkomster (eller båda). På grund av att experimenten var relativt småskaliga, bara mellan cirka 300 och 1000 arbetslösa ingick i respektive experiment, var emellertid precisionen i skattningarna i regel låg.

⁴⁰ En närmare beskrivning av insatserna ges i Hägglund (2006).

Beräkningar av kostnader och intäkter inom arbetslöshetsförsäkringen respektive statsbudgeten uppvisade överlag positiva siffror.

En slutsats från utvärderingarna av insatserna var att de positiva resultaten inte enbart var ett resultat av insatserna i sig. En grov uppskattning visade att endast mellan 60-75 procent av de arbetsökande i försöksgrupperna i någon grad deltog i insatserna, resterande andel hann lämna öppen arbetslöshet under de 3-6 veckor som passerade mellan kallelse till insatserna och insatsernas start. I Hägglund (2014) konstateras att enbart kallelsen till insatserna tydligt påskyndade utflödet från arbetslösheten i ett av projekten.

Anvisningsexperimentet (2007)

Hösten 2007 genomfördes ett experiment i samarbete mellan Arbetsförmedlingen, som stod för utförandet, och IFAU, som stod för utformningen av experimentet. Bakom experimentet stod den tidigare nämnda FUT-delegationen. Syftet var att undersöka överutnyttjande av ersättning kopplad till arbetslöshet (arbetslöshetsersättning och aktivitetsstöd) till följd av att arbetssökande inte söker de jobb som Arbetsförmedlingen anvisar dem. Arbetet med platsanvisningar har både ett matchande och kontrollerande syfte.

I undersökningen informerades en slumpmässigt utvald grupp arbetssökande som mottagit en platsanvisning om att deras ansökan med stor sannolikhet skulle komma att följas upp hos arbetsgivaren, det vill säga arbetsgivaren skulle kontaktas för att bland annat kontrollera att ansökan kommit in. En kontrollgrupp mottog inte denna information men genomgick samma kontroll. Resultaten visade i korthet att en hög andel, 30 procent, av dem som enligt reglerna var skyldiga att söka anvisat arbete inte gjorde det. Vidare hade informationen om kontroll endast en marginell inverkan på benägenheten att söka arbetet. En möjlig förklaring, enligt författarna, kan vara att hotet om sanktion inte upplevs som särskilt trovärdigt. Trots att en förhållandevis hög andel (13-14 procent) av de anvisningar som söktes också ledde till ett erbjudande om arbete, tycktes således en högre kontroll av förmedlade anvisningar inte ha någon inverkan på arbetslösheten.

Förmedlingsexperimentet (2015)

Under 2015 pågick en storskalig försöksverksamhet på många av landets arbetsförmedlingar i syfte att undersöka betydelsen av tidigare och aktivare förmedlingsinsatser i form av fler möten mellan arbetslösa och arbetsförmedlare. Sammanlagt ingick 72 slumpmässigt utvalda lokala förmedlingskontor under sex månaders tid i försöket, varav hälften av de nyinskrivna arbetssökande på hälften av dessa kontor slumpmässigt valdes ut för de utökade insatserna. Arbetssökande i kontrollgruppen var hänvisade till Arbetsförmedlingen reguljära verksamhet. Experimentet var utformat på sådant sätt att man på ett trovärdigt sätt skulle kunna analysera insatsernas *undanträngningseffekter*, det vill säga det faktum att ökade jobbchanser för vissa jobbsökanden (deltagarna) skulle kunna uppstå på bekostnad av försämrade jobbchanser för andra (icke-deltagarna). Därutöver slumpades deltagarna in i olika behandlingsgrupper för att göra det möjligt att undersöka effekterna av olika typer av utökade förmedlingsinsatser.

Resultaten presenteras i Cheung m.fl. (2019) och visar att insatserna hade tydligt positiva effekter för deltagarna, vars jobbchanser ökade med 10 procent under de första tre månaderna. Emellertid var undanträngningseffekterna betydande. När hänsyn tagits till de negativa effekter insatserna hade för icke-deltagarna kvarstod en visserligen positiv, men betydligt mindre, effekt motsvarande 0,7 procent. Författarna konstaterar att en fullskalig implementering av den förstärkta förmedlingsverksamheten skulle minska arbetslösheten med 0,2 procentenheter. Vidare konstateras att individuella möten har mer positiva effekter än gruppmöten, samt att distansmöten via telefon eller webb kan vara lika effektiva som fysiska möten. Avslutningsvis konstateras att en viktig förklaring till de positiva resultaten är den information om lediga platser (platsanvisningar, platsförslag) som förmedlas av förmedlarna till de arbetssökande.

5.3 Diskussion och summering

I avsnittet diskuteras de femton experimenten i föregående avsnitt med utgångspunkt i den perfekta utvärderingen, det vill säga den

ideala experimentdesignen (avsnitt 5.3.1) respektive den ideala analysen (avsnitt 5.3.2). En summering ges i avsnitt 5.3.3.

5.3.1 Experimentdesign

I avsnitt 3.1.2 diskuterades det ideala experimentet och de problem som följer av att experiment kan skapa beteendeeffekter som kan störa analysen. Frågan är hur nära det ideala experimentet ovan experiment kommer och i vilken utsträckning beteendeeffekter kan ha orsakat problem.

I majoriteten av experimenten har försökspersonerna inte kännedom om att experimentet genomförts, det vill säga experimentet har varit blindade för försökspersonerna. Det betyder att försökspersonernas beteende inte har förändrats som en konsekvens av att de observerats. Undantagen är Göteborgs- och Jämtlandsexperimentet (1988), "Supported employment"-experimentet (2014-15) och möjligtvis Eskilstuna-experimentet (1975). Även vid blindade experiment kan dock oönskade spridningseffekter förekomma. Exempelvis kan en arbetsmarknadspolitisk åtgärd reducera möjligheten att få jobb för dem som inte får ta del av åtgärden och som ingår i kontrollgruppen. Resultatet från åtgärden blir då positiv för dem som deltar men kan helt utebli, eller till och med bli negativ, på samhällsnivå eftersom de positiva resultaten sker på bekostnad av dem som inte deltar. Intressant att notera är att både Tandvårds-experimentet (2015) och Förmedlingsexperimentet (2015) explicit tar hänsyn till potentiella överspillningseffekter mellan försökspersoner som kan uppkomma på grund av experimentet. Den senare studien är även designad för att kunna beräkna storleken på dessa effekter.

Vad gäller Göteborgs- och Jämtlandsexperimentet (1988) visade Hesselius m.fl. (2008) och Hartman m.fl. (2013) att även individerna i kontrollgruppen ökade sin sjukfrånvaro som en följd av att experimentet genomfördes. Det innebär att den ökning av sjukfrånvaron som blev följden av den förlängda läkarintygsbefriade perioden i själva verket underskattades i deras analyser. "Supported employment"-experimentet (2014-15) föregicks av informerat samtycke bland experimentdeltagarna och kan därför potentiellt lida av beteendeeffekter. Dock riktade sig insatserna till personer mycket

långt från arbetsmarknaden (unga med aktivitetsersättning/förtidspension), vars alternativ till att delta i experimentet och till att förändra beteendet som en konsekvens av experimentet i sig, torde ha varit begränsade. Detta talar för att problemet med att experimentet inte var blindat kanske inte var så stort. Vad gäller Eskilstuna-experimentet (1975) är informationen kring hur experimentet genomfördes bristfällig varför vi inte vet om individerna informerades om att de deltog i ett experiment.

Oönskade beteendeeffekter kan också uppstå om experimentet inte blindats för dem som är satta att utföra det. Handläggare med vetskap om att de observeras skulle exempelvis kunna vara extra angelägna om att göra ett bra jobb. Det skulle också kunna vara så att handläggare hyser tilltro till den insats som ska studeras och därför anstränger sig extra för att den ska visa sig generera positiva effekter för de behandlade. I båda dessa fall finns en uppenbar risk att de observerade effekterna inte är representativa för de effekter som skulle uppstå vid reguljärt utförande. På motsvarande sätt skulle ansvariga handläggare för de icke-behandlade kunna påverka experimentet genom att erbjuda dessa samma eller liknande insatser som de behandlade. Detta benämns *substitutionsbias* i utvärderingslitteraturen. I en översikt av amerikanska arbetsmarknadspolitiska experiment genomförda mellan 1975 och 1992 visar Heckman m.fl. (1999) att andelen kontrollgruppspersoner som fick ta del av försöksverksamheterna i vissa fall var så hög som 40 procent.

Med undantag för AMS-experimenten (Hägglund 2009, 2014) diskuteras inte potentiella problem med handläggareffekter explicit. Hägglund (2009) konstaterar att en okänd andel av de icke-behandlade erbjöds insatser som liknade de behandlades. Detta kan sannolikt ha förekommit även i några av de andra experimenten. Risker torde vara som störst i de experiment där det fanns incitament inom organisationen eller myndigheten att uppvisa positiva resultat, i detta sammanhang främst Rehabiliteringsexperimentet (2008-10) och "Supported employment"-experimentet (2014-15). I Rehabiliteringsexperimentet, som analyseras av Laun och Skogman Thoursie (2014), konstateras att deltagarnas jobb- och utbildningsmöjligheter efter 1–2 år inte påverkades av huruvida aktören bakom rehabiliteringen var privat eller offentlig. Det är fullt möjligt att handläggare vid Arbetsförmedlingen såg de privata initiativen som ett hot och därför ansträngde sig extra för att visa positiva resultat.

I Fogelgren m.fl. (2019) konstateras tydligt positiva effekter av de insatser som erbjöds inom ramen för "Supported employment" på sannolikheten att vara i arbete efter 18 månader, och då främst i subventionerat arbete. Eftersom Arbetsförmedlingen i hög grad kan påverka vilka som får en subventionerad anställning kan handläggarna även här ha agerat på ett sätt som gynnade dem som ingick i behandlingsgruppen.

Det kan även finnas andra typer av beteendeeffekter som kan störa analysen, exempelvis det faktum att inte alla som slumpas in i behandling till fullo tar del av den, eller ens alls. Inom forskningslitteraturen används begreppen "no shows" för att beskriva försöksgruppsdeltagare som inte tar del av behandlingen överhuvudtaget, och "drop outs" för dem som bara tar del av delar av behandlingen. Eftersom avisering om ett experiment sker med viss framförhållning finns en risk att tänkta försöks- och kontrollgruppsdeltagare hunnit lämna, i detta fall sjukskrivning eller arbetslöshet, innan behandlingen inletts. På motsvarande sätt finns ingen möjlighet att hålla kvar behandlade och icke-behandlade i sjukskrivning och arbetslöshet bara för att ett experiment pågår. Det statistiskt korrekta sättet att hantera detta bortfall⁴¹ är att inkludera samtliga experimentdeltagares utfall i analysen, oavsett om de deltagit eller inte. Anledningen är att det endast är den ursprungliga gruppen av behandlade och icke-behandlade som är jämförbara. Snarare än effekten av behandlingen i sig mäts då intentionen att behandla ("intent to treat" i forskningslitteraturen). Hur stort problemet med "no shows" och "drop outs" är beror dels på vilken typ av insats som studeras, dels på vilken målgrupp insatsen riktar sig till. Om insatsen upplevs som krävande och meningslös kan den ha en avskräckande inverkan – arbetslösa och sjukskrivna kan (om möjlighet finns) föredra att återgå i arbete framför att ta del av behandlingen. I Heckman m.fl. (1999) finner man att andelen deltagare i försöksgrupperna ofta understiger 70 procent och ibland är lägre än 50 procent.

Problemet med no shows och drop outs var relativt utbrett bland de beskrivna experimenten. I AMS-experimenten mottog som sagt bara 60–75 procent av deltagarna behandling och i ett av experimenten kunde en tydlig anvisningseffekt konstateras. I Anvisnings-

⁴¹ Notera att detta bortfall inte är ett bortfall i statistisk mening, i så måtto att information om individerna i fråga saknas.

experimentet (2007) skickades aviseringar om kontroll av sökaktivitet ut till 1 502 individer, av dessa var endast 1 235 (82 procent) arbetslösa när väl aviseringen kom fram (1 240 av 1 485 i kontrollgruppen). I Förmedlingsexperimentet (2015), som analyserar effekterna av intensifierade förmedlingsinsatser bland arbetslösa, var det endast 23 procent som tog del av de faktiska insatserna.

5.3.2 Analyser

Som diskuterades i kapitel 3 finns alltid en möjlighet att ett enskilt experiment genererar statistiskt säkerställda resultat trots att ingen effekt finns. Risken för detta är större i småskaliga experiment eftersom dessa oftare innehåller obalanser mellan behandlade och icke-behandlade. Detta trots den randomiserade tilldelningen av behandling. Eftersom det är lättare att få statistiskt säkerställda resultat publicerade utgör småskaliga experiment ett stort problem. Vidare visar Ioannidis (2005) och Open Science Collaboration (2015) att resultat från småskaliga experiment ofta inte går att upprepa. Inga av de ovan nämnda experimenten är riktigt småskaliga. Det minsta experimentet, Informationsexperimentet (2005-06), bestod av 275 individer varav hälften i behandlingsgruppen. Den del av Förmedlingsexperimentet (2015) där undanträngningseffekter studeras kan potentiellt betraktas som småskaligt eftersom det endast omfattar 36 lokala förmedlingskontor. Detsamma gäller Läkarintygsexperimentet (2007) där behandlingen i form av kontrollinsatser omfattade 40 "tvillingkommuner" inom varje landsting. Antalet individer som behandlas är dock stort i båda fallen och eftersom analyserna omfattar individdata är detta en viss fördel gentemot om endast aggregerade utfall varit tillgängliga.

Som diskuterades i avsnitt 3.2 finns en risk att forskare genom att godtyckligt välja vilka kovariater som ska inkluderas i analysen, aktivt söker ett statistiskt säkerställt resultat, så kallat p-hackande. I princip alla av de genomgångna studierna använder sig av någon form av regressionsanalys i analysarbetet (Eskilstuna-experimentet, 1975, och Informationsexperimentet, 2005-06, är undantagen). Sannolikt är p-hackande inte något stort problem i de redovisade studierna. I de flesta studier presenteras resultat också utan justering för kovariater vilka i typfallet ger samma kvalitativa resultat. Dessutom

är problemet med kovariatjustering som störst vid småskaliga experiment.

Forskare har också drivkrafter att lära sig så mycket som möjligt av experimentet, det vill säga undersöka effekter av behandlingen med avseende på ett flertal utfall. Därutöver kan det vara intressant att studera effekterna för olika delgrupper, exempelvis för män och kvinnor separat, unga och gamla respektive hög- och lågutbildade etc. Med en mängd analyser och hypoteser att pröva kommer vissa med säkerhet att bli statistiskt säkerställda, detta benämns i analys-sammanhang för multisignifikansproblem. Problemet kan åtgärdas genom att justera nivån för vad som kan anses vara statistiskt säkerställt (det vill säga vilken signifikansnivå som väljs).

Om forskaren inte redan på förhand har definierat vilka utfall som ska studeras eller vilka delgruppsanalyser som ska göras har denne möjlighet att välja ut de analyser där man finner statistiskt säkerställda resultat. Detta kan likställas med p-hackande. Eftersom inga av de redovisade experimenten har presenterat någon föranalysplan kan vi inte bedöma om forskarna valt ut specifika utfall eller delgrupper för att stödja sina hypoteser. Vid genomgång av experimenten finner vi att merparten endast studerar ett utfall. I de fall där multipla utfall studerats har man dock inte tagit hänsyn till multisignifikans i de statistiska analyserna. Vid delgruppsanalyser har man inte heller tagit hänsyn till att många statistiska tester genomförs. I merparten av studierna redovisas endast ett fåtal delgruppsanalyser varför problemet inte är stort i gemen. I Förmedlingsexperimentet (2015) genomförs dock mer än 50 olika tester och detta också med olika utfall.

Att hantera multisignifikans är ett delikat problem. Om vi tar hänsyn till att en stor mängd tester ska göras genom att korrekt justera signifikansnivån kommer vi i princip aldrig kunna hitta några effekter av en intervention. Eftersom p-hackande och multisignifikans är återkommande problem inom den empiriska forskningen bör dock forskarna inom området bli mer tydliga med vad som utgör huvudresultat, och där statistiska tester kan genomföras, och vad som utgör explorativa eller hypotesgenererande resultat, där inga formella statistiska tester ska genomföras. När så är möjligt ska forskarna i en föranalysplan presentera vilka utfall och vilka regressionsmodeller som ska användas i analysen, och detta helst innan utfallsdata finns tillgängliga för analys.

5.3.3 Summering

En sammanfattande slutsats är att de senaste decenniernas fält-experiment inom social- och arbetslöshetsförsäkringen varit av hög kvalitet, åtminstone på så vis att de med hög trovärdighet lyckats mäta effekterna av aviseringen av en insats eller behandling, det som benämns avsikten att behandla eller "intent to treat" i forskningslitteraturen. Detta är den mest policyrelevanta parametern i så måtto att den bäst överensstämmer med den effekt vi kan förväntas få av insatsen vid reguljärt utförande. Genomgången visar emellertid att det i de fall där syftet varit att studera effekten av en aktiv insats (dvs. inte effekten av en kontrollinsats) ofta varit svårt att härleda mekanismerna bakom effekten, det vill säga vilka delar av en given behandling som varit verkningsfull.

Den mest genomarbetade experimentdesignen är Förmedlings-experimentet (2015) som, vid sidan av ansatsen för att mäta effekterna av en arbetsmarknadspolitisk insats, tillämpat en design som även medger möjligheten att analysera undanträngningseffekter, det vill säga i vad mån en ökning av jobbchanserna för deltagarna i insatsen sker på bekostnad av försämrade jobbchanser för dem som inte deltar. Experimentet, som analyseras i Cheung m.fl. (2019), visar emellertid på svårigheterna med fältexperiment i praktiken. Faktum är att det finns problem med båda nivåerna i analysen. Vad gäller behandlingseffekten består problemet i att endast 23 procent av de som tilldelades behandlingen i själva verket mottog den. Via kompletterande analyser konkluderar forskarna att den 10-procentiga ökningen av jobbchanserna härstammar från att de arbets-sökande fick information om lediga jobb som de också sökte. Med tanke på att endast 23 procent kunde få denna information innebär det att den faktiska effekten av insatsen för dem som mottog den var drygt fyra gånger så stor, det vill säga drygt 40 procent. Något förvånande bortser man i analysen från möjligheten att de som kallats till insatsen kanske tagit ett jobb eller lämnat arbetskraften innan start för att slippa ta del av de aktivare förmedlingsinsatserna i form av fler möten med arbetsförmedlare som insatsen innebar.⁴²

⁴² Det finns starkt stöd för att förprogramseffekter finns, se till exempel Black m.fl. (2003), Dolton och O'Neill (1996), Rosholm och Svarer (2004), Carling m.fl. (1996) och Hägglund (2006).

Problemet med beräkningen av undanträngningseffekten består i bristande precision på grund av att endast 72 förmedlingskontor ingick i studien. För att kunna få statistiskt säkerställda resultat adderar forskarna data utanför experimentet. Genom att addera data från tre år innan experimentet ökade antalet observationer tio gånger men det innebär också att resultaten inte kan analyseras utan modellantagande. Eftersom variationen i behandling genererats via randomisering innebär detta dock att studien är en mycket välgjord kvasiexperimentell observationsstudie.

6 Statistik, kausalitet och evidens inom olika utbildningar

En förutsättning för att politiken ska kunna vara evidensbaserad är att kunskapen om vad som är att betrakta som evidens finns på olika nivåer i samhället, både bland utförare (forskare) och mottagare (beslutsfattare, media, allmänhet). Som konstaterades ovan definierar vi evidens som ett flertal väl utförda studier med en kausal strategi där resultaten pekar i samma riktning.

I det här avsnittet undersöks i vilken utsträckning ämnet statistik, resonemang om kausala strategier och evidens förekommer på utbildningar vanliga inom det politiska beslutsfattandet. Regeringskansliet lät 2015 Arbetsgivarverket göra en kartläggning av utbildningsnivå och utbildningsinriktning bland de då knappt 5 000 anställda. De inriktningar som var mest representerade var juridik (21 procent), företagsekonomi (19 procent), humaniora och konst (14 procent), statskunskap (14 procent), samhälls- och beteendevetenskap (8 procent) samt nationalekonomi och ekonomisk historia (7 procent). De utbildningar vi tittar närmare på i detta avsnitt är nationalekonomi, statskunskap, sociologi och företagsekonomi. Grundutbildningarna (kandidatexamen) i dessa discipliner innehåller omkring en halv termin obligatoriska kurser i grundläggande statistik. Dessa kurser ges av statistiska institutioner (eller liknande) vid de olika lärosätena där undervisningen bedrivs.

Vi har undersökt kursplanerna för de grundläggande kurserna i statistik vid de fem största universiteten i Sverige.⁴³ Kursplanerna är enhetliga och beskriver samtliga att det övergripande syftet med kurserna är att studenten ska lära sig grundläggande begrepp inom statistik och på ett teoretiskt plan ha kunskaper i sannolikhetslära

⁴³ Vi har undersökt universiteten i Göteborg, Lund, Stockholm, Umeå och Uppsala. Institutionen vid Göteborgs universitet benämns Nationalekonomi med statistik.

och inferens. Studenten ska också behärska några statistiska metoder för datainsamling, bearbetning och analys av kvantitativa och kvalitativa data. Typiska moment i kurserna är datainsamling, urvalsmetoder, punktskattningar och konfidensintervall, hypotesprövning, analys av samband och enkel linjär regression. I endast en av kursplanerna (Kursplan för grundkurs i statistik vid Uppsala universitet) nämns ordet orsak. I denna kursplan skriver man ”Analys av samband mellan variabler. Orsaker och korrelation” och i endast en kursplan (Kursplan för Tillämpad statistik A5 Uppsala universitet) nämns ordet experiment och då endast som en typ av insamling av data. Utifrån denna genomgång är det svårt att tro att studenterna som endast läser de grundläggande kurserna i statistik får en förståelse för kausala strategier och evidens.

Det är dock sannolikt att en hög andel av dem som arbetar inom Regeringskansliet har en master- eller forskarutbildning och därför kan ha bättre förutsättningar att kunna bedöma evidens inom sina respektive sakområden. Masterutbildningarna omfattar två års studier varav en termin (30 högskolepoäng) är uppsatsskrivande. Forskarutbildningarna omfattar fyra års studier varav 1–2 år består av obligatoriska, respektive valbara, kurser. Den inhämtade informationen i kapitlet är i huvudsak avgränsad till hemsidor och kontakter inom respektive disciplin på ett lärosäte, Uppsala universitet.

Utöver ovan nämnda discipliner redovisar vi också vilka kunskaper i statistik som studenter inom medicin och journalistik tillgodogör sig. När det gäller medicin avser framställningen masterutbildningarna i medicinsk forskning och immunologi vid Uppsala universitet, forskarutbildningarna avser samtliga inriktningar inom ramen för Medicin och Farmaci vid Uppsala universitet. Vad gäller journalistik avses magisterprogrammet och forskarutbildningen i journalistik och medievetenskap vid JMG på Göteborgs universitet. Tanken var också att statistikkraven bland jurister skulle redovisas, emellertid framkom att juristutbildningen inte innehåller några obligatoriska moment i ämnet statistik.

Noterbart är att eftersom granskningen endast omfattar ett lärosäte per disciplin bör generella slutsatser dras försiktigt. Även om kursinnehållet på olika lärosäten i huvudsak kan förväntas överlappa kan variationer förekomma.

6.1 Nationalekonomi, statskunskap, sociologi, företagsekonomi och journalistik

Nationalekonomi

Masterutbildningen i nationalekonomi är relativt jämnt fördelad mellan kurser i metod och teori. Första terminen består dels av kurser i mikro- och makroteori, dels av kurser i matematisk analys och spelteori. Den matematiska analysen syftar till att ge studenterna nödvändiga kunskaper i matematik för att tillgodogöra sig avancerade studier i mikroekonomi och mikroekonometri. Den andra terminen ägnas uteslutande åt metodologiska frågor i form av statistik, ekonometrisk teori och tillämpad ekonometri. Under terminen avhandlas de kausala ansatser som beskrevs i avsnitt 3 både teoretiskt och i praktiken. Termin tre syftar till att bredda kunskaperna från de två inledande terminerna, studenterna kan välja mellan tillämpade kurser i makroekonomi, politisk ekonomi och arbetsmarknadsekonomi med mera. Termin fyra ägnas åt uppsatsskrivande där såväl de teoretiska som metodologiska kunskaperna som inhämtats förväntas användas. En masteruppsats i nationalekonomi innehåller således inte sällan policyutvärderingar med utgångspunkt i en kvasiexperimentell ansats.

På forskarutbildningen består hälften av det första årets obligatoriska kurser av dels metod, i form av matematik, statistik och ekonometri, dels teori, i form av mikro- och makroekonomi. Metodkurserna behandlar multivariat analys såväl teoretiskt som praktiskt i form av övningar där studenterna får analysera riktiga data. Det samma gäller experiment och de kvasiexperimentella ansatserna, exempelvis ingår i övningarna att replikera resultat från faktiska studier. Även det andra året av doktorandstudier, som består av valbara kurser, ställer höga krav på kvantitativa metodkunskaper.

Statskunskap

Till skillnad från i nationalekonomi erbjuds studenter på masterutbildningen i statskunskap möjligheten att pröva sina kunskaper på arbetsmarknaden. Via praktik på en myndighet eller i en organisation tillåts studenten att använda sina kunskaper och färdigheter som

utredare samtidigt som de kan få viktig erfarenhet och knyta värdefulla kontakter. Den första terminen består till hälften av kurser i antingen Internationell politik, "Governance" (statlig styrning och organisation) eller "State" (stat) beroende på intresse, och till hälften av metod. Metodkursen är obligatorisk och uppdelad i ett kvalitativt (7,5 poäng) och ett kvantitativt (7,5 poäng) block. Syftet med det kvantitativa blocket är att stärka studenternas förmåga att tillämpa statistiska metoder som linjär och logistisk regressionsanalys samt analysera olika typer av data (tvärsnitts- eller paneldata). Studenterna får också bekanta sig med de olika ansatserna för kausal analys som beskrivs i avsnitt 3, deras respektive styrkor och svagheter samt när det kan vara lämpligt att använda en viss ansats för att besvara en viss forskningsfråga. Studenterna får också möjlighet att själva prova att genomföra vissa analyser. Huvudfokus ligger på att studenterna ska lära sig förstå ansatserna och deras antagande på ett principiellt plan, snarare än den bakomliggande matematiken. Den andra terminen erbjuder, i likhet med den första, möjlighet att välja olika kurser vid institutionen, och den tredje terminen skriver studenten uppsats.

På forskarutbildningen avhandlas multivariat analys liksom experimentella och kvasiexperimentella metoder på ett övergripande plan. Det främsta syftet är således att doktoranderna ska ha kännedom om dessa analysmetoder. Fördjupade och mer konkreta kunskaper kan nås om doktoranden använder sig av någon av dessa metoder i sitt avhandlingsarbete, och då eventuellt också läser valbara kurser i syfte att fördjupa sig. Som statsvetare kan man också välja att ägna sig åt kvalitativa metoder eller politisk teori.

Sociologi

Masterutbildningen i samhällsvetenskap med inriktning på sociologi består till hälften av kurser i sociologi och till hälften av dels uppsatsskrivande, dels så kallade färdighetskurser med inriktning mot metod och hur man lägger upp och genomför en forskningsuppgift. Färdighetskurserna ger kunskap i att tillämpa grundläggande statistiska analysmetoder för samhälls- och beteendevetare, exempelvis linjär och logistisk regression. Man ska också kunna tolka och vara medveten om begränsningar och eventuella felkällor i olika analyser,

samt kunna redovisa resultaten i såväl muntlig som skriftlig form. Utöver detta innehåller grundutbildningen undervisning i bland annat enkätkonstruktion, hypotesprövning, statistiska test, linjär och multipel regression samt logistisk regression.

Den kvantitativa metoddelen på forskarutbildningen består av såväl föreläsningar och seminarier som av praktiska dataövningar. Fokus är linjär och logistisk regressionsanalys men undervisningen omfattar också multilevel regressionsanalys, faktor- och klusteranalys samt tidsserieanalys. Varken experimentell eller kvasiexperimentell design ingår som kursmoment och endast ett fåtal doktorander genomför kvantitativa analyser.

Företagsekonomi

Det finns två olika masterprogram i företagsekonomi i Uppsala, ett med inriktning på företagande och ledning, med delinriktningarna internationellt företagande, marknadsföring och organisation, och ett med inriktningen redovisning och finansiering, med delinriktningarna ekonomistyrning, finansiering och redovisning. Delinriktningarna inom respektive inriktning har överlappande kurser som exempelvis "Business Analysis and Managerial Action" och "The Global Firm" inom inriktningen Företagande och ledning, men också delinriktningsspecifika kurser. Den tredje terminen är vikt för valfria kurser alternativt praktik (internship) eller utbytesstudier, och under den fjärde terminen ska uppsats skrivas. Av de kurser som erbjuds, oavsett inriktning och delinriktning, är en fjärdedels termin "Vetenskapliga metoder i företagsekonomi". Syftet med denna kurs är att studenterna på ett övergripande plan ska kunna förstå och jämföra olika metodologiska ansatser, såväl kvantitativa som kvalitativa, samt själva tillämpa olika metoder för urval, datainsamling och analys. I den kvantitativa delen ingår grundläggande moment inom statistik som korrelationer, t-test och regressionsanalys.

På forskarutbildningen ingår en obligatorisk kurs i kvantitativ metod, därutöver är det upp till doktoranderna själva att välja. Den obligatoriska kursen går igenom linjär och logistisk regression, klusteranalys samt faktor- och variansanalys. Därutöver behandlas datainsamling och aspekter av surveyanalys. Målet är att studenterna

ska få kännedom om och, i viss mån, kunna genomföra denna typ av kvantitativa analyser.

Journalistik, media och kommunikation

På Göteborgs universitet och JMG, Institutet för journalistik, media och kommunikation, erbjuds en av landets mer välrenommerade utbildningar i journalistik. Där finns både en grundutbildning (kandidat) motsvarande tre års studier, samt en ettårig magisterutbildning i undersökande journalistik. Under utbildningens andra år erbjuds kurser som syftar till att studenterna ska nå en fördjupad förståelse för och färdighet i vetenskaplig metod. Kurserna omfattar frågeundersökningar, urval, datainsamling och analys. Särskild vikt läggs vid felkällor och att kritiskt kunna granska opinionsundersökningar. Bland kurserna återfinns också moment som kvantitativ innehållsanalys och fördjupad statistisk analys. Termin 5 fokuseras på visualisering och hantering av databaser. Magisterutbildningen är ett år lång och består av fyra kurser varav en, "Datajournalistik och visualisering", innehåller vissa statistiska moment. Dessa handlar dock främst om att lära sig hantera och kritiskt granska aggregerade data samt kunna presentera dessa på ett effektivt sätt.⁴⁴

På JMG:s forskarutbildning ingår en halv termins obligatorisk metodutbildning uppdelad på hälften kvantitativ (7,5 hp) respektive kvalitativ metod. Det står doktoranderna fritt att välja vilka kurser de ska läsa. Enligt ansvarig studierektor för forskarutbildningen är mönstret att de som använder sig av statistiska metoder i sitt avhandlingsarbete läser avancerade kurser i statistisk metod. Exempelvis erbjuder både Statsvetenskapliga och Psykologiska institutionen vid Göteborgs universitet kurser med moment som handlar om experimentell och kvasiexperimentell metod.

6.2 Medicin

Uppsala universitet erbjuder masterprogram i medicinsk forskning, som ett av tio masterprogram inom vetenskapsområdet Medicin och

⁴⁴ Vi har också varit i kontakt med representanter för journalistutbildningarna på Mittuniversitetet och på Södertörns högskola. Kunskapskraven i statistik på dessa utbildningar förefaller likna dem på JMG.

farmaci. De första två terminerna erbjuder fem kurser som ger grundliga kunskaper om aktuella biomedicinska forskningskoncept, tekniker och analysmetoder. Analysmetoderna består av bioinformatiska verktyg för analyser. Mycket fokus läggs på experimentell design, hur man redan i planeringsfasen ska tänka på utförande och analys av resultat. Målet är att man ska förstå och kunna förklara principer för studiedesign och experimentella modeller, inklusive aspekter på grundläggande och avancerad laborativ forskningsmetodik. Med experimentell design avses här inte kvasiexperimentella designer, även om studenterna får bekanta sig med principerna bakom "difference in difference"-ansatsen. Mer handgripligen får studenterna endast lära känna de kvasiexperimentella ansatserna om de väljer projekt som kräver det.

Det andra året består dels av ett 15 veckor långt individuellt projekt under den tredje terminen, dels av en masteravhandling på cirka 20 veckor under den fjärde terminen. Vid sidan av detta kommer en biostatistikkurs och en vetenskaplig presentationskurs att förbereda studenten för viktiga aspekter av vetenskaplig datahantering och presentation.

Folkhälsovetenskap är ett mångvetenskapligt ämne och kan locka studenter med varierande grundutbildningar, allt från sjuksköterskor och sjukgymnaster till samhällsvetare och jurister. Masterprogrammet består av fyra terminer varav den sista är uppsatsskrivande. Inslaget av metod återfinns den andra terminen då en tredjedel ägnas åt dels epidemiologi,⁴⁵ och dels "Forskningsdesign och metod inom folkhälsoområdet". Den sistnämnda kursen behandlar både kvantitativa och kvalitativa datainsamlings- och analysmetoder. I det kvantitativa momentet ingår multivariat analys såväl teoretiskt som praktiskt. Detsamma gäller experimentell design där studenterna bland annat får i uppgift att upprätta en forskningsplan med experimentell design. Kvasiexperiment, däremot, förekommer inte i undervisningen.

På forskarutbildningen vid vetenskapsområdet för medicin och farmaci, med institutioner inom exempelvis folkhälso- och vårdvetenskap, kirurgiska och medicinska vetenskaper, medicinsk biokemi och mikrobiologi och neurovetenskap, finns (förutom avhandlingsarbetet) ett antal kurser och andra moment som är obligatoriska för alla doktorander oavsett inriktning. Omfattningen av kurser och

⁴⁵ Epidemiologi omfattar läran om hur sjukdomar i allmänhet yttrar sig på befolkningsnivå.

andra moment ska motsvara minst 30 högskolepoäng för doktors-examen. De obligatoriska kurserna omfattar totalt 10,5 av sammanlagt 30 högskolepoäng och består av bland annat en introduktion till vetenskaplig forskning som, om doktoranden själv vill, kan bytas ut mot momenten, i) forskningsetik och vetenskapsteori, ii) vetenskaplig presentation, och iii) introduktionskurs i biostatistik. I kurserna ingår experimentell design som moment och studenterna får öva ”skarpt” genom att utforma designen på egna forskningsproblem. Multivariat analys ingår som moment i kurserna men endast teoretiskt och genom tolkning av ett antal vetenskapliga artiklar. Kvasiexperimentella metoder ingår inte som ett specifikt moment, åtminstone inte som ett obligatorisk sådant.

6.3 Summering

I kapitlet har kunskaperna i ämnet statistik i allmänhet och kausala strategier i synnerhet kartlagts med utgångspunkt i discipliner vanliga inom det politiska beslutsfattandet och dess omgivning. Kunskaper inom dessa områden utvecklar förmågan till kritiskt tänkande kring frågor om kontrafaktiska utfall och orsaks- och verkanssamband, kort sagt kring frågor kopplade till evidens.

Grundutbildningarna (kandidatexamen) inom dessa discipliner innehåller omkring en halv termin obligatoriska kurser i grundläggande statistik. Vid genomgång av kursplaner från de fem största universiteten kan vi konstatera att det är svårt att tro att studenterna som endast läser de grundläggande kurserna kan utveckla ett kritiskt tänkande kring sambandsanalyser med regressionsanalys. I endast en kursplan används ordet ”orsak” men detta nämns helt utan sammanhang.

På de mer avancerade utbildningsnivåerna inkluderar genomgången endast ett lärosäte per disciplin varför slutsatser bör dras med försiktighet. Granskningen av master- och forskarutbildningar visar att nationalekonomi sticker ut som ett ämne där studenterna inte bara ges övergripande utan också djupgående kunskaper i statistik och experimentella metoder. Redan på masternivå får studenterna bekanta sig både teoretiskt och praktiskt med de kvasiexperimentella ansatserna, och på forskarutbildningen ingår ytterligare fördjupning

kring dessa i de obligatoriska kurserna. Även statsvetare får i utbildningen lära känna metoderna, dock på ett ytligare plan. Djupare metodologiska kunskaper är möjliga att nå men är i hög grad beroende av studentens egna val, exempelvis när det gäller valbara kurser och val av uppsats eller avhandlingsämne.

Genomgången av de medicinska vetenskaperna är inte komplett, intrycket är dock att dessa, varken på master- eller forskarnivå, fördjupar sig i kvasiexperimentella ansatser. Med andra ord finns en betydande förbättringspotential för denna typ av experimentell metodik inom dessa områden.

Journalistutbildningarna förefaller innehålla statistikundervisning som i stora drag liknar dem på övriga samhällsvetenskapliga utbildningar, åtminstone på grundnivå. Med andra ord tyder ingenting på att studenterna får en tillräcklig grund för att tänka kritiskt kring frågor av kausal natur. Detta är en brist inte minst i granskningen av politiken där ju olika alternativ för att uppnå givna mål ställs mot varandra och där slutsatser kring konsekvenserna av en viss politik eller vissa förslag dras från olika politiska läger. Vidare följer, av oförmåga att kritiskt kunna granska innehåll, en risk att trovärdigheten i utsagor kommer att bedömas utifrån avsändarens renommé snarare än relevans. Detta illustreras av en pågående diskussion kring " trovärdiga" och " icke-trovärdiga" källor. En risk med en svartvit indelning i dessa kategorier är att de som klassificeras som trovärdiga kan missbruka makten och styra informationen på ett sätt som gynnar dem själva, det vill säga ägarna. Journalister och en allmänhet måste ges möjlighet att kritiskt granska uppgifter oavsett vilken källan är. På JMG:s forskarutbildning ges förvisso utrymme att läsa kurser som handlar om experimentell och kvasi-experimentell metod. Vi har dock ingen information om hur många studenter som utnyttjar denna möjlighet.

7 Hinder för evidensbaserad politik

Föregående kapitel har på olika sätt undersökt aspekter av förut-sättningarna för en politik baserad på evidens och ett gott kunskaps-läge. I detta avsnitt diskuteras aspekter av motsatsen, det vill säga vad vi anser står i vägen för en evidensbaserad politik.

I en festskrift från nya ESO:s 10-årsjubileum diskuterar Lars Calmfors denna fråga med utgångspunkt i den ekonomiska politi-ken, han ställer sig den retoriska frågan: ”Hur får forskningsunderlag genomslag i politiken?”. Den övergripande frågan som diskuteras är den administrativa organisation kring vilken forskningsbaserade resultat ska kunna nå ut till politiken. Calmfors betonar att det finns ett viktigt samspel mellan extern analys (via primärt myndigheter) och intern inom departement. Enligt Calmfors tvingar extern analys fram bättre intern analys för att matcha den externa. Nästa fråga gäller om oberoende analys av ekonomisk politik enbart ska vara positiv, det vill säga bara analysera politikens effekter, eller om den också bör vara normativ och rekommendera, eller avråda från, olika åtgärder för att nå uppställda mål. I denna fråga tar Calmfors inte ställning.

I skriften berör Calmfors vidare många redan kända problem när det gäller ny kunskap att få genomslag i politiken, exempelvis beslutsfattaress svårigheter att ta till sig oberoende, eller externa, utvärderingar och analyser samt forskare som endast ger stöd för redan gjorda ställningstaganden. Båda dessa problem är kända och beskrivs underhållande via anekdotisk bevisföring. Ett högst aktuellt exempel på det förstnämnda är ointresset från politiskt håll att sätta sig in i kunskapsläget inom migrationspolitiken. En rapport från Riksrevisionen (2017c) granskade underlagen till 26 migrations-politiska propositioner mellan 2004 och 2015, och konstaterade att föreskrivna analyser av de förväntade ekonomiska effekterna endast

hade redovisats i undantagsfall.⁴⁶ I majoriteten av fallen saknades en beskrivning av möjliga kommunala konsekvenser helt. I de fall där effekterna på statlig och kommunal nivå redovisades saknades i regel en förklaring till beräkningarna och en beskrivning av vilka antaganden som låg bakom.

Diskussionen i föreliggande kapitel är snävare än den som förs i Calmfors (2017) eftersom den utgår från den definition av evidensbaserad politik som redovisas i kapitel 2, och har sin utgångspunkt i experimentet som metod för att utgöra underlag för politiskt beslutsfattande. Diskussionen rör problem som finns inom akademien men eftersom dessa fått genomslag i samhällsdebatten och inom politiken blir denna diskussion viktig att föra.

I avsnitt 7.1 och 7.2 diskuteras andra hinder för en evidensbaserad politik. I 7.1 diskuteras kritisk teori ("Critical theory"), post-strukturalism och postmodernism, vilka alla är direkt kritiska mot positivistisk vetenskapsfilosofi. I avsnittet diskuteras hur dessa idéer har kommit att påverka både akademi och politik. Avsnitt 7.2 behandlar forskarens roll i kunskapsbildningen, åsiktskonformitet inom akademien och hur enskilda forskningsstudier kan få stort genomslag och prägla politiken och den politiska debatten. I avsnitt 7.3 sammanfattas iakttagelserna.

7.1 Kritik av positivism

Den vetenskapsfilosofi som utgör grunden för randomiserade experiment har sin grund i upplysningens idé om tron på människans förnuft och kan mer konkret hänföras till filosofer som Francis Bacon, John Locke, David Hume och Karl Popper. Positivismen är i sin tur en benämning av olika filosofiska idéer med en strävan att utgå från empiriska fakta vid analys.

Inom matematik kan man bevisa påståenden inom en matematisk modell men i det verkliga livet (det vill säga i empirin) kan ingenting bevisas, däremot kan vi förkasta vissa förklaringar för alternativa och mer troliga förklaringar. Denna tanke benämndes av Karl Popper den hypotetisk-deduktiva metoden, eller kritisk rationalism. Insikten att det inte går att bevisa förklaringar om verkligheten har delvis

⁴⁶ Brister i konsekvensanalyser som underlag för nationell politikutformning har påtalats länge. Se till exempel Riksdagens revisorer (1997), Gunnarsson och Lemne (1998) och Riksrevisionen (2004).

tagits som utgångspunkt för att kritisera själva grunden inom positivismen, det vill säga att sträva efter en objektiv beskrivning av verkligheten. Till exempel är forskarens uppgift enligt Poppers kritiska rationalism att vara kritisk till sin egen teori och att hitta motbevis till denna teori i data.

Enligt den kritiska teorin ("Critical theory") är den vetenskapliga metoden i sig en ideologi som främjar passivitet och reducerar den sociala världen till mönster av orsak och verkan (Agger 1991).⁴⁷ Den vetenskapliga metoden anses inte göra det möjligt att uppnå "nya fakta", till exempel slutet på klassamhället, patriarkatet, rasism och naturens dominans (ibid.). Poststrukturalism och postmodernism utvecklar den kritiska teorin framför allt genom att fokusera på språkets betydelse. Teorierna kan sammanfattas med att språket skapar samhället, inte att språket används för att beskriva en verklighet (ibid.). Det primära i kritiken mot den vetenskapliga metoden från dessa tre teorier är att alla individer har sin egen (subjektiva) verklighet och att det därför inte är meningsfullt att söka efter en "sanning" eller verklighet. Enligt den kritiska teorin skapar kapitalismen ett falskt medvetande om att det befintliga sociala systemet är både oundvikligt och rationellt (ibid.). Innebörden är att individerna behöver medvetandegöras om det förtryck och den maktordning som ligger bakom den rådande ordningen.

Noterbart är att nämnda teorier är kritiska till den vetenskapliga metoden som sådan och inte specifikt till den definition av evidens som specificeras i kapitel 2. Teorierna innehåller relevant kritik av den vetenskapliga metoden: frågeställningar kan styras av ideologi och dessutom är det svårt att som forskare vara objektiv i sin analys, i synnerhet om det i förlängningen finns implikationer för politiken. Det kan exempelvis visa sig i vilka slutsatser som dras och vilka resultat man framhäver som viktiga. Även resultat från randomiserade experiment tolkas i en kontext som kan styras av forskarens åsikter. I kapitel 3 diskuterades problemet med "p-hackande", det vill säga oviljan att som forskare, i enlighet med Poppers filosofi, söka stöd för det som förkastar snarare än bekräftar en teori. Slutsatsen är dock drastisk: i stället för att sträva efter en ökad objektivitet och mångfald i perspektiv eller ideologi, avvisas

⁴⁷ Ben Agger (1952-2015) skrev mer än 20 böcker, en stor mängd artiklar och bokkapitel om bland annat Frankfurtskolan, Marxism, postmodernism, intersektionalism och feminism. Agger var en skarp kritiker av positivism. Han motsatte sig tendensen att inordna sociologin inom den naturvetenskapliga vetenskapsmetodiken (Antonio 2015).

den vetenskapliga metoden eftersom den inte anses användbar i syfte att nå ”nya fakta”. Således är kritisk teori inte en teori så mycket som en ideologi med ett tydligt syfte, nämligen att förändra en rådande maktordning. Eftersom man inte kan använda den vetenskapliga metoden för att analysera den rådande maktordningen finns heller inga målkonflikter vid politiska beslut som fattas för att ändra denna maktordning.

Ur den kritiska teorin har under de senaste decennierna vuxit fram nya och applicerade former av postmoderna teorier med fokus på bland annat ras (kritisk rasteori) och sexualitet (queerteori). Idéer om att samhället är uppbyggt på maktordningar återkommer också inom genusvetenskapen och genusforskningen, då med utgångspunkt i könet. Enligt Jämställ.nu, som är en resurs under Jämställdhetsmyndigheten, bygger teorin om könsmaksordningen eller genussystemet på två principer, könens isärhållande och manlig överordning. *”Isärhållandet innebär att män och kvinnor, och det som är manligt respektive kvinnligt, hålls isär och ses som varandras motsatser. Det medför att kvinnor och män finns på olika arenor i samhället, både horisontellt och vertikalt. Uppdelningen mellan könen på arbetsmarknaden är ett exempel på detta, där kvinnor och män befinner sig i olika sektorer, vård – teknik, samt återfinns på olika positioner, manlig läkare – kvinnligt vårdbiträde. Den manliga överordningen och hierarkin yttrar sig bland annat i att det män gör betraktas som mer värdefullt. Därtill tjänar män mer, har mer makt än kvinnor och betraktas som norm, medan kvinnor ses som undantag och som det avvikande. Dessa könsmonster skapas och upprätthålls både på det personliga planet och på det strukturella planet av såväl kvinnor som män. Jämställdhetsarbetet syftar till att bryta denna ordning och det är först då som vi uppnått ett jämställt samhälle.”*⁴⁸

Könsmaksordningen består i att det finns yrken som är manligt respektive kvinnligt kodade samt att det män gör, bara genom att vara män, värderas högre. Det sistnämnda exemplifieras med att män i genomsnitt har högre inkomster och mer makt. Om man som forskare, i enlighet med Popper, aktivt ska verka för att försöka falsifiera sin teori skulle man kunna invända att det finns en mängd utfall där män som grupp kommer ut sämre än kvinnor, vilket alltså talar emot idén om en könsmaksordning. Exempelvis lyckas pojkar

⁴⁸ Genussystem | Jämställ.nu (jamstall.nu).

tydligt sämre än flickor i skolan och män har kortare livslängd än kvinnor, män utsätts också för mer våld än kvinnor och löper väsentligt större risk att hamna i fängelse.⁴⁹ Kontrasten till evidensbaserad vetenskap där man utifrån empiriska observationer vill pröva de teoretiska förutsägelseerna för att se i vad mån de håller, blir här tydlig. Inom kritisk teori finns inget intresse av att förkasta teorin. Noterbart är också att individers aktiva val av utbildning och yrke ovan beskrivs i passiva termer: män och kvinnor ”finns på olika arenor” och ”befinner sig i olika sektorer”. Eftersom individerna saknar agens består jämställdhetsarbetet i att politiskt bryta den rådande ordningen av isärhållande.

Genusforskningen har i olika sammanhang kritiserats för att vara alltför obenägen att ta intryck av forskningsfält utanför det egna.⁵⁰ Charlotta Stern, forskare i sociologi vid Stockholms universitet, undersökte i Stern (2016) saken närmare genom att granska 20 högt citerade uppsatser publicerade i genustidskrifter mellan 2004 och 2014. Urvalet av artiklar skedde med utgångspunkt i att samtliga handlade om könsskillnader och skillnader i utfall mellan könen och, enligt författaren, borde ha diskuterat biologiska skillnader. Ett kriterium var också att man i uppsatserna hade citerat ”Doing Gender” (West och Zimmerman 1987), en uppsats som ofta beskrivs som banbrytande inom genusforskningen och som är den mest citerade uppsatsen i den högst rankade genustidskriften *Gender & Society* någonsin. I uppsatsen gör författarna en åtskillnad mellan, å ena sidan, det biologiska könet (”sex”) som delar in människor i två kategorier, å andra sidan, det socialt konstruerade könet (”gender”), som med utgångspunkt i det biologiska könet oundvikligen och kontinuerligt återskapas till vardags i mänsklig interaktion.

Stern konstaterar att bara en av de 20 uppsatserna hänvisar till forskning utanför det egna fältet för att förstå observerade könsskillnader. Det begränsade underlaget medger inte slutsatser om all forskning inom området. Författarens urval av högt citerade artiklar kan exempelvis förväntas föra med sig att forskning som inte ligger i mittfåran av ämnet inte inkluderas. Om man gör antagandet att de högst citerade artiklarna är de mest betydelsefulla inom området

⁴⁹ Se exempelvis Medellivslängden i Sverige (scb.se) (uppdaterad 2020-09-02), Flickorna har gått om pojkar i alla ämnen | Läraren (lararen.se), Våld och misshandel – Brottsförebyggande rådet (bra.se), Anstalt - Forskning och statistik | Kriminalvården (kriminalvarden.se).

⁵⁰ Se exempelvis Avskaffa statens stöd till genusforskning Motion 2017/18:387 av Jan Ericson (M) – Riksdagen, eller Fel av staten att stötta genusforskningen | GP.

finns emellertid, på basis av underlaget, fog för att misstänka att det finns en kantring inom ämnet till att söka förklaringar till observerade könsskillnader i utfall i sociala processer snarare än i biologi.⁵¹ Stern påpekar att samlad forskning inom områden som antropologi, utvecklings- respektive evolutionspsykologi, neurovetenskap, genetik med flera visat på skillnader mellan könen i en mängd avseenden, utöver de uppenbart fysiska och biologiska.

Både nuvarande och tidigare regeringar har tagit intryck av teorin om en könsmaktsordning i utformningen av politiken. I skrivelsen ”Makt, mål och myndighet – en feministisk politik för en jämställd framtid (Skr. 2016/17:10) sammanfattas regeringens nuvarande inriktning på jämställdhetspolitiken. Där står bland annat:

Jämställdhetspolitiken utgår från en feministisk analys av samhället där de strukturer och processer som skapar ojämlikhet mellan kvinnor och män står i fokus. Maktbegreppet är centralt i jämställdhetspolitiken och tar sikte på bl.a. kvinnors och mäns inflytande, resurser och villkor, både kvantitativt och kvalitativt. De grundläggande särdragen i ett könsmaktsystem eller genusystem handlar om en uppdelning i två kön samt en isärhållning och segregering av dessa. Isärhållandet innebär att män och kvinnor, och det som ses som manligt respektive kvinnligt, hålls isär och ses som varandras motsatser. Denna segregering tar sig uttryck både horisontellt, bl.a. genom att kvinnor och män finns inom olika yrken och sektorer, och vertikalt, som ett hierarkiskt isärhållande där mannen/det manliga utgör normen och värderas högre. Jämställdhetspolitiken strävar efter att förändra denna strukturella ordning och de processer som resulterar i att kvinnor generellt har en svagare social, ekonomisk och politisk position i samhället än vad män har. De jämställdhetspolitiska målen formuleras därför utifrån grupperna kvinnor respektive män.

I linje med teorin om en könsmaktsordning är regeringens utgångspunkt för jämställdhetspolitiken att män och kvinnor hålls isär av strukturer och processer som leder till att kvinnor har en sämre position i samhället jämfört med män. Dessa strukturer och processer måste således förändras för att jämställdhet ska uppnås.

⁵¹ I detta sammanhang är det intressant att notera att Stern (2016) inte lyckats skapa en debatt inom akademien trots att studien är välkänd inom akademien. Stern (2016) har endast 16 citeringar enligt Google Scholar men har enligt tidskriften hela 34 176 nedladdningar. Noterbart är att ingen av dessa 16 citeringar är kritiska mot studien. Som en jämförelse har Charlotta Sterns högst citerade artikel enligt Google Scholar ”Political Diversity Will Improve Social Psychological Science” 533 citeringar och det med endast 11 879 nedladdningar.

De nu gällande målen för jämställdhetspolitiken fastställdes i huvudsak redan 1994 och har sedan dess bara genomgått smärre justeringar. Det övergripande målet för jämställdheten är att män och kvinnor ska ha samma makt att forma samhället och sitt eget liv. Därutöver finns sex jämställdhetspolitiska delmål.⁵² De handlar om att makt och inflytande ska vara jämnt fördelade mellan könen i samhällets alla sektorer, om att män och kvinnor ska ha samma möjligheter och villkor i fråga om betalt arbete som ger ekonomisk självständighet livet ut, att kvinnor och män samt pojkar och flickor ska ha samma möjligheter och villkor vad gäller utbildning, studieval och personlig utveckling, att det obetalda hem- och omsorgsarbetet ska vara jämnt fördelat, att män och kvinnor ska ha samma förutsättningar för en god hälsa, och att mäns våld mot kvinnor ska upphöra.

I målformuleringarna och i regeringens skrivelse definieras jämställdhet genomgående som att män och kvinnor ska ha samma villkor, förutsättningar och möjligheter. Återkommande är emellertid att det man i praktiken undersöker i stället är likhet i *utfall*, det vill säga en jämn representation av män och kvinnor i olika positioner, yrken och sektorer. Om olikhet i sådana utfall föreligger så drar man slutsatsen att det måste föreligga olikhet i villkor, förutsättningar och möjligheter. Att olikheter i utfall kan bero på andra faktorer, till exempel biologiska skillnader, är avskrivna på förhand.⁵³

Regeringen har sedan 2015 formulerat en rad mål om könsbaserad representation. Bland annat ska andelen kvinnor i börsbolagens styrelser öka och andelen män och kvinnor på politiskt

⁵² Delmål 1: En jämn fördelning av makt och inflytande. Kvinnor och män ska ha samma rätt och möjlighet att vara aktiva samhällsmedborgare och att forma villkoren för beslutsfattandet i samhällets alla sektorer. Delmål 2: Ekonomisk jämställdhet. Kvinnor och män ska ha samma möjligheter och villkor i fråga om betalt arbete som ger ekonomisk självständighet livet ut. Delmål 3: Jämställd utbildning. Kvinnor och män, flickor och pojkar ska ha samma möjligheter och villkor när det gäller utbildning, studieval och personlig utveckling. Delmål 4: Jämn fördelning av det obetalda hem- och omsorgsarbetet. Kvinnor och män ska ta samma ansvar för hemarbetet och ha möjligheter att ge och få omsorg på lika villkor. Delmål 5: Jämställd hälsa. Kvinnor och män, flickor och pojkar ska ha samma förutsättningar för en god hälsa samt erbjudas vård och omsorg på lika villkor. Delmål 6: Mäns våld mot kvinnor ska upphöra. Kvinnor och män, flickor och pojkar, ska ha samma rätt och möjlighet till kroppslig integritet. Källa: Mål för jämställdhet – Regeringen.se.

⁵³ Att likhet i utfall är en uttalad ambition bekräftas i Skr. 2016/17:10: "I regeringens proposition Makt att forma samhället och sitt eget liv (prop. 2005/06:155, bet. 2005/06:AU11, rskr. 2005/06:257) från 2006 slogs det fast att det övergripande målet och delmålen syftade till att uppnå jämställda förutsättningar, men också ett jämställt resultat eller tillstånd. Denna ambition står regeringen fast vid."

tillsatta poster inom statsförvaltningen fördelas jämnt.⁵⁴ På utbildningsområdet är målet att lika många kvinnor som män ska rekryteras som professorer år 2030. Därutöver ska forskningsanslagen fördelas mer lika mellan könen.⁵⁵ För att bryta den könssegregerade arbetsmarknaden har regeringen gett Skolverket i uppdrag att föreslå ändringar i läroplanerna för bland annat förskoleklassen och grundskolan för att förtydliga skolans uppdrag att motverka traditionella könsmönster. Därutöver ska universitet och högskola, genom så kallad jämställdhetsintegrering, arbeta för att bryta könsbundna studieval. Vidare ska löneskillnaderna minska genom bland annat krav på lönekartläggningar, förstärkta sanktionsmöjligheter vid osakliga löneskillnader samt mer resurser till Diskrimineringsombudsmannen.⁵⁶

Målen förtäljer att regeringen betraktar olikheter i arbetsmarknadsrelaterade utfall som tecken på brister i jämställdhet, samt att en jämnare könsrepresentation på ledande positioner och inom olika yrken och utbildningar skulle öka jämställdheten. Ingen forskning eller analys presenteras emellertid som stöd för slutsatsen att nuvarande situation skulle vara resultatet av olikheter i villkor, förutsättningar och möjligheter. Det saknas inte alternativa förklaringsmodeller till de observerade könsskillnaderna, det vill säga förklaringar som inte har med samhälleliga strukturer och normer som missgynnar endera könet att göra. Omfattande psykologiforskning om personliga egenskaper (femfaktorteorin) visar exempelvis på tydliga skillnader mellan män och kvinnor.⁵⁷ Studierna har visat att män i genomsnitt uppvisar en högre grad av självsäkerhet och öppenhet inför idéer än kvinnor, män är också mer riskbenägna. Kvinnor å sin sida uppvisar en högre grad av vänlighet, neuroticism och öppenhet inför känslor. För att få svar på om skillnaderna främst har sitt ursprung i biologi och evolutionära processer eller i samhälleliga normer och förväntningar analyserar Costa m.fl. (2001) svaren från 23 000 män och kvinnor från 26 olika kulturer. De finner, något överraskande, att skillnaderna mellan könen i genomsnitt är

⁵⁴ Regeringens mål om jämn könsrepresentation är en del av FN-kampanjen HeForShe, som UN Women lanserade 2014, <https://www.regeringen.se/artiklar/2015/06/jamstalld-representation/>.

⁵⁵ <https://www.regeringen.se/4b0b1f/contentassets/f26c798733cd41258ec06ff8bd8186d5/handlingsplan-jamstallda-livsinkomster>

⁵⁶ Handlingsplan för jämställda livsinkomster ([regeringen.se](https://www.regeringen.se)).

⁵⁷ Se exempelvis Weisberg m.fl. (2011) för en översikt.

större i de europeiska och amerikanska kulturerna, där de traditionella könsrollerna är svagare än i de asiatiska och afrikanska kulturerna. Liknande resultat har konstaterats i ett flertal efterföljande studier där fler länder inkluderats, se bland annat McCrae m.fl. (2005), Schmitt m.fl. (2008) och Falk och Hermle (2018). Även om resultaten i dessa studier inte medger kausala utsagor om bakomliggande orsaker talar de omvända sambanden emot idén att skillnaderna i egenskaper mellan män och kvinnor enbart skulle vara en konsekvens av sociala strukturer och processer. Motsvarande omvända samband har också konstaterats i en rad andra avseenden, bland annat när det gäller val av yrke och mer specifikt så kallade STEM-yrken, det vill säga arbete inom naturvetenskap (science), teknologi (technology), teknik (engineering) och matematik (mathematics). Kvinnor är generellt sett underrepresenterade inom dessa yrken och exempelvis Breda m.fl. (2020) samt Stoet och Geary (2018) har konstaterat att könsskillnaderna är större i rikare och mer jämställda länder.⁵⁸

Vidare finns frågetecken kring regeringens tolkning av observerade löneskillnader som osakliga.⁵⁹ Medlingsinstitutet (2020), som har i uppdrag att analysera löneutvecklingen ur ett jämställdhetsperspektiv, konstaterar att den oförklarade löneskillnaden, det vill säga den som kvarstår när justering gjorts för att män och kvinnor fördelar sig olika över yrken, sektorer, utbildning och arbetstid, uppgår till 4,2 procent. Man betonar emellertid att denna skillnad inte ska tolkas som ett mått på eventuell diskriminering eftersom alla potentiella faktorer med betydelse för lönen inte låter sig fångas i tillgängliga data. Att exempelvis fånga en individs faktiska erfarenhet kräver att forskaren har mycket god statistik över tjänsteomfattning och arbetstid. I många fall används ålder som en indikator för arbetslivserfarenhet, men denna indikator fungerar emellertid sämre för kvinnor eftersom de oftare gör uppehåll under arbetslivet på grund av familjeskäl. I en rapport konstaterar Riksrevisionen (2019) att lönekartläggningar har begränsade möjligheter att påverka löneskillnaderna mellan män och kvinnor, samt att kraven på genomförande av lönekartläggningar haft liten eller ingen effekt på könsskillnaderna hos arbetsgivare. Sammantaget torde satsningar på att

⁵⁸ Resultaten i Stoet och Geary (2018) har ifrågasatts av Richardson och Reiches (2020) samt av Richardson m.fl. (2020), Stoet och Geary (2020) står emellertid fast vid sina slutsatser.

⁵⁹ I Handlingsplan för jämställda livsinkomster (regeringen.se) slår regeringen fast: "Det förekommer osakliga löneskillnader mellan kvinnor och män."

minska eventuellt osakliga skillnader i löner mellan män och kvinnor ha små förutsättningar att minska könslönegapet.

På motsvarande sätt är stödet för att kvinnor skulle vara missgynnade när det gäller möjligheterna att få forskningsanslag och nå högre positioner inom akademien svagt. Vi känner bara till ett fåtal studier på svenska förhållanden. Wennerås och Wold (1997) analyserar poängsättningen som föregår tilldelningen av forskningsmedel från Medicinska forskningsrådet och konstaterar på basis av jämförelser mellan peer review-omdömen och de ansökandes forskningsmeriter att kvinnliga sökande på ett systematiskt sätt värderas lägre än jämförbara män. När Sandström och Hällsten (2008) ett antal år senare upprepade analysen på nyare ansökningar var resultaten emellertid de motsatta, nu föreföll män vara de som var något förfördelade.⁶⁰ I en experimentell studie av Carlsson m.fl. (2019) ville forskarna undersöka om underrepresentationen av kvinnor i ämnena nationalekonomi, statsvetenskap, juridik, sociologi och fysik på universitet i Norge, Island och Sverige kunde förklaras av diskriminering. Fiktiva CV:s bedömdes till en tänkt tjänst som docent inom respektive ämne. Resultaten visar att kvinnor bedömdes som såväl mer kompetenta som mer anställningsbara än män med jämförbara kvalifikationer. Därutöver återfanns ingen negativ effekt av att ha barn samt inga skillnader i avkastningen på ett starkt CV. Författarna drar därför slutsatsen att diskriminering inte förefaller vara förklaringen till kvinnors underrepresentation på dessa positioner.⁶¹ Madison och Fahlman (2020) undersöker i vad mån kvinnor väljs bort vid tillsättningen av professorstjänster genom att jämföra antalet vetenskapliga publiceringar och citeringar under åren efter utnämningen. Om underrepresentationen av kvinnor är ett resultat av att det krävs mer för att utnämnas borde detta visa sig genom en högre produktivitet de första åren som professor. Studien avser 1 345 tillsättningar vid de största universiteten inom såväl medicin som samhällsvetenskap mellan 2009 och 2014. Resultaten visar dels att kvinnor i genomsnitt hade lägre vetenskapliga meriter vid tillsättningen än män, dels att män i 6 av 12 jämförelser (6

⁶⁰ I Sandström och Hällsten (2008) konstateras att kvinnor i genomsnitt har en fördel motsvarande 10 procentenheter, detta korrigerades senare i Sandström och Hällsten (2020) då fördelen i stället beräknades till drygt 3 procentenheter.

⁶¹ Carlsson m.fl. (2019) sammanfattar det internationella forskningsläget när det gäller rekryteringar inom akademien som blandat. När det kommer till experimentella studier av rekryteringar till högre positioner visar de emellertid att kvinnliga sökanden har en fördel framför män med samma kvalifikationer.

discipliner och 2 jämförelsegrunder) hade signifikant fler publikationer och citeringar. Författarna förkastar därmed teorin om att det ställs högre krav på kvinnor vid professorstillsättningar som förklaring till varför fler män än kvinnor utses till professor.

Utan att göra anspråk på att ha gjort en fullständig genomlysning av jämställdhetspolitiken är vår slutsats att nuvarande utformning av politiken, som tar sin utgångspunkt i ett genusteoretiskt maktordningsperspektiv där strukturer och processer gör att män är överordnade kvinnor, saknar stöd i evidensbaserad forskning. Problemet med teorin om en könsmaktsordning för att förstå könsskillnader är att den med utgångspunkt i vetenskaplig metod inte låter sig prövas och falsifieras till förmån för konkurrerande förklaringsmodeller. Liksom kritisk teori i stort är idén om en könsmaktsordning således inte så mycket en teori som ett axiom eller, om man så vill, en lins genom vilken verkligheten ska förstås. Det här betyder inte att könsroller och normer inte kan vara betydelsefulla för att förklara observerade könsskillnader i olika avseenden, men det är ett alltför snävt perspektiv för att till fullo förstå skillnader i utfall mellan könen. En viktig poäng är därför att en politik som inte beaktar all relevant kunskap för att förstå könsskillnaderna riskerar att motverka sitt syfte. Ett (som i detta fall) alltför ensidigt fokus på jämn representation och oreflekterat tolkande av könsskillnader som resultatet av oönskade strukturer och manlig överordning riskerar att via politiska beslut försämma likheten i villkor, förutsättningar och möjligheter mellan män och kvinnor.

7.2 Forskarens roll

Att vara objektiv är svårt. Den evidensbaserade vetenskapliga metoden minskar dock utrymmet för forskarens egna, personliga värderingar att få genomslag i de resultat som genereras. Problemet är dock att värderingar i stor utsträckning kan styra vilka frågor som ställs, hur resultaten ska tolkas och vilka slutsatser som lyfts fram. Av denna anledning är det viktigt med forskare med en mångfald av värderingar och åsiktspluralism. Avsaknad av sådan kan innebära att perspektiv går förlorade i både undervisning och forskning.

van de Werfhorst (2019) konstaterar att professorer verksamma på universitet i Europa är mer vänsterorienterade än de flesta andra

yrkesgrupper. Emellertid finner han inget stöd för att professorer skulle vara väsentligt mer homogena i sina åsikter än andra. Ett undantag gäller frågor om migration, där en liberal inställning är dominerande och där avvikande åsikter är relativt ovanliga.⁶²

I en nyligen genomförd undersökning av politiska preferenser bland fakultetsanställda vid Harvard University konstaterades att endast 1,5 procent av respondenterna identifierade sig själva som konservativa eller mycket konservativa.⁶³ Att forskare inom olika discipliner tenderar att ha en kantring vad gäller partipolitiska sympatier är knappast överraskande, och har konstaterats även i Sverige. I en studie av Berggren m.fl. (2009) undersöktes partipolitiska sympatier bland forskare och doktorander vid universitet och högskolor i Sverige under december 2005 och januari 2006. Av 4 301 tillfrågade besvarade 1 512 enkäten, vilket ger svarsfrekvens på 35 procent. Med ett betydande bortfall måste resultaten tolkas försiktigt.⁶⁴ Även om det politiska landskapet ser delvis annorlunda ut i dag är det intressant att notera att sociologer och genusvetare då stod starkt till vänster. Bland sociologer föredrog 24 procent Vänsterpartiet och 22 procent Socialdemokraterna medan 33 procent av genusvetarna hade Feministiskt initiativ som största parti. Även statsvetare stod generellt sett till vänster men där var Socialdemokraterna största parti med 26 procent och Folkpartiet (Liberalerna) näst största parti med 19 procent. Bland nationalekonomer återfanns en viss lutning till höger med Folkpartiet som största parti (27 procent) och med Moderaterna som näst största parti (20 procent).

Även om partisympatierna i sig inte behöver påverka den enskilde forskaren finns en överhängande risk att åsiktskonforma miljöer kommer att prägla den forskning som bedrivs, inte minst genom de frågeställningar som studeras men också via de resultat som har störst förutsättningar att ge meritering.

Ett flertal exempel under senare år låter ana att ideologi och rådande åsiktsströmningar har påverkat forskningens innehåll och forskares ställningstaganden. I en debattartikel (DN-debatt 11

⁶² <https://onlinelibrary.wiley.com/doi/epdf/10.1111/1468-4446.12716>

⁶³ Undersökningen omfattade 476 anställda vid Harvard University varav 263, eller 55 procent, besvarade enkäten. <https://www.thecrimson.com/article/2020/3/3/faculty-support-warren-president/>

⁶⁴ Antalet svaranden inom respektive disciplin varierade mellan 39 (genusvetare) och 449 (företagsekonomer).

september 2020) ifrågasätter statsvetarna Bo Rothstein och Peter Esaiasson hur det kommer sig att det är journalistiska snarare än vetenskapliga insatser som belyst problemen med de släktbaserade nätverken, eller klanerna, som begår grova brott och genom detta lägger stora hinder för integration i samhället.⁶⁵ De anser att det är ett tydligt nederlag för den ambition regeringen uttryckt om att mer forskning ska kunna belysa och korrigera allvarliga samhällsproblem.⁶⁶ Författarna menar att en möjlig orsak till frånvaron av forskning kan vara "områdets ideologiska och politiska laddning". Den kan ha inneburit att forskare som velat ifrågasätta bilden av integrationsproblemen som orsakade av majoritetsbefolkningens diskriminering i slutändan valt att inte göra det. Vidare menar de att den kommitté som ger råd om hur forskningsagendan bör utformas, genomföras och uppdateras kan ha styrt forskare bort från detta område.⁶⁷

Även i samhällsdebatten i stort har det kunnat anas en ovilja att analysera och resonera kring kulturens betydelse för invandrades överrepresentation i brottsstatistiken. Återkommande har "socioekonomiska faktorer" lyfts fram som en förklaring, med tillägget att integrationen i samhället måste fungera bättre och nyanlända komma i arbete snabbare. En studie som ibland har refererats till i sammanhanget är Hällsten m.fl. (2013), som undersökt betydelsen av uppväxtförhållanden för att förklara skillnaden i brottsbenägenhet mellan invandrade och svenskfödda. Studien får här fungera som ett exempel på hur resultat från forskningsstudier kan tolkas alltför långtgående och kommuniceras felaktigt.

För att undersöka betydelsen av uppväxtförhållanden skattar man i uppsatsen en regressionsmodell av typen:

$$\text{Brottsbenägenhet} = \text{Utländsk bakgrund (Ja/Nej)} + \text{Socioekonomiska faktorer} + \text{Uppväxtområde},$$

där brottsbenägenhet mäts på olika sätt men bland annat som misstanke om brott. Utländsk bakgrund (Ja) avser antingen att vara född i Sverige med två utlandsfödda föräldrar eller tillhöra en familj som invandrat till Sverige. Socioekonomiska faktorer avser ett flertal

⁶⁵ <https://www.dn.se/debatt/forskningen-har-missat-de-kriminella-natverken/>

⁶⁶ <https://www.aftonbladet.se/debatt/a/MgmmoK/vi-backar-aldrig-fran-svenska-varderingar>

⁶⁷ <https://www.vr.se/uppdrag/framja-och-finansiera-forskning/forskning-om-migration-och-integration/nationellt-forskningsprogram-inom-migration-och-integration.html>

variabler som bland annat antalet syskon, föräldrarnas ålder, sysselsättning, högsta utbildningsnivå, disponibla inkomster samt huruvida de är skilda eller inte. Uppväxtområde fångar geografiskt bostadsområde vid 16 års ålder.⁶⁸

Resultaten i rapporten bekräftar tidigare resultat om att barn med utländsk bakgrund är kraftigt överrepresenterade i brott. Män (kvinnor) födda i utlandet hade i genomsnitt 2,6 (2,1) gånger fler brottsmisstankar jämfört med män (kvinnor) födda i Sverige med svenskfödda föräldrar. Motsvarande överrepresentation för män (kvinnor) födda i Sverige med utlandsfödda föräldrar är 2 (1,5).

När resultaten kontrolleras för resurser i uppväxtmiljö i form av familjesituation (frånskilda föräldrar eller inte, antal syskon), bostadsområde samt föräldrarnas utbildningsnivå och inkomster, försvinner emellertid 66-81 procent av skillnaderna mellan barn födda i Sverige och utlandet för män. För kvinnor är motsvarande minskning 64-79 procent. Skillnaderna mellan barn födda i Sverige med föräldrar från Sverige respektive från utlandet försvinner i ännu högre utsträckning.

På basis av resultaten drar författarna slutsatsen att den största delen av skillnaden i brottslighet mellan svenskfödda och personer födda i utlandet kan *förklaras* av skillnader i socioekonomisk uppväxtmiljö och bostadsområde. Vidare menar man att resultaten är förväntade med utgångspunkt i de teoretiska förutsägelser som kriminologisk forskning gör om brottslighetens orsaker.

Ett problem är emellertid att analysen ovan inte medger kausala utsagor vare sig om hur födelseland eller socioekonomi påverkar brottsbenägenheten. Svår fångade faktorer kopplade till uppväxtmiljö och kultur, associerade med såväl födelseland som socioekonomi, kan ha utelämnats i analysen vilket innebär att resultaten inte kan tolkas i kausala termer. För att undersöka socioekonomiska faktorerers betydelse för brottslighet hade krävts en design som erbjuder en variation i socioekonomisk status som är fränkopplad dessa bakomliggande faktorer. Den analys som nu presenteras ger

⁶⁸ Urvalet avser personer i storstockholmsområdet som lämnade grundskolan (klass 9) mellan 1990 och 1993. Noterbart är att föräldrarnas sysselsättning och inkomster endast avser ett enda år och inte under hela barnets uppväxt. Vad gäller utbildningsnivå för invandrade föräldrar används självrapporterade uppgifter och inte, som för svenskfödda föräldrar, uppgifter från offentliga register. Brottsbenägenheten avser åldern 15 till 28-31 år och mäts på olika sätt, antalet misstankar om brott, antalet fällningar i domstol, antalet fällningar för våldsbrott samt sammanlagd tid i fängelse. Studiepopulationen avser drygt 66 000 individer som gick ut grundskolan i Storstockholm mellan 1990 och 1993.

bara en beskrivning av hur olika i modellen ingående variabler samvarierar med benägenheten att begå brott.

I en separat analys gör forskarna också en ansats att uppskatta kulturens betydelse för brottslighet, detta genom att analysera inomgruppskorrelationer av ungdomar med bakgrund i samma land. Man konstaterar att kultur och brottslighet visserligen korrelerar men att sambandet är svagt och inte tillräckligt betydelsefullt för att kunna förklara det återstående (och oförklarade) gapet i brottsbenägenhet mellan de med och utan utländsk bakgrund när socio-ekonomiska uppväxtförhållanden och bostadsområde beaktats. I stället menar man att resterande del måste förklaras av antingen svår fångade individspecifika faktorer eller med selektion eller diskriminering i rättsväsendet.

Emellertid är inte heller analysen av kulturens betydelse utformad på ett sätt som möjliggör utsagor om orsak- och verkanssamband. Att åstadkomma det är heller inte helt enkelt bör tilläggas. Dels låter sig inte kultur enkelt definieras: man kan tänka sig ett flertal olika sätt att specificera kultur som skulle vara relevanta i sammanhanget, dels inryms i kultur ingen exogen variation, det vill säga en variation som möjliggör att man på ett trovärdigt sätt kan bryta den eventuella associationen som finns mellan, å ena sidan, brottsbenägenhet, å andra sidan, kulturell bakgrund. Sariaslan m.fl. (2013) utnyttjar variationen i uppväxtförhållanden mellan syskon som vuxit upp i olika bostadsområden för att analysera sambandet mellan ungdomsbrottslighet och bostadsområde. Genom att analysera syskon ges möjlighet att studera betydelsen av en aspekt av kultur, nämligen familjen. Man konstaterar att brottsligheten i hög grad samvarierar med bland annat ungdomarnas kön, ursprung (land) och betyg samt föräldrarnas inkomster, bidrag, utbildningsnivå och kriminell belastning. När därutöver betydelsen av icke-definierade familjefaktorer beaktas försvinner sambandet mellan bostadsområde och brottslighet. Tvärtemot Hällsten m.fl. drar författarna därför slutsatsen att "kultur" är betydelsefullt för att förstå brottsbenägenhet.

Hällsten m.fl. (2013) får här illustrera forskarens roll och ansvar i kunskapsbildningen. Trots att författarna förefaller medvetna om den valda metodens begränsningar dras långtgående slutsatser om

orsaks- och verkansamband, vilka i sin tur ger policyimplikationer.⁶⁹ Även om det inte kan uteslutas att studiens resultat hade fått uppmärksamhet och använts i politiska syften även utan dessa långtgående slutsatser, har kommunikationen av resultaten, där även policyförslag förekommer, bidragit till att förstärka snarare än nyansera studiens politiska implikationer.⁷⁰ Vi vill betona att exemplet inte valts för att det nödvändigtvis skulle vara det mest flagranta exemplet på hur resultat kan övertolkas, utan för att det är aktuellt och för att det på ett tydligt sätt illustrerar hur slutsatser från enskilda studier kan få stort genomslag i både politiken och samhällsdebatten.

7.3 Summering

Det finns en mängd faktorer som kan hindra evidens från att få genomslag i politiken. I kapitlet har givits exempel på när bristen på ideologisk mångfald i forskarsamhället leder till att frågor inte ställs och perspektiv inte belyses, eller när forskare drar alltför långtgående slutsatser av sina resultat.

Särskilt vill vi lyfta fram riskerna med diverse maktordnings-teorier med utgångspunkt i kritisk teori ("Critical theory") som under senare år fått genomslag i samhällsdebatt och politik. Gemensamt för dessa är att man förstår samhället utifrån olika maktordningar och att man tolkar gruppvisa utfallsskillnader som resultatet av olika gruppers över- och underordning. Att på förhand låsa sig vid en förklaringsmodell utan att undersöka det empiriska stödet vare sig för den valda modellen eller alternativa förklaringsmodeller är motsatsen till evidensbaserad politik. En evidensbaserad politik kräver att forskare och expertis är villiga att utmana såväl egna övertygelser som inomdisciplinära dogmer. Skrivningar på jämställdhetsområdet antyder att regeringen i viss mån låst sig vid en teori för att förstå observerade könsskillnader. Detta riskerar para-

⁶⁹ På s. 465 i artikeln skriver de: "In addition, as in any analysis based on observable control variables, we suffer from omitted variable bias. Immigrant parents may have specific characteristics that influence crime among their children that we cannot observe."

⁷⁰ Se exempelvis <https://www.aftonbladet.se/debatt/a/11vv2y/uppvaxten-bakom-hogrebrottslighet>, <https://sverigesradio.se/artikel/4417066> och <https://www.dn.se/debatt/okad-invandring-leder-inte-till-okat-antal-brott/>.

doxalt nog att få negativa konsekvenser för jämställdheten, definierat som mäns och kvinnors lika villkor, förutsättningar och möjligheter.

8 Slutdiskussion

Evidens kan sägas vara den bästa beskrivningen av nuvarande kunskapsläge med utgångspunkt i data och historiska händelser. Evidens är inte teorier, ideologier eller resultat från diverse modellsimuleringar. När det gäller frågor om orsak- och verkanssamband utgör sammanställningar av studier med en experimentell design, det vill säga studier som på ett framgångsrikt sätt kan antas ha isolerat betydelsen av behandlingen från andra faktorer som är associerade med det observerade utfallet, den bästa evidensen.

Bland ansatser med ambitionen att belägga kausala samband har det randomiserade, eller klassiska, experimentet länge ansetts vara idealet, i vissa kretsar till och med det enda alternativet. Erfarenheter har emellertid visat att det randomiserade experimentet dras med en rad problem i dess praktiska utförande, problem som gör att epitetet "gold standard" får anses oförtjänt. En genomgång av genomförda fältexperiment inom social- och arbetslöshetsförsäkringen under senare decennier visar att experimenten visserligen genomgående varit av hög kvalitet, men också att de haft brister. En slutsats är att experimenten varit framgångsrika i att mäta de totala effekterna av insatserna, det vill säga dem som även beaktar erbjudandet om att delta i insatserna, men sämre på att ge klarhet i vilka delar av insatserna som haft verkan. Vi menar att om randomiserade experiment ska anses motiverade att genomföra vid utvärdering av policy, ska de förmå skapa förståelse för de bakomliggande mekanismerna om framtida liknande åtgärder, och även i andra kontexter.

I takt med den teknologiska utvecklingen, den ökade tillgången på data och inte minst den metodologiska utvecklingen inom framför allt nationalekonomi, har det naturliga experimentet, eller kvasiexperimentet, vuxit fram som ett många gånger bättre alternativ till det klassiska experimentet. Fördelarna med kvasiexperimentet är

många, inte minst blir kostnaden för utvärderingen lägre och ingreppet i eventuellt reguljära verksamheter betydligt mindre. Därutöver försvinner problem med etiska aspekter och oönskade beteendeffekter till följd av att experimentet ibland inte kan blindas. Nackdelarna är dels att vissa antaganden måste vara uppfyllda för att en kvasiexperimentell ansats ska anses giltig, dels att användning av dessa metoder ställer krav på kunskaper i statistik, kausal inferens och i analys av kvantitativa data. Vår genomgång visar att kraven på kunskaper i statistik och kausala ansatser är grunda inom de flesta samhällsvetenskapliga discipliner, nationalekonomi undantaget. En kartläggning av i vilken utsträckning randomiserade experiment och kvasiexperiment förekommer bland vetenskapliga publikationer inom ekonomi och medicin, ger att studier med en experimentell ansats ökat över tid, vilket torde ha ökat möjligheterna för en väl underbyggd och evidensbaserad politik. Det som talar emot att så är fallet är att dessa studier fortfarande utgör en mycket liten andel av det totala antalet studier, och att citeringsgraden bland dem är låg. Detta försvårar arbetet med att skilja ut bra studier från mindre bra studier och det ställer högre krav på metodkompetens på mottagarsidan. Bristande metodologiska kunskaper, i såväl utförar- som mottagarled, är således ett hinder för evidensbaserad politik.

Ett annat exempel på hinder som ges i rapporten är det växande intresset för maktordningsteorier som ett sätt att förstå samhället och utforma politiken. Gemensamt för förespråkarna för dessa teorier är skepsisen inför den vetenskapliga metoden och alternativa förklaringsmodeller. En politik som av ideologiska skäl väljer att bortse från fakta och kunskap som strider mot en förutbestämd uppfattning om verkligheten riskerar att motverka sitt syfte.

Med anledning av de slutsatser som dragits i rapporten lämnar vi följande rekommendationer i syfte att åstadkomma en mer evidensbaserad politik:

Stärk den metodologiska undervisningen inom samhällsvetenskapliga ämnen

En evidensbaserad politik ställer krav på kritiskt tänkande och kunskaper inom statistik och vetenskaplig metod. Vår genomgång visar att grundutbildningarna (kandidatexamen) inom de samhällsvetenskapliga ämnen som ofta förekommer inom det politiska

beslutsfattandet i regel innehåller en halv termin i ämnet statistik. Det övergripande syftet med kurserna är att få grundläggande kunskaper i bland annat datainsamling, sannolikhetslära, inferens och hypotesprövning. Mycket lite tyder på att studenterna får en förståelse för skillnaden mellan korrelation och kausalitet. Endast i undantagsfall nämns begrepp som orsak, korrelation och experiment i kursplanerna.

Vidare finner vi att det på master- och forskarnivå finns en stor variation i vilken utsträckning studenterna får komma i kontakt med olika kausala ansatser. Exempelvis får statsvetare, om studenten inte själv aktivt väljer att fördjupa sig på området, endast på ett ytligt plan bekanta sig med kvasiexperimentella ansatser. Inom de medicinska utbildningar vi undersökt förekommer kvasiexperimentella ansatser inte alls i kursutbudet, varken på master- eller forskarnivå. Det är egentligen bara i nationalekonomi som krav ställs på att studenterna ska behärska dessa metoder på djupet, vilket sannolikt kan förklaras av nationalekonomers intresse för just orsaks- och verkanssamband.

Utan att göra anspråk på en komplett bild visar vår genomgång att kraven bland studenter på att förstå principerna för hur kausalitet ska förstås och mätas överlag är låg. Vi menar att detta är central kunskap för en evidensbaserad politik och att en översyn av målen med de grundläggande kurserna i statistik är nödvändig. Redan i detta skede bör studenterna introduceras för experiment och tänkande kring det kontrafaktiska samt ges möjlighet att förstå de begränsningar som finns med användning av icke-experimentella data och traditionella regressionsansatser vid kausala utsagor.

Ifrågasätt forskare och forskning

Kartläggningar av politisk hemvist bland forskare har visat att tydliga åsiktskluster kan förekomma inom discipliner, inte sällan till vänster på den politiska höger-vänsterskalan. Forskares ideologiska hemvist kan gömma sig bakom vilka frågeställningar som behandlas, vilka slutsatser som dras och vilka resultat man framhäver som viktiga, i synnerhet om det finns politiska implikationer av det som studeras. Vi menar att ifrågasättande av akademisk auktoritet är nödvändigt för en evidensbaserad politik.

För att undvika risken att inomdisciplinära dogmer får genomslag i politiken och tvärtom säkerställa att beslutsfattare ges en relevant och rättvis bild av bevisläget inom ett visst område, kan en idé vara att använda sig av expertis utanför det aktuella forskningsområdet för att granska forskningsläget. Denna expertis ska bestå av oberoende forskare med god metodkompetens.

Fler analysmyndigheter med tydliga uppdrag behövs

En evidensbaserad politik nås inte genom finansiering av enskilda randomiserade experiment utan genom en långsiktig strategi för kunskapsuppbyggnad inom olika ämnesområden och sektorer, exempelvis genom att bygga upp forskningsmiljöer med fokus på empirisk analys. Ett bra sådant exempel är bildandet av Institutet för arbetsmarknads- och utbildningspolitisk utvärdering (IFAU) i Uppsala 1997. Nästan alla på IFAU är disputerade forskare och institutets verksamhet är av hög kvalitet (Hjälmarsson m fl. 2004).

I en översyn av regeringens analys- och utvärderingsresurser (SOU 2018:79) konstateras att de sju analys- och utvärderingsmyndigheter som finns idag (Brå, ISF, IFAU, Kulturanalys, Tillväxtanalys, Trafikanalys och Vårdanalys) inte tillgodoser regeringens behov av analyser och utvärderingar på olika områden. Man anser att regeringen inte i tillräcklig utsträckning får underlag för att bedöma effekter av olika insatser, kostnadseffektivitet eller hur den samlade måluppfyllelsen ser ut inom ett område. Utredningens kartläggning visar att det saknas oberoende analys- och utvärderingsresurser inom åtta utgiftsområden där det finns särskilt starka skäl för regeringen att ha sådana resurser.

I utredningen konstateras också att analys- och utvärderingsmyndigheter och deras kombination av djupgående sak- och metodkunskaper samt oberoende i förhållande till politikens genomförande är särskilt väl lämpade för att ta fram de underlag som regeringen behöver för att bedöma effekter av beslutade insatser, justera och förbättra utformningen av olika insatser och bedöma behovet av omprövning och effektivitet.⁷¹ Vidare har analys- och

⁷¹ Alternativen vore att utöka resurserna inom Regeringskansliet och kommittéväsendet, att förlägga uppgiften till universitet och högskolor eller att köpa upp analyser och utvärderingar från privata aktörer. Utredningen menar att myndighet är den organisationsform som bäst

utvärderingsmyndigheterna goda förutsättningar att bygga upp kunskap om vad forskning och tidigare utvärderingar visat. Detta är resultat som regeringen behöver få analyserade och bearbetade för att kunna genomföra välunderbyggda reformer.

Vi delar utredningens bedömning att analys- och utvärderingsmyndigheter genom en kombination av oberoende sak- och metodkompetens är särskilt väl lämpade för långsiktig kunskapsuppbyggnad, och att fler myndigheter med tydliga uppdrag att utvärdera effekterna av, och kostnadseffektiviteten i, olika satsningar skulle öka förutsättningarna för en väl underbyggd och evidensbaserad politik. Utmaningen ligger i att bemanna verksamheterna med tillräcklig kompetens inom kausal design och analys. Kompetenta forskare kan arbeta med en mängd olika frågeställningar förutsatt att de får samarbeta med medarbetare med nödvändig sakkunskap. Av denna anledning kan det finnas samordningsvinster med att skapa ett antal större utvärderingsmyndigheter. Utvärderingsmyndigheterna bör placeras i en bredare forskningsmiljö eftersom avgörande för att kunna leverera ett högkvalitativt kunskapsunderlag till politiken är möjligheten att rekrytera personal med rätt kompetens.

tillgodoser regeringens behov av stadigvarande stöd och ger goda förutsättningar för att bygga upp kunskaper över tid. "Myndighetsformen ger också goda möjligheter till ett väl fungerande samspel med regeringen och utvärderaren och tydliga ansvarskedjor, vilket är viktigt både för verksamhetens relevans och dess transparens."

Referenser

- Agger, B. (1991). Critical Theory, Poststructuralism, Postmodernism: Their Sociological Relevance, *Annual Review of Sociology*, 17, 105-131.
- Anderegg, W. R. L., Prall, J. W., Harold, J., och S. H. Schneider (2010). Expert credibility in climate change. *PNAS*, 107, 12107–12109, <https://doi.org/10.1073/pnas.1003187107>.
- Angelov, N., P. Johansson och E. Lindahl (2013). Det envisa könsgapet i inkomster och löner – Hur mycket kan förklaras av skillnader i familjeansvar?, IFAU rapport 2013:2.
- Antonio, R.J (2015). Remembering Ben Agger (1952–2015). *Critical Sociology*, 41, 825-827.
- Banerjee, A. V., Chassangx, S., och Snowberg, E. (2017). Decision theoretic approaches to experiment design and external validity. i *Handbook of Economic Field Experiments*, Kap 4, Vol 1.
- Banerjee, A., L. Pasea, S. Harris, A. Gonzalez-Izquierdo, A. Torralbo, L. Shallcross, m.fl. (2020). Estimating excess 1-year mortality associated with the COVID-19 pandemic according to underlying conditions and age: a population-based cohort study. *Lancet*, 395, 10238, 1715-1725, Maj 30, 2020.
- Barda, N., D. Riesel, A. Akriv, J. Levy, U. Finkel, G. Yona, D. Greenfeld, S. Sheiba, J. Somer, E. Bachmat, G. N. Rothblum, U. Shalit, D. Netzer, R. Balicer och N. Dagan (2020). Developing a COVID-19 mortality risk prediction model when individual-level data are not available. *Nature Communications* 11:4439. <https://doi.org/10.1038/s41467-020-18297-9>.
- Benson, K. och A.J. Hartz (2000). A comparison of observational studies and randomized, controlled trials. *New England Journal of Medicine*, 342, 1878-86.

- Berggren, N., Jordahl, H., och C. Stern (2009). The Political Opinions of Swedish Social Scientists. Finnish Economic Papers, Volume 22, No 2 Autumn 2009.
- Berlinski, D. (2009). The Deniable Darwin and Other Essays. David Klinghoffer (editor), Discovery Institute Press. Seattle, WA ISBN 978-0-979-01412-3.
- Black, D. A., J. A. Smith, M. C. Berger och B. J. Noel (2003). Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System. American Economic Review, 93(4), s. 1313-1327.
- Box, G.E.P och K.B. Wilson (1951). On the experimental attainment of optimum conditions (with discussion). Journal of the Royal Statistical Society (B), 13, 1-15.
- Bray, D och H. von Storch (2010). CliSci2008: A survey of the Perspectives of Climate Scientists Concerning Climate Science and Climate Change. GKSS Working Paper 2010/9.
- Breda, T., Jouini, E., Napp, C. och G. Thebault (2020). Gender stereotypes can explain the gender-equality paradox. PNAS, December 8 2020 117 (49) 31063-31069.
- Brodeur, A., M. Lé, M. Sangnier, och Y. Zylberberg (2016). Star Wars: The Empirics Strike Back. American Economic Journal: Applied Economics, 8(1): 1–32.
- Calmfors L. (2017). Hur får forskningsunderlag genomslag i politiken? Reflektioner utifrån egna Erfarenheter. Rapport till Expertgruppen för studier i offentlig ekonomi 2017:8 Att vara brygga mellan forskning och politik – en festskrift från nya ESO:s 10-årsjubileum, 43-58.
- Camerer, C. F., Dreber, A., Forsell, E., Ho, T.-H., Huber, J., Johannesson, M., Kirchler, M., Almenberg, J., Altmejd, A., Chan, T., Heikensten, E., Holzmeister, F., Imai, T., Isaksson, S., Nave, G., Pfeier, T., Razen, M., and Wu, H. (2016). Evaluating replicability of laboratory experiments in economics. Science 351, 1433-1436.

- Card, D. (1995). Using Geographical Variation in College Proximity to Estimate the Return to Schooling, i Christofides, N.L, E.K. Grant, and R. Swidinsky (red.) *Aspects of Labor Market Behaviour: Essays in Honour of John Vanderkamp*, University of Toronto Press, Toronto.
- Carling, K., P-A Edin, A. Harkman och B. Holmlund (1996). Unemployment Duration, Unemployment Benefits, and Labor Market Programs in Sweden. *Journal of Public Economics*, 59, 313-334.
- Carlsson, M., H. Finseraas, A. H. Midtbøen och G. L. Rafnsdottir (2019). Gender Bias in Academic Recruitment? Evidence from a Survey Experiment in the Nordic Region. *European Sociological Review*, 2020, 1–12 doi: 10.1093/esr/jcaa050 Original Article.
- Cheung, M., Egebark, J., Forslund, A., Laun, L., Rödin, M. och J. Vikström (2019). Effekter av förstärkta förmedlingsinsatser. *Lärdomar från en försöksverksamhet. IFAU rapport 2019:27*.
- Cook, J., Nuccitelli, D., Green, S. A., Richardson, M., Winkler, B., Painting, R., Way, R., Jacobs. P., och A. Skuce (2013). Quantifying the consensus on anthropogenic global warming in the scientific literature. *Environmental research letters*, 8.
- Costa, P. T., Terracciano, A., och McCrae, R. R. (2001). Gender differences in personality traits across cultures: Robust and surprising findings. *Journal of Personality and Social Psychology*, 81, 322–331.
- Crépon B., E. Duflo, M. Gurgand, R. Rathelot, P. Zamora (2013). Do Labor Market Policies have Displacement Effects? Evidence from a Clustered Randomized Experiment. *The Quarterly Journal of Economics*, 128, 2, 531–580. <https://doi.org/10.1093/qje/qjt001>.
- Ding P. och VanderWeele TJ. (2016). Sensitivity Analysis Without Assumptions. *Epidemiology*. 27(3):368-77. doi: 10.1097/EDE.0000000000000457.
- Dolton, P. och D. O'Neill (1996). Unemployment Duration and the Restart Effect: Some Experience Evidence. *The Economic Journal*, 106(435), 387-400.

- Doran, P. T., och M. K. Zimmerman (2011). Examining the Scientific Consensus on Climate Change. EOS, Science News by AGU <https://doi.org/10.1029/2009EO030002>.
- Engström, P., Hesselius, P. och M. Persson (2006). Överutnyttjande i tillfällig föräldrapenning för vård av barn. IFAU rapport 2006:9.
- Engström, P., H. Goine, P. Johansson och E. Palmer (2009). Påverkas sjukskrivning och sjukfrånvaro av information om förstärkt granskning av läkarnas sjukskrivning? IFAU rapport 2009:13.
- Engström, P. och P. Johansson (2012). The medical doctors as gatekeepers in the sickness insurance? Applied Economics, 44(28), 3615–3625.
- Engström, P., P. Hägglund och P. Johansson (2015). Early Interventions and Disability Insurance: Experience from a Field Experiment. Economic Journal, s. 363-392.
- European Commission (2013). Ethics for researchers Facilitating Research Excellence in FP7. European Commission, Directorate-General for Research and Innovation Directorate B – European Research Area Unit B.6 – Ethics and gender.
- Falk, A., och J. Hermle (2018). Relationship of gender differences in preferences to economic development and gender equality. Science 362, Issue 6412, eaas9899. DOI: 10.1126/science.aas9899.
- Fisher, R.A. (1926). The arrangement of field experiments. J. Minist. Agric, 33, 503-513.
- Fisher, R.A. (1935). Design of Experiments. Edinburgh, United Kingdom: Oliver and Boyd. Current 5th ed., Oxford University Press.
- Fogelgren, M., Ornstein, P., Rödin, M. och P. Thoursie (2019). Is Supported Employment Effective for Young Disability Pensioners? Evidence from a Randomized Experiment. Working Paper.
- Fredriksson, P., Oosterbeek, H. och B. Öckert, (2012). Långsiktiga effekter av mindre klasser, IFAU rapport 2012:05.

- Fredriksson, P., Öckert, B., och Oosterbeek, H. (2013). Long-Term Effects of Class Size. *The Quarterly Journal of Economics*, 128 (1), 249-285.
- Goine H., E. Söderberg, P. Engström, och E. Palmer (2009) Effekter av information om förstärkt granskning av medicinska underlag. IFAU rapport 2009:14.
- Gunnarsson, V. och Lemne, M. (1998). Kommittéerna och bofinken – Kan en kommitté se ut hur som helst? Rapport till ESO. Ds 1998:57.
- Haavelmo, T. (1943). The Statistical Implications of a System of Simultaneous Equations *Econometrica* 11, 1—12.
- Hartman, L., P. Hesselius och P. Johansson (2013). "Effects of eligibility screening in the sickness insurance: Evidence from a field experiment". *Labour Economics*, 20, 48– 56.
- Heckman, J., R. LaLonde och J. Smith (1999). "The Economics and Econometrics of Active Labor Market Policies." I *Handbook of Labor Economics*, Vol. III, edited by O. Ashenfelter och David Card.
- Hendry, D. (1980). "Econometrics – Alchemy or Science?" *Economica*, 7(188): 387-406.
- Hesselius, P., Johansson P. och Vikström J. (2008) Påverkas individen av omgivningens sjukfrånvaro? IFAU Rapport 2008:11.
- Hesselius, P., P. Johansson och P. Nilsson (2009). "Sick of your colleagues's absence?" *Journal of European Economic Association*, 7, 1–12.
- Hesselius, P., P. Johansson och J. Vikström (2013). "Monitoring and norms in sickness insurance: empirical evidence from a natural experiment." *The Scandinavian Journal of Economics*, vol. 115(4), s. 995–1019.
- Hjalmarsson, L., A. Sorensen och S. Ström (2004). *Utvärdering av Institutet för arbetsmarknadspolitisk utvärdering* (IFAU) 1997-2002, FAS.
- Holland, P. (1986). "Statistical and Causal Inference" (with discussion)." *Journal of the American Statistical Association*, 81: 945-970.

- Hägglund, P. (2006). "Anvisningseffekter" – finns dom? Resultat från tre arbetsmarknadspolitiska experiment. Institutet för arbetsmarknads- och utbildningspolitisk utvärdering. Rapport 2006:3.
- Hägglund, P. (2009). Effekter av intensifierade förmedlingsinsatser vid Arbetsförmedlingen – erfarenheter från randomiserade experiment. Institutet för arbetsmarknads- och utbildningspolitisk utvärdering. Rapport 2009:15.
- Hägglund, P. (2011). "Are there pre-programme effects of Swedish active labour market policies? Evidence from three randomised experiments." *Economics Letters*, vol. 112(1), s. 91–93.
- Hägglund, P. (2014). Experimental Evidence from Intensified Placement Efforts among Unemployed in Sweden. *Evaluation Review*. Sage Journals. Vol. 38, Issue 3. p. 191-216.
- Hägglund, P., P. Johansson och L. Laun (2020). The Impact of CBT on Sick Leave and Health. *Evaluation Review*, Vol. 44 (2-3), s. 185-217.
- Häkkinen-Skans, I. och P. Johansson (2018). Self-screening effects of monitoring: evidence from a quasi experiment in the Swedish temporary parental benefit program. *Oxford Bulletin of Economics and Statistics*, 2018, 80, 893-904. doi/abs/10.1111/obes.12230.
- Hällsten, M., R. Szulkin och J. Sarnecki (2013). Crime as a price of inequality? The Gap Registered Crime between Childhood Immigrants, Children of Immigrants and Children of Native Swedes. *The British Journal of Criminology*, Vol. 53(3), May 2013, s. 456-481.
- Imbens, Guido, W. (2003). "Sensitivity to Exogeneity Assumptions in Program Evaluation." *American Economic Review*, 93 (2): 126-132.
- Imbens, Guido W. and Rubin, Donald B (2015). *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*, Cambridge University Press.
- Ioannidis, J. (2005). Why Most Published Research Findings Are False. *PLOS Med*, 2(8):e124.

- Ioannidis, J. och Doucouliagos, C. (2013). What's to know about the credibility of empirical economics? *Journal of Economic Surveys*, 27: 997-1004.
- Ioannidis, J. P. A., Stanley, T. D., och Doucouliagos, H. (2017). The Power of Bias in Economics Research. *The Economic Journal*, 127(605):F236-F265.
- ISF (2011). Effekter av tidiga insatser för sjukskrivna. Inspektionen för socialförsäkringen, rapport (2011:17).
- ISF (2013). Slopät frånvarointyg: Kontroll och uttag av tillfällig föräldrapenning. Inspektionen för socialförsäkringen, rapport (2013:19).
- Johansson, P. och M. Palme (2005). Moral hazard and sickness insurance", *Journal of Public Economics* 89, 1879-1890.
- Johansson, P. och E. Lindahl (2012). Locking-in effects due to early interventions? – an evaluation of a multidisciplinary screening programs for avoiding long-term sickness". *Evaluation Review* 2012, 36, 323 – 345.
- Johansson, P. och P. Hägglund (2013). Höjd metodkunskap viktigare än experiment. Debattartikel i *Läkartidning* 2013-07-16.
- Johansson, P. och E. Lindahl (2013). "Can sickness absence be affected by information meetings? Evidence from a social experiment." *Empirical Economics* 2013, vol. 44, 3, s. 1673–1695.
- Kapelner, A., Krieger, A. M., Sklar, M., Shalit, U., och Azriel, D. (2020). Harmonizing optimized designs with classic randomization in experiments. *The American Statistician*, 75:2, 195-206. <https://doi.org/10.1080/00031305.2020.1717619>.
- Laun, L. och P. Skogman Thoursie (2014). "Does privatisation of vocational rehabilitation improve labour market opportunities? Evidence from a field experiment in Sweden". *Journal of Health Economics*, vol. 34, March 2014, s. 59-72.
- Leamer E. (1978) Regression Selection Strategies and Revealed Priors, *Journal of the American Statistical Association*, 73:363, 580-587.
- Leamer, E. (1983). Let's Take the Con Out of Econometrics. *American Economic Review*, 73: 31-43.

- Legates, D. R., Soon, W., Briggs, W. M., och C. Monckton of Brenchley (2015). Climate Consensus and 'Misinformation': A Rejoinder to *Agnotology, Scientific Consensus, and the Teaching and Learning of Climate Change*. *Science & Education*, vol. 24, sid. 299-318.
- Levitt, S.D., och J. A. List (2009). Field experiments in economics: The past, the present, and the future. *European Economic Review*, vol. 53, Issue 1, January 2009, p. 1-18.
- Lomborg, B. (2018). *Prioritizing Development: A Cost Benefit Analysis of the United Nations' Sustainable Development Goals*. Editor Bjorn Lomborg. Cambridge University Press, 2018.
- Lucas, R. (1976). "Econometric Policy Evaluation: A Critique". In Brunner, K.; Meltzer, A. (eds.). *The Phillips Curve and Labor Markets*. Carnegie-Rochester Conference Series on Public Policy. 1. New York: American Elsevier. p. 19-46.
- Madison, G. och P. Fahlman (2020). Sex differences in the number of scientific publications and citations when attaining the rank of professor in Sweden. *Studies in higher education* Published online: 06 Feb 2020
<https://doi.org/10.1080/03075079.2020.1723533>.
- Marschak, J. (1953). "Econometric Measurements for Policy and Prediction". In Wood, W. C.; Koopmans, T. C. (eds.). *Studies in Econometric Methods*. New York: John Wiley & Sons.
- McCrae, R. R., Terracciano, A., Sanchez-Bernados, M. L, D. Djuric Jovic. (2005). Universal Features of Personality Traits From the Observer's Perspective: Data From 50 Cultures. April 2005 *Journal of Personality and Social Psychology* 88(3):547-61.
DOI: 10.1037/0022-3514.88.3.547.
- Medlingsinstitutet (2020). Löneskillnader mellan kvinnor och män 2019 – Vad säger den officiella lönestatistiken?
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, 349 (6251).
- Oreskes, N. (2004). The Scientific Consensus on Climate Change. *Science*, Vol. 306, Issue 5702, s. 1686.
- Pearl Judea (2000). *Cusality*. Cambridge University Press.

- Pearson, K. (1899). Mathematical contributions to the theory of evolution. V. On the reconstruction of the stature of prehistoric races, *Philosophical Transactions of the Royal Society of London Series A* 192, 169-244.
- Permut, T. and Hebel, J. (1989). Simultaneous-equation estimation in a clinical trial of the effect of smoking on birth weight. *Biometrics*, 45(2):619–622.
- Richardson, S. S. och M. W. Reiches (2020). We Dug Into Data to Disprove a Myth About Women in STEM. *Slate Magazine*, mottagen 2020-03-03.
- Richardson, S. S., Reiches, M. W., Bruch, J., Boulicault, M., Noll, N. och H. Shattuck-Heidorn (2020). Is There a Gender-Equality Paradox in Science, Technology, Engineering, and Math (STEM)? Commentary on the Study by Stoet and Geary (2018)”, *Psychological Science* 31 (3): 338-341.
- Riksdagens revisorer (1997). Kommittéväsendet Förslag till riksdagen 1997/98:RR3.
- Riksrevisionen (2004). Förändringar inom kommittéväsendet. RiR 2004:2.
- Riksrevisionen (2017a). Inkomsteffekter av bruksvärdessystemet. Rapport från Riksrevisionen, RiR 2017:24.
- Riksrevisionen (2017b). Överutnyttjande av tandvårdsstödet till följd av överbehandling. Rapport från Riksrevisionen, RiR 2017:19.
- Riksrevisionen (2017c). Konsekvensanalyser inför migrationspolitiska beslut. Rapport från Riksrevisionen, RiR 2017:25.
- Riksrevisionen (2019). Diskrimineringslagens krav på lönekartläggning – ett trubbigt verktyg för att minska löneskillnader mellan könen. Rapport från Riksrevisionen, RiR 2019:16.
- Roelfs, D. J., E. Shor, K. W. Davidson, J. E. Schwartz (2011). Losing life and livelihood: a systematic review and meta-analysis of unemployment and all-cause mortality *Social Science in Medicine* 6,840-54. doi: 10.1016/j.socscimed.2011.01.005.

- Rosenbaum P. R. & D. B. Rubin (1984) Reducing Bias in Observational Studies Using Subclassification on the Propensity Score, *Journal of the American Statistical Association*, 79:387, 516-524.
- Rosenbaum, P. R. (1999). Choice as an alternative to control in observational studies. *Statistical Science*, 14(3):259–304.
- Rosenzweig, M. and K. Wolpin (1980). "Testing the Quantity-Quality Fertility Model: The Use of Twins as a Natural Experiment." *Econometrica*, 48(1): 227-240.
- Roser M., H. Ritchie, E. Ortiz-Ospina och J. Hasell (2020). "Coronavirus Pandemic (COVID-19)". Published online at OurWorldInData.org. Retrieved from: <https://ourworldindata.org/coronavirus> [Online Resource].
- Rosholm, M. och M. Svarer (2004). "Estimating the Threat Effect of Active Labour Market Programmes." Discussion Paper, 1300, Institute for the Study of Labor (IZA).
- Rubin, D.B. (1974). "Estimating Causal Effects of Treatments in Randomized and Non-randomized Studies." *Journal of Educational Psychology*, 66: 688-701.
- Sandström, U. och Hällsten, M. (2008). Persistent nepotism in peer-review". *Scientometrics*, vol 74, nr 2, s. 175-189.
- Sandström, U. och Hällsten, M. (2020). Correction to: Persistent nepotism in peer-review". *Scientometrics*, 17 October 2020.
- Sariaslan, A., N. Långström, B. D'Onofrio, J. Hallqvist, J. Franck och P. Lichenstein (2013). The impact of neighbourhood deprivation on adolescent violent criminality and substance misuse: A longitudinal, quasiexperimental study of the total Swedish population. *International Journal of Epidemiology*, Aug; 42(4): 1057-1066.
- Schmitt D.P., Realo A, Voracek M, J. Allik (2008). Why can't a man be more like a woman? Sex differences in Big Five personality traits across 55 cultures. *J Pers Soc Psychol*, 94(1):168-182. doi: 10.1037/0022-3514.94.1.168.
- Schuklenk, U. (2003). AIDS: Bioethics and public policy, *New Review of Bioethics*, 1:1, 127-144.
- Schulz, K. (1995). Subverting Randomization in Controlled Trials, *JAMA* 274, p. 1456-1458.

- Sims, C.A. (1980). "Macroeconomics and Reality." *Econometrica*, 48(1): 1-48.
- Stern, C. (2016). Undoing Insularity: A Small Study of Gender Sociology's Big Problem *Econ Journal Watch: Sex differences, doing gender*, Steven Pinker, blank slate (econjwatch.org).
- Stock, J. H. and Trebbi, F. (2003). Retrospectives who invented instrumental variable regression? *Journal of Economic Perspectives*, 17(3):177–194.
- Stoet, G. och D.C. Geary (2018). The gender-equality paradox in science, technology, engineering, and mathematics education. *Psychological Science*, 29(4), 581–593.
- Stoet, G. och D.C. Geary (2020). The Gender-Equality paradox Is Part of a Bigger Phenomenon: Reply to Richardson and Colleagues (2020):" *Psychological Science*.
- Strengers, B., Verheggen, B., och K. Vringer (2015). *Climate Science Survey – Questions and Responses*.
- Student (Gosset, W. S.) (1938). Comparison between balanced and random arrangements of field plots. *Biometrika*, 29(3/4):363–378.
- Thistlewaite, D.L. and D.T. Campbell (1960). "Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment." *Journal of Educational Psychology*, 51: 309-317.
- Valentin Amrhein, Sander Greenland, Blake McShane (2019). Scientists rise up against statistical significance. *Nature* 567, 305-307 doi: <https://doi.org/10.1038/d41586-019-00857-9>.
- van de Werfhorst, H. G. (2019). Are universities left-wing bastions? The political orientation of professors, professionals, and managers in Europe. *The British Journal of Sociology*, Vol. 71 (1), s. 47-73.
- Wasserstein, R. L., Schirm, A. L., och Lazar, N. A. (2019). Moving to a World Beyond $p < 0.05$ ". *The American Statistician*, 73(sup1):1-19.
- Weisberg Y.J., Deyoung C.G., J.B. Hirsh. (2011). Gender Differences in Personality across the Ten Aspects of the Big Five. *Front Psychol*, 2, 178. doi: 10.3389/fpsyg.2011.00178. eCollection 2011.

- Wennerås, C. och A. Wold (1997). Nepotism and sexism in peer-review", *Nature*, vol. 387, nr 6631, s. 341-343.
- West, C. och D. H. Zimmerman (1987). *Doing Gender*. *Gender & Society* 1(2): 125–151.
- Wright, P.G. (1928). *The Tariff on Animal and Vegetable Oils*, Macmillan, New York.
- Yates, E. (1937). *Design and Analysis of Factorial Experiments*. Technical Communication 35. Harpenden, United Kingdom: Imperial Bureau of Soil Science.
- Ziliak, S. T. (2014). "Balanced versus Randomized Field Experiments in Economics: Why W. S. Gosset aka "Student Matters", *Review of Behavioral Economics*: Vol. 1: No. 1–2, pp 167-208.

Bilaga A

Tabell A.1 Beskrivning av behandlade inom KBT och matchade icke-behandlade

	Ej sjukskrivna				Sjukskrivna			
	Behandlad		Icke-behandlad		Behandlad		Icke-behandlad	
Man	0,3	(0,5)	0,3	(0,5)	0,3	(0,4)	0,3	(0,4)
Gift	0,3	(0,5)	0,3	(0,5)	0,4	(0,5)	0,5	(0,5)
Ålder	37,5	(11,2)	37,7	(15,4)	43,2	(10,0)	42,5	(15,1)
Utrikes född	0,1	(0,3)	0,1	(0,3)	0,1	(0,3)	0,1	(0,3)
PGI (tkr)	215,0	(163,2)	211,1	(184,3)	270,4	(114,9)	261,4	(122,4)
Högskola	0,5	(0,5)	0,5	(0,5)	0,5	(0,5)	0,5	(0,5)
Arbetslös	0,2	(0,4)	0,2	(0,4)	0,1	(0,4)	0,1	(0,3)
Antal förskrivningar	32,4	(43,2)	32,3	(38,8)	49,1	(68,8)	46,7	(58,6)
Läkemedelskost. (tkr)	10,8	(51,4)	10,3	(13,6)	16,2	(43,4)	17,1	(63,6)
Antal besök i specialistvård	10,5	(15,5)	10,0	(14,8)	16,4	(25,0)	15,8	(22,2)
Antal besök i primärvård	22,1	(21,4)	22,3	(20,6)	30,0	(28,9)	29,0	(27,9)
Antal dagar i slutenvård	0,8	(3,0)	0,7	(2,5)	2,9	(14,0)	3,0	(10,7)
Antal besök med läkare	16,3	(14,1)	16,4	(14,5)	24,4	(21,4)	23,3	(18,5)
Antal vårdbesök, totalt	32,3	(30,2)	32,1	(28,7)	45,9	(43,0)	44,1	(41,0)
Antal vårdbesök, KBT-diagnos	1,2	(1,7)	1,2	(1,6)	1,8	(2,0)	1,8	(1,8)
Antal vårdbesök, MMR-diagnos	0,4	(1,5)	0,4	(1,5)	0,9	(3,2)	0,7	(2,6)
Sjukskrivning, 1 kvartal bakåt	1,9	(9,3)	1,8	(8,9)	45,8	(30,8)	44,2	(32,0)
Sjukskrivning, 2 kvartal bakåt	2,0	(10,7)	1,7	(9,8)	19,0	(31,9)	19,3	(32,6)
Sjukskrivning, 3 kvartal bakåt	2,1	(11,2)	1,9	(10,5)	12,3	(27,8)	13,8	(28,0)
Sjukskrivning, 4 kvartal bakåt	2,1	(11,6)	1,9	(10,9)	10,5	(26,0)	10,3	(25,1)
Sjukskrivning, senaste 3 åren	30,8	(104,3)	26,2	(94,5)	156,5	(214,8)	152,3	(213,8)
Antal obs.	5 124		5 124		1 148		1 148	

Anm: Fetstil markerar statistiskt säkerställd skillnad mellan behandlade och icke-behandlade på 5-procentsivån.

Tabell A.2 Publikationer, randomiserade experiment inom socialförsäkringar 1989-2019

Titel	Författare	Tidskrift	År	Cit.
Do Bonus Offers Shorten Unemployment-Insurance Spells – Results From the Washington Experiment	Oleary, CJ; Spiegelman, RG; Kline, KJ	JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	1995	2
Testing a financial incentive to promote re-employment among displaced workers: The Canadian Earnings Supplement Project (ESP)	Bloom, HS; Schwartz, S; Lui-Gurr, S; Lee, SW; Peng, J; Bancroft, W	JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	2001	3
Housing mobility programs and economic self-sufficiency: Evidence from a randomized experiment	Ludwig, J; Duncan, GJ; Pinkston, JC	JOURNAL OF PUBLIC ECONOMICS	2004	25
Activating unemployed workers works; Experimental evidence from Denmark	Graversen, Brian Krogh; van Ours, Jan C.	ECONOMICS LETTERS	2008	12
The effect of monitoring unemployment insurance recipients on unemployment duration: Evidence from a field experiment	Micklewright, John; Nagy, Gyula	LABOUR ECONOMICS	2010	11
The medical doctors as gatekeepers in the sickness insurance?	Engstrom, Per; Johansson, Per	APPLIED ECONOMICS	2012	4
Social Behaviour in Work Absence	Hesselius, Patrik; Johansson, Per; Vikstrom, Johan	SCANDINAVIAN JOURNAL OF ECONOMICS	2013	12
Effects of eligibility screening in the sickness insurance: Evidence from a field experiment	Hartman, Laura; Hesselius, Patrik; Johansson, Per	LABOUR ECONOMICS	2013	8
Can sickness absence be affected by information meetings? Evidence from a social experiment	Johansson, Per; Lindahl, Erica	EMPIRICAL ECONOMICS	2013	6
The impacts of vocational rehabilitation	Markussen, Simen; Roed, Knut	LABOUR ECONOMICS	2014	11
Does privatisation of vocational rehabilitation improve labour market opportunities? Evidence from a field experiment in Sweden	Laun, Lisa; Thoursie, Peter Skogman	JOURNAL OF HEALTH ECONOMICS	2014	8
Heterogeneous impacts on earnings from an early effort in labor market programs	Sorensen, Kenneth Lykke	LABOUR ECONOMICS	2016	2
Optimal Unemployment Insurance with Endogenous Negative Duration Dependence	Lehr, Brandon	PUBLIC FINANCE REVIEW	2017	1

Titel	Författare	Tidskrift	År	Cit.
Bridging the Gap: Evidence from Externally Hired CEOs	Ertimur, Yonca; Rawson, Caleb; Rogers, Jonathan L.; Zechman, Sarah L. C.	JOURNAL OF ACCOUNTING RESEARCH	2018	3
Imperfect Monitoring of Job Search: Structural Estimation and Policy Design	Cockx, Bart; Dejemeppe, Muriel; Launov, Andrey; Van der Linden, Bruno	JOURNAL OF LABOR ECONOMICS	2018	2
The effects of computers and acquired skills on earnings, employment and college enrollment: Evidence from a field experiment and California UI earnings records	Fairlie, Robert W.; Bahr, Peter Riley	ECONOMICS OF EDUCATION REVIEW	2018	1
Self-screening Effects of Monitoring: Evidence from a Quasi Experiment in the Swedish Temporary Parental Benefit Program	Skans, Iida Hakkinen; Johansson, Per	OXFORD BULLETIN OF ECONOMICS AND STATISTICS	2018	0
Raising Take-Up of Social Assistance Benefits through a Simple Mailing: Evidence from a French Field Experiment	Chareyron, Sylvain; Gray, David; L'Horty, Yannick	REVUE D ECONOMIE POLITIQUE	2018	0
Person-Centred Care in Patients with Acute Coronary Syndrome: Cost-Effectiveness Analysis Alongside a Randomised Controlled Trial	Pirhonen, Laura; Bolin, Kristian; Olofsson, Elisabeth Hansson; Fors, Andreas; Ekman, Inger; Swedberg, Karl; Gyllenstein, Hanna	PHARMACOECONOMICS- OPEN	2019	0
Liquidity Affects Job Choice: Evidence from Teach For America	Coffman, Lucas C.; Conlon, John J.; Featherstone, Clayton R.; Kessler, Judd B.	QUARTERLY JOURNAL OF ECONOMICS	2019	0

Tabell.A.3 Publikationer, naturliga och kvasiexperiment inom området socialförsäkringar 1989-2019

Titel	Författare	Tidskrift	År	Citeringar
Unnatural experiments? Estimating the incidence of endogenous policies	Besley, T; Case, A	ECONOMIC JOURNAL	2000	190
The effects of the unemployment insurance payroll tax on wages, employment, claims and denials	Anderson, PM; Meyer, BD	JOURNAL OF PUBLIC ECONOMICS	2000	67

Titel	Författare	Tidskrift	År	Citeringar
Supply side hysteresis: the case of the Canadian unemployment insurance system	Lemieux, T; MacLeod, WB	JOURNAL OF PUBLIC ECONOMICS	2000	17
Spell durations with long unemployment insurance periods	Bratberg, E; Vaage, K	LABOUR ECONOMICS	2000	13
The impact of taxing unemployment insurance benefits on unemployment duration and post-unemployment earnings	Lee, CI	INTERNATIONAL TAX AND PUBLIC FINANCE	2000	1
Benefit duration and unemployment entry: A quasi-experiment in Austria	Winter-Ebmer, R	EUROPEAN ECONOMIC REVIEW	2003	29
Benefit entitlement and unemployment duration - The role of policy endogeneity	Lalive, R; Zweimuller, J	JOURNAL OF PUBLIC ECONOMICS	2004	41
Benefit entitlement and the labor market: Evidence from a large-scale policy change	Lalive, R; Zweimuller, J	LABOR MARKET INSTITUTIONS AND PUBLIC REGULATION	2004	9
Not so voluntary retirement decisions? Evidence from a pension reform	Hakola, T; Uusitalo, R	JOURNAL OF PUBLIC ECONOMICS	2005	28
Does early intervention help the unemployed youth?	Carling, K; Larsson, L	LABOUR ECONOMICS	2005	12
How shortening the potential duration of unemployment benefits affects the duration of unemployment: Evidence from a natural experiment	van Ours, JC; Vodopivec, M	JOURNAL OF LABOR ECONOMICS	2006	56
Fertility and the household's economic status: A natural experiment using Indian micro data	Gupta, ND; Dubey, A	JOURNAL OF DEVELOPMENT STUDIES	2006	3
Does reducing unemployment insurance generosity reduce job match quality?	van Ours, Jan C.; Vodopivec, Milan	JOURNAL OF PUBLIC ECONOMICS	2008	46
The employment effect of stricter rules for eligibility for DI: Evidence from a natural experiment in Sweden	Karistrom, Anders; Palme, Marten; Svensson, Ingemar	JOURNAL OF PUBLIC ECONOMICS	2008	31
Interest rate restrictions in a natural experiment: Loan allocation and the change in the usury laws in 1714	Temin, Peter; Voth, Hans-Joachim	ECONOMIC JOURNAL	2008	15
How changes in unemployment benefit duration affect the inflow into unemployment	Tuit, Sander; van Ours, Jan C.	ECONOMICS LETTERS	2010	11

Titel	Författare	Tidskrift	År	Citeringar
Do time limits in the sickness insurance system increase return to work?	Hagglund, Pathric	EMPIRICAL ECONOMICS	2013	11
Do Alternative Base Periods Increase Unemployment Insurance Receipt Among Low-Educated Unemployed Workers?	Gould-Werth, Alix; Shaefer, H. Luke	JOURNAL OF POLICY ANALYSIS AND MANAGEMENT	2013	6
Is there a threat effect of mandatory activation programmes for the long-term unemployed?	Graversen, Brian Krogh; Larsen, Brian	EMPIRICAL ECONOMICS	2013	4
Labor Market Effects of Unemployment Insurance Design	Tatsiramos, Konstantinos; van Ours, Jan C.	JOURNAL OF ECONOMIC SURVEYS	2014	46
Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore	Agarwal, Sumit; Qian, Wenlan	AMERICAN ECONOMIC REVIEW	2014	41
The Sick Pay Trap	Fevang, Elisabeth; Markussen, Simen; Roed, Knut	JOURNAL OF LABOR ECONOMICS	2014	22
Unemployment insurance and management of the older workforce in a dual labor market: Evidence from France	Baguelin, Olivier; Remillon, Delphine	LABOUR ECONOMICS	2014	5
Market Externalities of Large Unemployment Insurance Extension Programs	Lalive, Rafael; Landais, Camille; Zweimuller, Josef	AMERICAN ECONOMIC REVIEW	2015	30
Economic Insecurity and the Weight Gain of Canadian Adults: A Natural Experiment Approach	Watson, Barry; Osberg, Lars; Phipps, Shelley	CANADIAN PUBLIC POLICY-ANALYSE DE POLITIQUES	2016	2
Corporate Social Responsibility as an Employee Governance Tool: Evicence From a Quasi-Experiment	Flammer, Caroline; Luo, Jiao	STRATEGIC MANAGEMENT JOURNAL	2017	53
A methodological proposal to evaluate the cost of duration moral hazard in workplace accident insurance	Martin-Roman, Angel; Moral, Alfonso	EUROPEAN JOURNAL OF HEALTH ECONOMICS	2017	2
Temporary Disability and Economic Incentives	Fevang, Elisabeth; Hardoy, Ines; Roed, Knut	ECONOMIC JOURNAL	2017	1

Titel	Författare	Tidskrift	År	Citeringar
Can a search model predict the effects of an increase in the benefit duration? Evidence from the Portuguese unemployment insurance reform	Novo, Alvaro A.; Silva, Andre C.	IZA JOURNAL OF LABOR POLICY	2017	0
The Effects of the Complementary Compensation on Sickness Absence: an Approach Based on Collective Bargaining Agreements in France	Ben Halima, Mohamed Ali; Koubi, Malik; Regaert, Camille	LABOUR- ENGLAND	2018	2
Labor Misallocation and Mass Mobilization: Russian Agriculture during the Great War	Dower, Paul Castaneda; Markevich, Andrei	REVIEW OF ECONOMICS AND STATISTICS	2018	2
The Labour Market Effects of Increases in Social Insurance Premium: Evidence from Japan	Kodama, Naomi; Yokoyama, Izumi	OXFORD BULLETIN OF ECONOMICS AND STATISTICS	2018	1
Accelerating the transition to employment at benefit exhaustion: still possible after four years of unemployment?	Geerdsen, Lars Pico, Lyk-Jensen, Stephanie Vincent; Weatherall, Cecilie Dohlmann	EMPIRICAL ECONOMICS	2018	1
Self-screening Effects of Monitoring: Evidence from a Quasi Experiment in the Swedish Temporary Parental Benefit Program	Skans, Iida Hakkinen; Johansson, Per	OXFORD BULLETIN OF ECONOMICS AND STATISTICS	2018	0
Unemployment insurance and the labor market	Zweimueller, Josef	LABOUR ECONOMICS	2018	0
The Impact of Corporate Welfare Policy on Firm-Level Productivity: Evidence from Unemployment Insurance	Darrough, Masako; Kim, Heedong; Zur, Emanuel	JOURNAL OF BUSINESS ETHICS	2019	0
Forward-looking moral hazard in social insurance	Eliason, Marcus; Johansson, Per; Nilsson, Martin	LABOUR ECONOMICS	2019	0
Consumption and Debt Response to Unanticipated Income Shocks: Evidence from a Natural Experiment in Singapore	Agarwal, Sumit; Qian, Wenlan	AMERICAN ECONOMIC REVIEW	2014	41

Förteckning över tidigare rapporter till ESO

2021

- I en tid av pandemi – en ESO-antologi med samhällsvetenskapliga reflektioner.
- En beklaglig förlust? En ESO-rapport om erfarenheter och lärdomar av arvsskatt.
- Ingen reklam tack – en ESO-rapport om myndigheternas kommunikation.

2020

- Med framtiden för sig – en ESO-rapport om sociala investeringar.
- Jämställdhet räknas – en ESO-rapport om kvinnors förändrade position i arbetslivet.
- Morot utan piska – en ESO-rapport om stärkta incitament för kommunal effektivitet.
- Spänning på hög nivå – en ESO-rapport om elnätets roll för säkra elleveranser.
- Uppkopplad utbildning – en ESO-rapport om högskolans digitalisering.
- Avgörande mål – en ESO-rapport om sysselsättningspolitiska målformuleringar.
- Vårt framtida skattesystem – en ESO-rapport med förslag på en genomgripande skattereform.

2019

- Lika för alla? En ESO-antologi om skolans likvärdighet.

- Synd och skatt – en ESO-rapport om politiken inom områdena alkohol, tobak och spel.
- Skillnad på marginalen – en ESO-rapport om reformerad inkomstbeskattning.
- Rätt på EU:s sätt – en ESO-rapport om EU-rättens inverkan på svensk skattelagstiftning.
- Klimatmål på villovägar? En ESO-rapport om politiken för utsläppsminskningar i vägtrafiken.
- Stöta på patrull – en ESO-rapport om polisens problemorienterade arbete.
- Pang för pengarna – en ESO-rapport om Sveriges militära materielförsörjning.
- Vem vårdar bäst? En ESO-rapport om svensk sjukhusvård i ett jämförande perspektiv.

2018

- Grundlag i gungning? En ESO-rapport om EU och den svenska offentlighetsprincipen.
- Lönar sig arbete 2.0? En ESO-rapport med fokus på nyanlända.
- Tid för integration – en ESO-rapport om flyktingars bakgrund och arbetsmarknadsetablering.
- Skydda lagom – en ESO-rapport om miljömålet Levande skogar.
- Tänk efter före! En ESO-rapport om samhällsekonomiska konsekvensanalyser.
- Operation digitalisering – en ESO rapport om hälso- och sjukvården.
- Data i egna händer – en ESO-rapport om personliga hälsokonton.

2017

- Dags för omprövning – en ESO-rapport om styrning av offentlig verksamhet.
- Bygg mer för fler! En ESO-rapport om staten, kommunerna och bostadsbyggandet.
- Ankomst och härkomst – en ESO-rapport om skolresultat och bakgrund.
- Yes box! En ESO-rapport om en ny modell för kapital- och bostadsbeskattning.

- Olika kön, olika lön – en ESO-rapport om diskriminering på arbetsmarknaden.
- Makar som delar på kakan – en ESO-rapport om jämställda pensioner.
- Inspiration för integration – en ESO-rapport om arbetsmarknadspolitik för nyanlända i fem länder.
- Att vara brygga mellan forskning och politik – en festskrift från nya ESO:s 10-årsjubileum.

2016

- Boende med konsekvens – en ESO-rapport om etnisk bostadssegregation och arbetsmarknad.
- Sjukskrivningarnas anatomi – en ESO-rapport om drivkrafterna i sjukförsäkringssystemet.
- När skolan själv får välja – en ESO-rapport om friskolornas etableringsmönster.
- Digitaliseringens dynamik – en ESO-rapport om strukturomvandlingen i svenskt näringsliv.
- Grön tillväxt under lupp – en ESO-rapport om ett begrepp i tiden.
- Mer än tur i struktur – en ESO-rapport om kommunal effektivitet.
- När det rätta blir det lätta – en ESO-rapport om ”nudging”.

2015

- En ny giv? En ESO-rapport om regleringen av spelmarknaden.
- Maktutövningar under lagarna? En ESO-rapport om trotsiga kommuner.
- En förlorad generation? En ESO-rapport om ungas etablering på arbetsmarknaden.
- Verksamma insatser mot brott? En ESO-rapport om orsak och verkan.
- Familjepolitik för alla? En ESO-rapport om föräldrapenning och jämställdhet.

2014

- Goda år på ålders höst? En ESO-rapport om konkurrens i äldreomsorgen.
- 3:12-Corporations in Sweden: The Effects of the 2006 Tax Reform on Investments, Job Creation and Business Start-ups.
- Företagandets förutsättningar – En ESO-rapport om den svenska ägarbeskattningen.
- Kapital på krita? En ESO-rapport om företagandets finansiering.
- Konkurrens, kontrakt och kvalitet – hälso- och sjukvård i privat regi.
- Hållbara beräkningar – en ESO-rapport om att bedöma den offentliga sektorns finansiella hållbarhet.
- Med nya mått mätt – en ESO-rapport om indikationer på produktivitetsutvecklingen i offentlig sektor.
- Institutionsvård, incitament och information – en ESO-rapport om placering av ungdomar med sociala problem.

2013

- Bäste herren på täppan? En ESO-rapport om bostadsbyggande och kommunala markanvisningar.
- Allmän nytta eller egen vinning? En ESO-rapport om korruption på svenska.
- Var skapas jobben? En ESO-rapport om dynamiken i svenskt näringsliv 1990–2009.
- Transportinfrastrukturens framtida organisering och finansiering.
- Investeringar in blanco? En ESO-rapport om behovet av infrastruktur.
- Bonde söker bidrag – en ESO-rapport om effektivitet i det svenska landsbygdsprogrammet.
- The pension system in Sweden.
- Den offentliga sektorn – en antologi om att mäta produktivitet och prestationer.
- Utvinning för allmän vinning – en ESO-rapport om svenska mineralinkomster.
- Offentlig upphandling eller gröna nedköp? En ESO-rapport om miljöpolitiska ambitioner.

2012

- Svängdörr i staten – en ESO-rapport om när politiker och tjänstemän byter sida.
- En god start – en ESO-rapport om tidigt stöd i skolan.
- Den akademiska frågan – en ESO-rapport om frihet i den högre skolan.
- Income Shifting in Sweden. An empirical evaluation of the 3:12 rules.
- Samhällsekonomin på spåret – en ESO-rapport om att räkna på tunnelbanan.
- Hjälpa eller stjälpa? En ESO-rapport om kontrollfunktionen i arbetslöshetsförsäkringen.
- Lärda för livet? – en ESO-rapport om effektivitet i svensk högskoleutbildning.
- Forskning och innovation – statens styrning av högskolans samverkan och nyttiggörande.

2011

- UD i en ny sits – organisation, ledning och styrning i en globaliserad värld.
- Försvarets förutsättningar – en ESO-rapport om erfarenheter från 20 år av försvarsreformer.
- Kalorier kostar – en ESO-rapport om vikten av vikt.
- Avtalsbestämda ersättningar, andra kompletterande ersättningar och arbetsutbudet.
- Sysselsättning för invandrare – en ESO-rapport om arbetsmarknadsintegration.
- Kollektivtrafik utan styrning.
- Vägval i vården – en ESO-rapport om skillnader och likheter i Norden.
- Att lära av de bästa – en ESO-rapport om svensk skola i ett internationellt forskningsperspektiv.
- Rapport från ett ESO-seminarium – decenniets framtidsfrågor.

2010

- En kår på rätt kurs? En ESO-rapport om försvarets framtida kompetensförsörjning.

- Beskattning av privat pensionssparande.
- Polisens prestationer – En ESO-rapport om resultatstyrning och effektivitet.
- Swedish Tax Policy: Recent Trends and Future Challenges.
- Statliga bidrag till kommunerna – i princip och praktik.
- Revisionen reviderad – en rapport om en kommunal angelägenhet.
- Värden i vården – en ESO-rapport om målbaserad ersättning i hälso- och sjukvården.
- Enkelt och effektivt – en ESO-rapport om grundtrygghet i välfärdssystemen.
- Kåren och köerna. En ESO-rapport om den medicinska professionens roll i styrningen av svensk hälso- och sjukvård.

2009

- Den långsiktiga finansieringen – välfärdspolitikens klimatfråga?
- Regelverk och praxis i offentlig upphandling.
- Invandringen och de offentliga finanserna.
- Fyra dyra fonder? Om effektiv förvaltning och styrning av AP-fonderna.
- Lika skola med olika resurser? En ESO-rapport om likvärdighet och resursfördelning.
- En kår i kläm – Läraryrket mellan professionella ideal och statliga reform ideologier.

2003

- Skolmisslyckande – hur gick det sen?
- Politik på prov – en ESO-rapport om experimentell ekonomi.
- Precooking in the European Union – the World of Expert Groups.
- Förtjänst och skicklighet – om utnämningar och ansvarsutkrävande av generaldirektörer.
- Bostadsbyggandets hinderbana – en ESO-rapport om utvecklingen 1995–2001.
- Axel Oxenstierna – Furstespegel för 2000-talet.

2002

- ”Huru skall statsverket granskas?” – Riksdagen som arena för genomlysning och kontroll.
- What Price Enlargement? Implications of an expanded EU.
- Den svenska sjukan – sjukfrånvaron i åtta länder.
- Att bekämpa mul- och klövsjuka en ESO-rapport om ett brännbart ämne.
- Lärobok för regelnissar – en ESO-rapport om regelhantering vid avregleringar.
- Att hålla balansen – en ESO-rapport om kommuner och budgetdisciplin.
- The School’s Need for Resources – A Report on the Importance of Small Classes.
- Klassfrågan – en ESO-rapport om lärartätheten i skolan.
- Staten fick Svarte Petter – en ESO-rapport om bostadsfinansieringen 1985–1993.
- Hoten mot kommunerna – en ESO-rapport om ansvarsfördelning och finansiering i framtiden.

2001

- Mycket väsen för lite ull – en ESO-rapport om partnerskapen i de regionala tillväxtavtalen.
- I rikets tjänst – en ESO-rapport om statliga kårer.
- Rättvisa och effektivitet – en ESO-rapport om idéanalys.
- Nya bud – en ESO-rapport om auktioner och upphandling.
- Betyg på skolan – en ESO-rapport om gymnasieskolorna.
- Konkurrens bildar skola – en ESO-rapport om friskolornas betydelse för de kommunala skolorna.
- Priset för ett större EU – en ESO-rapport om EU:s utvidgning.

2000

- Att granska sig själv – en ESO-rapport om den kommunala miljötillsynen.
- Bra träffbild, fast utanför tavlan – en ESO-rapport om EU:s strukturpolitik.
- Utbildningens omvägar – en ESO-rapport om kvalitet och effektivitet i svensk utbildning.

- En svartvit arbetsmarknad? – en ESO-rapport om vägen från skola till arbete.
- Privilegium eller rättighet? – en ESO-rapport om antagningen till högskolan
- Med många mått mätt – en ESO-rapport om internationell benchmarking av Sverige.
- Kroppen eller knoppen? – en ESO-rapport om idrotts-gymnasierna.
- Studiebidraget i det långa loppet.
- 40-talisternas uttåg – en ESO-rapport om 2000-talets demo-grafiska utmaningar.

1999

- Dagens och drivkrafter – en ESO-rapport om 2000-talets demografiska utmaningar.
- Återvinning utan vinning – en ESO-rapport om sopor.
- En akademisk fråga – en ESO-rapport om rankning av C-uppsatser.
- Hederlighetens pris – en ESO-rapport om korruption.
- Samhällets stöd till de äldre i Europa – en ESO-rapport om fördelningspolitik och offentliga tjänster.
- Regionalpolitiken – en ESO-rapport om tro och vetande.
- Att snubbla in i framtiden – en ESO-rapport om statlig omvandling och avveckling.
- Att reda sig själv – en ESO-rapport om rederier och subventioner.
- Bostad sökes – en ESO-rapport om de hemlösa i folkhemmet.
- Att ta sig ton – en ESO-rapport om svensk musikexport 1974–1999.
- Med backspegeln som kompass – en ESO-rapport om stabiliseringspolitiken som läroprocess.
- Med backspegeln som kompass – ett ESO-seminarium om stabiliseringspolitik som läroprocess.

1998

- Staten och bolagskapitalet – om aktiv styrning av statliga bolag.

- Kommittéerna och bofinken – kan en kommitté se ut hur som helst?
- Regeringskansliet inför 2000-talet – rapport från ett ESO-seminarium.
- Att se till eller titta på – om tillsynen inom miljöområdet.
- Arbetsförmedlingarna – mål och drivkrafter.
- Kommuner Kan! Kanske! – om kommunal välfärd i framtiden.
- Vad kostar en ren? – en ekonomisk och politisk analys.

1997

- Fisk och Fusk – Mål, medel och makt i fiskeripolitiken.
- Ramar, regler, resultat – vem bestämmer över statens budget?
- Lönar sig arbete?
- Egenföretagande och manna från himlen.
- Jordbruksstödet – efter Sveriges EU-inträde.

1996

- Kommunerna och decentraliseringen – Tre fallstudier.
- Novemberrevolutionen – om rationalitet och makt i beslutet att avreglera kreditmarknaden 1985.
- Kan myndigheter utvärdera sig själva?
- Nästa steg i telepolitiken.
- Reglering som spel – Universiteten som förebild för offentliga sektorn?
- Hur effektivt är EU:s stöd till forskning och utveckling? – En principdiskussion.

1995

- Kapitalets rörlighet Den svenska skatte- och utgiftsstrukturen i ett integrerat Europa.
- Generationsräkenskaper.
- Invandring, sysselsättning och ekonomiska effekter.
- Hushållning med knappa naturresurser Exemplet sportfiske.
- Kostnader, produktivitet och måluppfyllelse för Sveriges Television AB.

- Vad blev det av de enskilda alternativen? En kartläggning av verksamheten inom skolan, vården och omsorgen.
- Hushållning med knappa naturresurser Exempler allemansrätten, fjällen och skotertrafik i naturen.
- Företagsstödet Vad kostar det egentligen?
- Försvarets kostnader och produktivitet.

1994

- En effektiv försvarspolitik? Fredsvinst, beredskap och återtagning.
- Skatter och socialförsäkringar över livscykeln En simuleringsmodell.
- Nettokostnader för transfereringar i Sverige och några andra länder.
- Fördelningseffekter av offentliga tjänster.
- En Social Försäkring.
- Valfrihet inom skolan Konsekvenser för kostnader, resultat och segregation.
- Skolans kostnader, effektivitet och resultat En branschstudie.
- Bensinskatteförändringens effekter.
- Budgetunderskott och statsskuld Hur farliga är de?
- Den svenska insolvensrätten Några förslag till förbättringar inom konkurshanteringen m.m.
- Det offentliga stödet till partierna Inriktning och omfattning.
- Den offentliga sektorns produktivitetsutveckling 1980–1992.
- Kvalitet och produktivitet – Teori och metod för kvalitetsjusterande produktivitetsmått.
- Kvalitets- och produktivitetsutvecklingen i sjukvården 1960–1992.
- Varför kulturstöd? Ekonomisk teori och svensk verklighet.
- Att rädda liv Kostnader och effekter.

1993

- Idrott åt alla? Kartläggning och analys av idrottsstödet.
- Social Security in Sweden and Other European Countries Three Essays.

- Lönar sig förebyggande åtgärder? Exempel från hälso- och sjukvården och trafiken.
- Hur välja rätt investeringar i transportinfrastrukturen?
- Presstödens effekter en utvärdering.

1992

- Statsskulden och budgetprocessen.
- Press och ekonomisk politik tre fallstudier.
- Kommunerna som företagsägare – aktiv koncernledning i kommunal regi.
- Slutbudsmetoden ett sätt att lösa tvister på arbetsmarknaden utan konflikter.
- Hur bra är vi? Den svenska arbetskraftens kompetens i internationell belysning.
- Statliga bidrag motiv, kostnader, effekter?
- Vad vill vi med socialförsäkringarna?
- Fattigdomsfällor.
- Växthuseffekten slutsatser för jordbruks-, energi- och skattepolitiken.
- Frihandeln ett hot mot miljöpolitiken eller tvärtom?
- Skatteförmåner och särregler i inkomst- och mervärdesskatten.

1991

- SJ, Televerket och Posten bättre som bolag?
- Marginaleffekter och tröskeleffekter barnfamiljerna och barnomsorgen.
- Östyrga projekt att styra stora kommunala satsningar.
- Prestationsbaserad ersättning i hälso- och sjukvården vad blir effekterna?
- Skogspolitik för ett nytt sekel.
- Det framtida pensionssystemet två alternativ.
- Vad kostar det? Prislista för statliga tjänster.
- Metoder i forskning om produktivitet och effektivitet med tillämpningar på offentlig sektor.
- Målstyrning och resultatuppföljning i offentlig förvaltning.

1990

- Läkemedelsförmånen.
- Sjukvårdskostnader i framtiden vad betyder åldersfaktorn?
- Statens dolda kapital. Aktivt ägande: exemplet Vattenfall.
- Skola? Förskola? Barnskola?
- Bostadskarriären som en förmögenhetsmaskin.

1989

- Arbetsmarknadsförsäkringar.
- Hur ska vi få råd att bli gamla?
- Kommunal förmögenhetsförvaltning i förändring – city-kommunerna Stockholm, Göteborg och Malmö.
- Bostadsstödet – alternativ och konsekvenser.
- Produktivitetmätning av folkbibliotekens utlåningsverksamhet.
- Statsbidrag till kommuner: allt på en check eller lite av varje? En jämförelse mellan Norge och Sverige.
- Vad ska staten äga? De statliga företagen inför 90-talet.
- Beställare-utförare – ett alternativ till entreprenad i kommuner.
- Lönestrukturen och den "dubbla obalansen" – en empirisk studie av löneskillnader mellan privat och offentlig sektor.
- Hur man mäter sjukvård – exempel på kvalitet- och effektivitetmätning.

1988

- Vad kan vi lära av grannen? Det svenska pensionssystemet i nordisk belysning.
- Kvalitet och kostnader i offentlig tjänsteproduktion.
- Alternativ i jordbrukspolitiken.
- Effektiv realkapitalanvändning i kommuner och landsting.
- Hur stor blev tvåprocentaren? Erfarenheten från en besparings-teknik.
- Subventioner i kritisk belysning.
- Prestationer och belöningar i offentlig sektor.
- Produktivitetsutveckling i kommunal barnomsorg.
- Från patriark till part – spelregler och lönepolitik för staten som arbetsgivare.
- Kvalitetsutveckling inom den kommunala barnomsorgen.

1987

- Integrering av sjukvård och sjukförsäkring.
- Produktkostnader för offentliga tjänster – med tillämpningar på kulturområdet.
- Kvalitetsutvecklingen inom den kommunala äldreomsorgen 1970–1980.
- Väger ut ur jordbruksregleringen – några idéskisser.
- Att leva på avgifter – vad innebär en övergång till avgiftsfinansiering?

1986

- Offentliga utgifter och sysselsättning.
- Produktions-, kostnads-, och produktivitetsutveckling inom den offentliga finansierade utbildningssektorn 1960–1980.
- Socialbidrag. Bidragsmottagarna: antal och inkomster. Socialbidragen i bidragssystemet.
- Regler och teknisk utveckling.
- Kostnader och resultat i grundskolan – en jämförelse av kommuner.
- Offentliga tjänster – sökarljus mot produktivitet och användare.
- Svensk inkomstfördelning i internationell jämförelse.
- Byråkratiseringstendenser i Sverige.
- Effekter av statsbidrag till kommuner.
- Effektivare sjukvård genom bättre ekonomistyrning.
- Samhällsekonomiskt beslutsunderlag – en hjälp att fatta bättre beslut.
- Produktions-, kostnads- och produktivitetsutveckling inom armén och flygvapnet 1972–1982.

1985

- Egen regi eller entreprenad i kommunal verksamhet – möjligheter, problem och erfarenheter.
- Sociala avgifter – problem och möjligheter inom färdtjänst och hemtjänst.
- Skatter och arbetsutbud.
- Produktions-, kostnads- och produktivitetsutveckling inom vägsektorn.

- Organisationer på gränsen mellan privat och offentlig sektor – förstudie.
- Frivilligorganisationer alternativ till den offentliga sektorn?
- Transfereringar mellan den förvärvsarbetande och den äldre generationen.
- Produktions-, kostnads- och produktivitetsutveckling inom den sociala sektorn 1970–1980.
- Produktions-, kostnads- och produktivitetsutveckling inom offentligt bedriven hälso- och sjukvård 1960–1980.
- Statsskuldräntorna och ekonomin effekter på den samlade efterfrågan i samhället.

1984

- Återkommande kostnads- och prestationsjämförelser – en metod att främja effektivitet i offentlig tjänsteproduktion.
- Parlamentet och statsutgifterna hur finansmakten utövas i nio länder.
- Transfereringar och inkomstskatt samt hushållens materiella standard.
- Marginella expansionsstöd ekonomiska och administrativa effekter.
- Är subventioner effektiva?
- Konstitutionella begränsningar i riksdagens finansmakt – behov och tänkbara utformningar.
- Perspektiv på budgetunderskottet, del 4. Budgetunderskott, utlandsupplåning och framtida konsumtionsmöjligheter. Budgetunderskott, efterfrågan och inflation.
- Vem utnyttjar den offentliga sektorns tjänster.

1983

- Administrationskostnader för våra skatter.
- Fördelningseffekter av kommunal barnomsorg.
- Perspektiv på budgetunderskottet, del 3. Budgetunderskott, portföljeval och tillgångsmarknader. Modellsimuleringar av offentliga besparingar m.m.
- Produktivitet i privat och offentliga tandvård.
- Generellt statsbidrag till kommuner – modellskisser.

- Administrationskostnader för några transfereringar.
- Driver subventioner upp kostnader – prisbildningseffekter av statligt stöd.
- Minskad produktivitet i offentlig sektor – en studie av patent- och registreringsverket.
- Perspektiv på budgetunderskottet, del 2. Fördelningseffekter av budgetunderskott. Hushållsekonomi och budgetunderskott.
- Enhetligt barnstöd? några variationer på statligt ekonomiskt stöd till barnfamiljer.
- Staten och kommunernas expansion några olika styrmedel.

1982

- Ökad produktivitet i offentlig sektor – en studie av de allmänna domstolarna.
- Offentliga tjänster på fritids-, idrotts- och kulturområdena.
- Perspektiv på budgetunderskottet, del 1. Budgetunderskottens teori och politik. Statens budgetfinansiering och penningpolitiken.
- Inkomstomfördelningseffekter av livsmedelssubventioner. Perspektiv på besparingspolitiken.