

RAPPORT

2023

Randomiserte og kvasieksperimentelle studier i den statlige forvaltningen: en kartlegging

Rapport

Randomiserte og kvasieksperimentelle studier i den statlige forvaltningen: en kartlegging

Unni Gopinathan

Atle Fretheim

Utgitt av:

Senter for forskning på epidemiltak (CEIR), Folkehelseinstituttet, juni 2023.

Tittel:

Randomiserte og kvasieksperimentelle studier i den statlige forvaltningen: en kartlegging.

Forfatter(e):

Unni Gopinathan, *Folkehelseinstituttet*

Atle Fretheim, *Folkehelseinstituttet*

Oppdragsgiver:

Helse- og omsorgsdepartementet/Kjernegruppen for et bedre kunnskapssystem for håndtering av kriser.

Publikasjonstype:

Rapport

Bestilling:

Rapporten kan lastes ned som pdf på Folkehelseinstituttets nettsider: www.fhi.no

Grafisk design omslag:

Fete Typer

ISBN elektronisk utgave:

978-82-8406-394-2

Sitering:

Gopinathan U, Fretheim A. Randomiserte og kvasieksperimentelle studier i den statlige forvaltningen: en kartlegging. [Randomised and quasi-experimental studies in the public sector: a mapping study]. Rapport 2023. Oslo: Folkehelseinstituttet, 2023.

Innhold

Innhold	3
Hovedbudskap	5
Sammendrag	6
Key messages (English)	10
Executive summary (English)	11
Background and process	11
Main findings	11
Conclusion	13
Forord	15
1 Innledning	16
2 Metode	17
2.1 Definisjoner	17
2.1.1 Virkemiddel	17
2.1.2 Randomiserte studier	18
2.1.3 Kvasieksperimentelle studier	18
3 Resultater	19
3.1 Effektforskning med randomiserte og kvasieksperimentelle studier: etatenes vurderinger av nytteverdi og sentrale forutsetninger	19
3.2 Randomiserte og kvasieksperimentelle studier: metodiske styrker og svakheter	21
3.3 Organisering, gjennomføring og formidling av effektevalueringer	22
3.4 Omfang og eksempler på evaluerte virkemidler	23
3.4.1 Randomiserte studier	23
3.4.2 Kvasieksperimentelle studier	24
3.5 Vitenskapelig kvalitet	25
3.6 utfordringer ved gjennomføring av RCTer og kvasieksperimentelle studier	26
3.6.1 Tilstrekkelig antall deltakere	26
3.6.2 Datagrunnlag	26
3.6.3 Juridiske utfordringer	27
3.6.4 Etske hensyn	30
3.6.5 Politiske	31
3.6.6 Andre utfordringer	32
3.7 Muligheter	34
3.7.1 Forankring og politisk aksept for tidsavgrenset forskjellsbehandling	34
3.7.2 Bygge kultur og heve kompetanse	36
3.7.3 Tydelig juridisk og etisk rammeverk	36
3.7.4 Bedre utnytting av eksisterende og utvikling av nye datakilder	37

3.7.5 God medvirkning av berørte parter	37
3.7.6 Kunnskapsorientert mål- og resultatstyring i forvaltningen	37
3.7.7 Finansiering og langsiktige forskningsprogram	37
3.7.8 Veiledning	38
4 Konklusjon	39
Vedlegg 1: Spørreskjema	42
Vedlegg 2: Barne-, ungdoms- og familiedirektoratets (Bufdir) erfaringer med randomiserte og kvasiekperimentelle studier	47
Vedlegg 3: Folkehelseinstituttets erfaringer med randomiserte og kvasiekperimentelle studier	50
Vedlegg 4: NAVs erfaringer med randomiserte og kvasiekperimentelle studier	53
Vedlegg 5: Norads erfaringer med randomiserte og kvasiekperimentelle studier	56
Vedlegg 6: Skatteetatens erfaringer med randomiserte og kvasiekperimentelle studier	59
Vedlegg 7: Utdanningsdirektoratets erfaringer med randomiserte og kvasiekperimentelle studier	62

Hovedbudskap

De fleste etatene som bidro til denne kartleggingen, har gjennomført eller anskaffet randomiserte og kvasiexperimentelle studier. Eksempler på tiltak som er blitt evaluert ved hjelp av slike metoder er organisatoriske, pedagogiske og økonomiske virkemidler.

Tett dialog mellom fagetater, departement og politisk ledelse har bidratt til bedre effektevalueringer. Potensialet er stort for å dele kunnskap mellom sektorer om hvordan effektevalueringer kan forberedes og gjennomføres. Flere etater erfarer imidlertid politisk tilbøyelighet til storskala iverksetting av virkemidler framfor gradvis innføring som er nødvendig for å kunne evaluere effekter ved hjelp av randomiserte eller kvasiexperimentelle metoder.

Randomiserte og kvasiexperimentelle studier møter mange juridiske, etiske, politiske og praktiske utfordringer, og de kan være ressurskrevende. Flere etater påpeker behovet for bedre tilrettelagt lovgivning for effektstudier, med tydeligere kriterier for a) når forskjellsbehandling kan være forsvarlig, b) når det kan gis unntak fra kravet om informert samtykke, og c) håndtering av personvern i forbindelse med datatilgang.

Sammendrag

Bakgrunn og prosess

I tildelingsbrevet for 2023 fikk Folkehelseinstituttet (FHI) i oppdrag å koordinere en kartlegging av hvordan relevante statlige etater bruker randomiserte studier (RCTer) og kvasiexperimentelle metoder for å gjøre virkemiddelbruken mer kunnskapsbasert. Kartleggingen tar sikte på å avdekke styrker og svakheter ved å bruke slike evalueringsmetoder, og hva som fremmer og hindrer bruken av dem. Kartleggingen omfatter i hovedsak Barne-, ungdoms- og familiedirektoratet (Bufdir), NAV, Direktoratet for Utviklingssamarbeid (Norad), Skatteetaten, Utdanningsdirektoratet (Udir) og FHI.

Hovedfunn

Etatene som har deltatt i denne undersøkelsen påpeker at **randomiserte og kvasiexperimentelle studier har styrker og svakheter som påvirker i hvilken grad metodene gir pålitelige konklusjoner**. Det vises til at i RCTer gjøres fordelingen til tiltaks- og kontrollgrupper tilfeldig, som minimerer systematiske forskjeller mellom de som mottar og ikke mottar tiltaket. Dermed kan det trekkes pålitelige konklusjoner om at observerte forskjeller mellom gruppene skyldes virkningen av tiltaket som blir studert. Samtidig understrekes det at kvasiexperimentelle metoder også kan gi gyldige effektestimater gitt visse forutsetninger, som variasjon i iverksettelse av virkemiddelet, og tilgang til detaljerte data og datakoblinger.

En av de største metodiske utfordringene ved RCTer er at resultatene ikke nødvendigvis er overførbare utenfor den sammenhengen der studien ble gjennomført. Etatene har også erfart at effekten av et tiltak rettet mot grupper (f.eks. skoler eller kommuner) kan «smitte» mellom tiltaks- og kontrollgruppene, og dermed gi misvisende resultater. Den viktigste metodiske svakheten ved kvasiexperimentelle studier er faren for at resultatene er påvirket av systematiske skjevheter (bias) pga. andre forskjeller mellom gruppene enn det aktuelle tiltaket.

De fleste etatene som har bidratt til kartleggingen har erfaringer med å gjennomføre eller anskaffe randomiserte og kvasiexperimentelle studier. Organisatoriske, pedagogiske og økonomiske virkemidler, samt virkemidler for offentlig myndighetsutøvelse (f.eks. tiltak for å øke etterlevelsen av lover og regler) har blitt evaluert med disse metodene. Hvor mange studier etatene anslo å ha gjennomført i løpet av de siste fem årene varierte fra 1-5 studier til over 15. Forvaltningsorganene som har gjennomført over 15 er de som stort sett utformer og gjennomfører studier selv.

Etatene rapporterer mange av de samme utfordringene for gjennomføring av RCTer og kvasiexperimentelle studier. Det gjelder bl.a. datagrunnlag og politisk aksept for gradvis utrulling og forskjellsbehandling når virkemidler iverksettes. I noen tilfeller er det spesifikke utfordringer knyttet til en av metodene. Det gjelder spesielt gjennomføring av RCTer, som ofte støter på juridiske, etiske og praktiske utfordringer. En betydelig praktisk utfordring, særlig for mange RCTer, er oppnåelse av tilstrekkelig statistisk styrke. Kvasiexperimentelle studier er ofte mer gjennomførbare, noe som understreker viktigheten av å betrakte disse metodene som supplerende tilnærminger.

For begge metodene er **tilgang til registerdata og kobling av registre** fremtredende utfordringer. Godt gjennomførte kvasiexperimentelle studier forutsetter ofte detaljerte

data og omfattende datakoblinger. Mange slike studier forhindres av at det er tidkrevende og vanskelig å få tilgang til slike data. Når spørreundersøkelser benyttes som datagrunnlag er utilstrekkelig oppslutning et vanlig problem. Videre er fravær av sammenlignbare og presise mål for effekt også en utfordring, spesielt dersom det er manglende hyppighet og kontinuitet i målingene som benyttes og eksisterende datakilder er ufullstendige eller ikke utviklet for forskningsformål.

Juridiske utfordringer inkluderer

- krav om informert samtykke som kan bety at befolkningsrettede studier er umulige å gjennomføre
- at studier der deltakerne blir fordelt i to grupper som behandles forskjellig kan komme i konflikt med prinsipper om likebehandling og lovbestemmelser som beskytter grunnleggende rettigheter og tjenester
- sektorspesifikke utfordringer som forhindrer finansiering eller praktisk gjennomføring av effektevalueringer

Etiske hensyn inkluderer

- motstand mot å ikke tilby kontrollgruppen et tiltak som forventes å ha positive virkninger
- at studier kan innebære uforholdsmessig risiko eller byrde for deltakerne
- innvendinger mot studier der deltakerne ikke har mulighet til å trekke sin deltakelse, enten fordi de ikke er klar over at de deltar i en studie, eller fordi det ikke er praktisk mulig å trekke seg

Forskningsetiske vurderinger av ett og samme virkemiddel blir for øyeblikket vurdert ulikt, avhengig av hvilket lovverk som kommer til anvendelse, f.eks. når det gjelder krav om informert samtykke. Tydeligere juridiske og etiske rammer er nødvendig for å kunne avveie samfunnsnyttene av å gjennomføre en studie opp mot risiko, byrder og personvern hensyn.

Politiske utfordringer inkluderer

- beslutningstagernes ønske om å utvise handlekraft, som ikke gir tid og rom for grundige effektevalueringer før et virkemiddel iverksettes i stor skala
- behovet for økt politisk forståelse for at systematisk utprøving i form av randomiserte eller kvasiexperimentelle studier kan være et viktig verktøy for å redusere usikkerheten rundt effekten av et virkemiddel
- politiske prosesser, f.eks. regjeringendringer, budsjettforlik og politiske vedtak, som kan endre konteksten virkemiddelet testes i, og innebære at kontrollgruppen blir eksponert for tiltaket som blir evaluert

Planlegging og gjennomføring av flere og bedre effektevalueringer kan skje gjennom tett dialog mellom fagetater, departementer og forskningsmiljøer. Informasjons- og holdningsarbeid kan bidra til å skape større tverrpolitisk aksept for tidsavgrenset, systematisk og gradvis – og om mulig, randomisert – utrulling av reformer og virkemidler.

Tverrpolitisk enighet kan også fremme langsiktig finansiering for større forskningssatsninger som kan gi kunnskap om myndighetenes politikk og virkemiddelbruk. Kartleggingen fremhever positive eksempler, som LÆREEFFEKT-programmet finansiert av Forskningsrådet, og Kompetanseprogrammet som ledes av Direktoratet for høyere

utdanning og kompetanse, hvor innsats og ressurser er prioritert for systematisk utprøving av virkemidler.

For RCTer spesielt gjelder også andre utfordringer av praktisk, økonomisk og psykologisk art, som

- at gjennomføring i komplekse organisasjoner og tjenester kan være praktisk utfordrende for personalet ute i tjenestene, og kan påvirke vanlig drift
- at studiene ofte er kostnadskreven
- skepsis blant tjenesteytere og i befolkningen mot forskningsstudier som innebærer en forskjellsbehandling mellom de som får og de som ikke får tiltaket

Kostnaden ved å gjennomføre randomiserte og kvasiexperimentelle studier bør vurderes opp mot besparelsene som oppnås ved å unngå å iverksette tiltak med usikker effekt. Medvirkning av dem som berøres av studien, som undervisningspersonale, helsepersonell, saksbehandlere eller befolkningen generelt, er nødvendig for å skape aksept for en tidsavgrenset forskjellsbehandling som kan bidra til mer og bedre kunnskap om effekten av virkemidler. God medvirkning støttet av implementeringsstudier kan bidra til at evalueringer ikke medfører uforholdsmessig byrde for de berørte tjenestene. Bruk av tilgjengelig infrastruktur, f.eks. nasjonale registre, kan gjøre studier mer administrativt og praktisk gjennomførbare.

For å begrense bruk av ressurser på studier som det ikke kan trekkes pålitelige konklusjoner fra, er det viktig at planleggingsfasen omfatter et godt utredet kunnskapsgrunnlag, og at det legges vekt på høy vitenskapelig standard. Dette innebærer f.eks. utvikling og forhåndsregistrering av protokoller med pre-spesifiserte analyser, i tråd med internasjonale retningslinjer. Videre er det viktig at det investeres i utvikling av datakilder og at det gis mulighet for tilgang til, og sammenkobling av disse.

Kompetansen som ønskes for å gjennomføre flere effektevalueringer varierer mellom forvaltningsorganene, men på to områder var de samstemte: For det første trengs mer kunnskap om forskningsmetodikk, inkludert utforming av effektstudier. For det andre er mer juridisk kompetanse nødvendig for å forstå gjeldende regelverk, avklare rettslige rammer og for å kunne utnytte handlingsrommet som regelverket gir.

Konklusjoner

Kartleggingen har identifisert flere eksempler på hvordan tett dialog mellom fagetater, departement og politisk ledelse har tilrettelagt for bedre effektevalueringer. Potensialet er stort for å dele kunnskap mellom sektorer om hvordan effektevalueringer kan forberedes og gjennomføres. Flere etater erfarer imidlertid politisk tilbøyelighet til å rulle ut programmer med usikker effekt i stor skala heller enn gradvis innføring som kan tilrettelegge for effektevaluering med randomiserte eller kvasiexperimentelle metoder, før eventuell oppskalering.

Randomiserte og kvasiexperimentelle studier møter mange juridiske, etiske, politiske og praktiske utfordringer, og de kan være ressurskrevende. For mange virkemidler er det mer gjennomførbart med kvasiexperimentell evaluering enn en RCT. For å oppnå troverdige resultater med kvasiexperimentelle metoder må flere forutsetninger være oppfylt, inkludert variasjon i implementeringen av virkemiddelet, tilgang på detaljerte data og større datakoblinger, og solid metodekompetanse. Flere etater påpeker behovet for bedre tilrettelagt lovgivning for effektstudier, med tydeligere kriterier for a) når

forskjellsbehandling kan være forsvarlig, b) når det kan gis unntak fra kravet om informert samtykke, og c) håndtering av personvern i forbindelse med datatilgang.

En større politisk aksept for systematisk, gradvis og – om mulig – randomisert, utrulling av reformer vil kunne tilrettelegge for bedring læring om virkemidlenes effekt. Dette vil kunne føre til en mer effektiv og målrettet politikkutforming og øke sjansene for at virkemidlene faktisk fører til ønsket effekt, og at uvirksomme virkemidler prioriteres vekk.

Key messages (English)

Most government agencies that contributed to this survey have carried out or commissioned randomized and quasi-experimental studies. Organizational, educational, and financial interventions are examples of measures that have been evaluated using such methods.

Close dialogue between implementers, ministries and political leadership has facilitated better impact evaluations. The potential is great for sharing knowledge between sectors about how impact evaluations can be prepared and carried out. However, several agencies experience a political inclination to large scale roll out of interventions with uncertain effects rather than gradual implementation, which could allow for evaluating effectiveness with randomized or quasi-experimental methods, before scale-up.

Randomized and quasi-experimental studies face many legal, ethical, political and practical challenges, and they may be resource intensive. Several agencies point out the need for legislation that facilitates the conduct of impact evaluations, with clearer criteria for a) when differential treatment may be considered justifiable, b) when informed consent requirements can be waived, and c) managing privacy concerns related to data access.

Executive summary (English)

Background and process

In the allocation letter for 2023, the Norwegian Institute of Public Health (FHI) was commissioned to coordinate a survey of how relevant government agencies use randomized trials (RCTs) and quasi-experimental methods to strengthen the evidence base for public policy decisions. The mapping exercise aims to identify strengths and weaknesses with the use of such evaluation methods, and barriers and facilitators to their use. The main contributors to this mapping were the Directorate for Children, Youth and Family Affairs (Bufdir), the Norwegian Labour and Welfare Administration (NAV), the Norwegian Agency for Development Cooperation (Norad), the Norwegian Tax Administration, the Directorate for Education and Training (Udir) and FHI.

Main findings

The agencies that participated in this survey emphasized that **RCTs and quasi-experimental studies have strengths and weaknesses that affect the extent to which the methods provide reliable conclusions**. For RCTs, it is pointed out that random allocation to intervention or control group minimizes systematic differences between those who receive and those who do not receive the intervention. Thus, reliable conclusions can be drawn about whether observed differences between the groups can be attributed to the effect of the measure that is being studied. At the same time, it is emphasized that quasi-experimental methods can also provide valid effect estimates given certain assumptions, such as variation in the implementation of the intervention and access to detailed data and linking of different data sources.

One of the main methodological challenges with RCTs is that the results are not necessarily transferable outside the context in which the study was carried out. The agencies have also experienced that the effect of an intervention aimed at groups (e.g. schools or municipalities) may "leak" between the intervention and control groups, and thus produce misleading results. The primary methodological limitation of quasi-experimental studies is the risk for biased results due to differences between the groups, not related to the intervention being studied.

Most of the agencies that contributed to the survey have experience conducting or commissioning randomized and quasi-experimental studies. Organizational, educational, and financial interventions, as well as interventions for public policy governance (e.g. measures to improve adherence to laws and regulation) have been evaluated using such methods. Over the last five years, the estimated number of such evaluations ranged from 1 to 5 studies for some agencies, and exceeded 15 for others. The agencies that had completed >15, were those that for that usually designed and carried out studies themselves.

The agencies report similar barriers for conducting RCTs and quasi-experimental studies, like access to necessary data and political acceptance for gradual roll-out and differential treatment when introducing new measures. In some cases, there may be specific challenges associated with one of the methods. Implementing RCTs often encounters legal, ethical, and practical challenges. A significant practical challenge, especially for many RCTs, is to include a large enough sample to ensure the necessary statistical power. Quasi-

experimental studies are often more feasible, which highlights the importance of considering RCTs and quasi-experimental methods as complementary approaches.

For both methods, **access to registry data and linking of registries** are key challenges. Well-conducted quasi-experimental studies often require detailed data and extensive linking of data sources. Many such studies are prevented by the fact that it is time-consuming and difficult to access such data. When surveys are used as data source, achieving sufficient response rates is a common problem. Furthermore, the absence of comparable and precise data that can serve as outcome measures is also a challenge, especially if there is a lack of frequent and continuous measurements or if existing data sources are incomplete or not developed for the purpose of research.

Legal challenges include

- informed consent requirements, which can mean that population-based studies are impossible to carry out
- that studies where the participants are divided into two differentially treated groups may come into conflict with principles of equal treatment and legal provisions that protect basic rights and services
- sector-specific challenges that prevent funding or the practical implementation of impact evaluations

Ethical considerations include

- resistance to not offering the control group a measure that is expected to have positive effects
- that studies may entail disproportionate risk or burden for the participants
- objections to studies where the participants do not have the opportunity to withdraw their participation, either because they are not aware that they are participating in a study, or because it is not practically possible to withdraw

Ethical assessments of the one and the same intervention are currently assessed differently, depending on the legislation that is applied, e.g. with regard to informed consent requirements. Well-defined legal and ethical frameworks are needed to provide guidance for assessing the societal benefit of a study against its potential risks, burdens, and privacy implications.

Political challenges include

- the decision-makers' desire to appear efficient and action oriented, which does not allow for the time and space needed to carry out thorough impact evaluations before a measure is implemented on a large scale
- the need for heightened political awareness about how systematic testing, using randomized or quasi-experimental studies, can be an important tool for reducing the uncertainty about the effects of a measure
- political processes, such as changes in government, budget settlements and political decisions, can change the context in which the measure is evaluated, which may lead to the control groups being exposed to the measure under study

Planning and carrying out more and better impact evaluations can be done through close dialogue between agencies, ministries, and researchers. Enhanced exchange of information, along with promoting changes in attitudes, can lead to greater cross-party acceptance for a time-limited, systematic, and gradual – and if possible, randomized – rollout of reforms and measures.

Cross-party agreement can also facilitate longer-term funding, which are necessary for major research initiatives that can generate knowledge about public policy interventions and programs. The survey findings highlight positive examples, such as the LÆREEFFEKT program funded by the Norwegian Research Council and the Competencies Program led by the Directorate for Higher Education and Skills, where resources for systematic evaluation have been prioritized.

Especially for RCTs, other challenges of a practical, economic, and psychological nature also apply, e.g.

- that implementation in complex organizations and services can be practically challenging for the staff involved in service delivery, and may affect day-to-day operations
- that conducting the studies can be expensive
- skepticism among service providers and in the general population towards research studies that involve differential treatment between those who receive and those who do not receive the measure

The cost of carrying out randomized and quasi-experimental studies should be assessed against the savings achieved by not implementing measures that have uncertain or undesirable effects. The participation of those who are affected by the study, such as teaching staff, health care workers, case managers or the general population, is necessary to ensure acceptance for time-limited differential treatment to gain more and better knowledge about the effect of interventions. Appropriate user involvement supported by implementation studies can help ensure that studies do not entail a disproportionate burden on the services affected. Use of available data infrastructure, e.g. national registers, can make studies more administratively and practically feasible.

To limit the use of resources on studies from which reliable conclusions cannot be drawn, it is important that the planning phase includes a review of the existing evidence base, and emphasis on sound scientific standards. This includes development and pre-registration of protocols with pre-specified analyses, in line with international guidelines. Furthermore, it is important that investments are made in the development of data sources and that access to and linkage across data sources is made possible.

The expertise needed to carry out impact evaluations varies across agencies, but views align on two areas. First, higher competency in research methodology is required, including the design of impact studies. Second, more legal expertise is needed to understand the current regulations, clarify the legal basis for studies, and to make use of the room to maneuver within the existing legal framework.

Conclusion

Our mapping exercise has identified several examples of how close dialogue between agencies, ministries and the political leadership has facilitated better impact evaluations.

There is a potential for improved sharing of knowledge between sectors about how impact evaluations can be prepared and carried out. However, several agencies experience a political inclination to large scale roll out of interventions with uncertain effects rather than gradual introduction. which could allow for evaluating effectiveness with randomized or quasi-experimental methods, before scale-up. Randomized and quasi-experimental studies face many legal, ethical, political, and practical challenges, and they can be resource intensive. For many interventions, quasi-experimental evaluation is more feasible than conducting an RCT. To achieve credible results with quasi-experimental methods, several prerequisites must be met, including variation in the implementation of the interventions, access to detailed data and linkages across data sources, and sufficient methodological expertise. Several agencies point out the need for legislation that facilitates the conduct of impact evaluations, with clearer criteria for a) when differential treatment may be considered justifiable, b) when informed consent requirements can be waived, and c) managing privacy concerns related to data access.

A greater political acceptance for systematic, gradual and – if possible – randomized roll-out of reforms would facilitate learning about the effects of interventions. This could lead to more effective and targeted policy making and increase the chances that implemented measures lead to the desired results, and that ineffective measures are discontinued.

Forord

I tildelingsbrevet for 2023 fikk Folkehelseinstituttet (FHI) i oppdrag å koordinere en kartlegging av hvordan relevante statlige etater bruker randomiserte studier og kvasiekperimentelle metoder for å gjøre virkemiddelbruken mer kunnskapsbasert. Kartleggingen tar sikte på å avdekke svakhetene og begrensningene ved å bruke slike metoder, og hva som fremmer og hindrer bruken av dem. Kartleggingen skulle omfatte Barne-, familie- og ungdomsdirektoratet (Bufdir), NAV, Utdanningsdirektoratet (Udir), Skatteetaten og Direktoratet for Utviklingssamarbeid (Norad) i tillegg til FHI. I tillegg skulle andre etater i statlig og kommunal sektor vurderes. Denne rapporten presenterer kartleggingens funn.

Oslo, mai 2023

Jan Frich

Områdedirektør, Helsetjenester, FHI

1 Innledning

Koronapandemien understreket betydningen av et sterkt kunnskapssystem for at myndighetene skal kunne ta kunnskapsbaserte beslutninger. En interdepartemental gruppe – Kjernegruppen for et bedre kunnskapssystem for håndtering av kriser – er etablert for å bidra til at norske myndigheter har tilgang til rask og pålitelig kunnskap til støtte for beslutninger under fremtidige kriser. I 2022 nedsatte kjernegruppen to ekspertgrupper som produserte følgende rapporter: (1) Effektiv og sikker infrastruktur for tilgang til, deling og bruk av relevant statistikk og data i kriser; (2) Juridiske og etiske problemstillinger knyttet til innsamling, tilgjengeliggjøring, deling og bruk av data, samt bruk av randomiserte forsøk, i kriser.^{1,2} Denne kartleggingen bør leses i sammenheng med disse rapportene.

Kjernegruppen har pekt ut fire områder for videre arbeid på dette temaet.³ Ett av dem er at departementene bør stimulere til bedre evalueringer, inkludert bedre tilrettelegging for randomiserte og kvasiexperimentelle studier. Et tilbakevendende problem under pandemien var mangel på kunnskap om effektene av ulike smitteverntiltak. Flere randomiserte studier kunne ikke gjennomføres, bl.a. av etiske, juridiske og politiske grunner. Ekspertgruppen som utredet juridiske og etiske problemstillinger knyttet til bruk av randomiserte forsøk i kriser konkluderte at myndighetene under pandemien ikke iverksatte tiltak på en måte som tilrettela for læring om effekter. Videre var det usikkerhet rundt forvaltningens hjemler for å iverksette randomiserte forsøk eller tilrettelegge for kvasiexperimentelle studier. Et viktig moment var at helseforskningslovens krav om individuelt samtykke gjorde det umulig å gjennomføre flere planlagte effektevalueringer av smitteverntiltak.

Mer og bedre bruk av randomiserte og kvasiexperimentelle studier for å gjøre myndighetenes virkemiddelbruk mer kunnskapsbasert vil være i tråd med utredningsinstruksen, som instruerer forvaltningen til å legge et godt kunnskapsgrunnlag for beslutninger om statlige tiltak. Flere offentlige utredninger, som produktivitetskommisjonens utredning og utredningen til ekspertgruppen om livslang læring, har tatt til orde for bruk av kontrollerte eksperimenter for systematisk utprøving av virkemidler. En viktig lærdom fra pandemien er at kompetanse og kapasitet, samt et tydelig juridisk og etisk grunnlag for effektevalueringer, bør være under kontinuerlig utvikling før kriser inntreffer. En effektiv forskningsberedskap krever grundig erfaring med å gjennomføre randomiserte og kvasiexperimentelle studier, også når det gjelder virkemidler som iverksettes utenom krisetider. Denne kartleggingen gir et innblikk i den statlige forvaltningens erfaringsgrunnlag og forutsetninger, samt sentrale utfordringer og muligheter for å fremme kunnskapsbasert virkemiddelbruk gjennom bruk av randomiserte og kvasiexperimentelle studier.

¹<https://www.regjeringen.no/contentassets/436542ebbc77473b96e4c45c5668b07a/effektiv-og-sikker-infrastruktur-for-deling-av-relevant-statistikk-og-data-i-kriser.pdf>

²<https://www.regjeringen.no/contentassets/436542ebbc77473b96e4c45c5668b07a/juridiske-og-etiske-problemstillinger-knyttet-til-innsamling-tilgjengeliggjoring-deling-og-bruk-av-data-samt-bruk-av-randomiserte-forsok-i-kriser.pdf>

³Mandat for kjernegruppen for et bedre kunnskapssystem for håndtering av kriser.

<https://www.regjeringen.no/no/aktuelt/mandat-for-kjernegruppen-for-et-bedre-kunnskapssystem-for-handtering-av-kriser/id2965367/>

2 Metode

FHI utviklet et utkast til spørreskjema som ble delt med forvaltningsorganene som var nevnt i tildelingsbrevet: Bufdir, NAV, Norad, Skatteetaten og Udir. FHI hadde også digitale en-til-en møter med representanter fra disse forvaltningsorganene for å avklare spørsmål og oppnå en felles forståelse om formålet med kartleggingen. Etter tilbakemeldinger ble skjemaet ferdigstilt (Vedlegg 1), og sendt ut med svarfrist på tre uker.

Besvarelsene ble deretter sammenfattet i et førsteutkast til rapport, som ble sendt til alle forvaltningsorganene for tilbakemeldinger innen en uke. Et revidert utkast ble så sendt med mulighet for en siste runde med tilbakemeldinger innen to virkedager.

Supplerende erfaringer og perspektiver ble også hentet fra Helsedirektoratet og Statistisk Sentralbyrå gjennom muntlige samtaler og kortere skriftlige innspill.

Proessen la opp til å skaffe et innblikk i hvordan randomiserte og kvasieksperimentelle metoder brukes. Vi har ikke hatt noen ambisjon om å lage en komplett oversikt over bruken av disse metodene, verken i offentlig forvaltning generelt, eller i de forvaltningsorganene som har bidratt i denne rapporten.

Vi takker følgende personer for bidrag og innspill: Gaute Eielsen (NAV), Bernhard Weigel (NAV), Elin Selboe (Udir), Eivind Moe Hammersmark (Norad), Eirik Mofoss (Norad), Silje Jenny Undahl (Skatteetaten), Nina Serdarevic (Skatteetaten), Arnstein Øvrum (Skatteetaten), Therese Dalsbøe (Bufdir), Sigurd Røtnes (Hdir), Ingunn Bakkene Cowan (Lånekassen), Magnus Fodstad Larsen (HK-dir), Olena Tkachenko (HK-dir), Bengt Oscar Lagerstrøm (SSB), Arild Ragnøy (Statens Vegvesen), Arnfinn Helleve (FHI), Tone Bruun (FHI), Petter Elstrøm (FHI), Jonas Minet Kinge (FHI), Rannveig Kaldager Hart (FHI), Ingeborg Margrete Rossow (FHI), Daniel Bergsvik (FHI), Maja Weemes Grøtting (FHI), Kjetil Telle (FHI), Hege Marie Gjefsen (FHI), Trygve Ottersen (FHI), Camilla Stoltenberg (FHI) og Mona Bjørnbæk (FHI),

2.1 Definisjoner

Definisjoner av tre sentrale begreper i denne kartleggingen ble utformet for å sikre en felles forståelse blant etatene som besvarte spørreskjemaet:

2.1.1 Virkemiddel

Statlige, kommunale eller fylkeskommunale etater eller virksomheter kan bruke en rekke virkemidler⁴ for å iverksette politikk. Slike virkemidler inkluderer:

- juridiske virkemidler, f.eks. nye lover og forskrifter
- økonomiske virkemidler, f.eks. endringer i skatter og avgifter eller insentiver for å påvirke atferd
- organisatoriske virkemiddel, f.eks. omorganisering, sentralisering og desentralisering av statlige oppgaver

⁴Vi har tatt utgangspunkt i klassifiseringen som benyttes av Direktoratet for forvaltning og økonomistyring: https://dfo.no/sites/default/files/fagomr%C3%A5der/Rapporter/Rapporter-Difi/hva_er_god_statlig_styring_difi-rapport_2019_9_0.pdf

- pedagogiske virkemidler, f.eks. veiledere, retningslinjer, kompetanseutviklingstiltak eller holdningskampanjer
- offentlig myndighetsutøvelse, f.eks. gjennom kontroll og sanksjon for å øke etterlevelsen av lover og regler

Disse virkemidlene, i større eller mindre skala, kan være utgangspunkt for effektevaluering med randomiserte eller kvasieksperimentelle studier. Vi bruker begrepene «virkemiddel» og «tiltak» som synonymer i denne rapporten.

2.1.2 Randomiserte studier

I randomiserte studier (RCTer) måles effektene av et virkemiddel ved å sammenligne deltakere som fordeles tilfeldig – ved loddtrekning eller tilsvarende – til å bli eksponert eller ikke bli eksponert for virkemiddelet. Randomisering kan også foregå gjennom gradvis og tilfeldig innfasing av et virkemiddel. Deltakerne kan være individer eller grupper av individer, f.eks. skoleklasser, skoler eller kommuner. Sistnevnte kalles vanligvis klyngerandomiserte studier. Randomisering av tilstrekkelig antall deltakere/klynger sannsynliggjør at intervensjonsgruppen og kontrollgruppen er like mht. kjennetegn som kan påvirke utfallet, slik at eventuelle forskjeller i utfallet kan tilskrives virkemiddelet.

2.1.3 Kvasieksperimentelle studier

I kvasieksperimentelle studier vil man i varierende grad kunne avdekke kausale effekter. Årsaken er at i kvasieksperimentelle studier er det ikke på forhånd gjort en tilfeldig tilordning av individer eller klynger i en tiltaks- og en kontrollgruppe. Årsakssammenhenger kan likevel la seg identifisere dersom visse forutsetninger, som variasjon i implementeringen av tiltaket, er oppfylt. I hvilken grad disse forutsetningene er oppfylt vil variere mellom ulike typer kvasieksperimentelle studier. Dermed vil også muligheten for sikker avdekking av kausale effekter variere.

3 Resultater

3.1 Effektforskning med randomiserte og kvasiexperimentelle studier: etatenes vurderinger av nytteverdi og sentrale forutsetninger

Etatene som deltok i denne kartleggingen var samstemte om nytten av og metodologiske utfordringer ved randomiserte og kvasiexperimentelle studier.

Disse metodene har en felles målsetning: å isolere effekten av et virkemiddel på ett eller flere definerte utfallsmål, og måle størrelsen på denne effekten. Viktige utfallsmål som etatene i denne kartleggingen har bidratt til å måle inkluderer helseindikatorer, fattigdomsreduksjon, ulike mål for barn og unges trivsel eller skoleprestasjoner, sykefravær og arbeidsdeltakelse, og korrekt rapportering av skattbar inntekt i selvangivelsen.

Den sentrale forskjellen mellom randomiserte og kvasiexperimentelle studier ligger i hvordan virkemiddelet er fordelt til tiltaks- og kontrollgruppene. En randomisert studie (RCT) innebærer at tiltaket er tilfeldig fordelt ved loddtrekning eller tilsvarende metode. Randomisering – dersom optimalt gjennomført – sikrer at tiltaks- og kontrollgruppene i gjennomsnitt er like på både observerbare og uobserverbare faktorer. Ved å minimere systematiske forskjeller mellom tiltaks- og kontrollgrupper kan det trekkes pålitelige konklusjoner om hvorvidt observerte forskjeller mellom gruppene kan tilskrives virkningen av tiltaket som blir studert.

Kvasiexperimentelle studier utgjør en annen viktig kilde til kunnskap om effekter. Begrepet «kvasiexperimentell» er et sekkebegrep for ulike analytiske tilnærminger, bl.a. forskjell-i-forskjeller, avbrutte tidsserier, instrument-variabler og regresjonsdiskontinuitet («regression discontinuity» - RD-design), som forsøker å måle kausale effekter av et virkemiddel. Siden tilordningen til tiltaks- og kontrollgrupper ikke har skjedd tilfeldig må visse forutsetninger være oppfylt for at kvasiexperimentelle metoder skal kunne gi gyldige effektestimater. Dette gjelder spesielt strategier for å oppnå sammenlignbare tiltaks- og kontrollgrupper. Etatene har fremhevet flere sentrale forutsetninger som bør være oppfylt for at kvasiexperimentelle metoder skal kunne gi pålitelige effektestimater:

- **Variasjon i implementeringen av virkemiddelet:** Muligheten for kvasiexperimentelle studier oppstår når et virkemiddel iverksettes for noen grupper før andre. Ofte identifiserer forskere slik naturlig variasjon retrospektivt, men kvasiexperimentelle studier kan også planlegges prospektivt gjennom tett samarbeid mellom forskere og beslutningstakere slik at det er mulig å sammenligne grupper som eksponeres og ikke eksponeres for virkemiddelet (Boks 1). Variasjon i implementeringen kan oppnås gjennom gradvis utrulling av virkemiddelet eller ved å definere klare grenseverdier som avgjør hvilke grupper som får tiltaket først. Sistnevnte kan legges til rette for regresjonsdiskontinuitetsstudier (Boks 2). I begge tilfeller er målet å oppnå tiltaks- og kontrollgrupper som er sammenlignbare.
- **Detaljerte data:** Å utnytte et kvasiexperiment krever ofte svært detaljerte data for å skille kausalitet fra seleksjon, og det kan være at dataene ikke finnes på tilstrekkelig detaljert nivå. Det kan f.eks. dreie seg om fødselsdato (for tiltak som innføres for en bestemt aldersgruppe) eller dato for et legebesøk.
- **Omfattende datakoblinger:** For å oppnå gyldige resultater krever kvasiexperimentelle studier ofte kobling av data fra flere datasett og/eller

sektorer, store befolkningsutvalg og lange tidsserier. Slike datakoblinger er bl.a. nødvendig for å konstruere sammenlignbare tiltaks- og kontrollgrupper.

- **Kompetanse:** Solide kvasiexperimentelle studier krever god metodekompetanse, innsikt i sektorene der virkemidlene ble iverksatt og kompetanse på avansert håndtering av datakilder. De som gjennomfører slike studier bør ha god kontakt med personer i tjenestene og sikre vitenskapelig kvalitet gjennom deltakelse i internasjonale nettverk og publisering i fagtidsskrift.

Boks 1. Eksempel på aktiv tilrettelegging for kvasiexperimentell evaluering¹

Stovner bydel i Oslo kommune ble særlig hardt rammet under koronapandemien, og en valgte der å iverksette en rekke tiltak bl.a. for å øke oppslutningen om testing. Ett av tiltakene var dør-til-dør-aksjoner og mobile testbussar. For å måle effekten av tiltaket ville en sammenlikne testhyppighet i områder hvor dør-til-dør-aksjoner ble iverksatt, med områder der dette ikke ble gjort. Data på forekomst av testing kunne hentes ut fra Beredskapsregisteret, men ettersom dette kun inneholdt aidentifiserte data var det ikke mulig å måle forekomst av testing på annet enn grunnkrets nivå (den minste geografiske enheten som SSB benytter). For å legge til rette for evaluering sørget derfor bydelsadministrasjonen for at tiltaket ble iverksatt innen utvalgte grunnkretser.

¹Vinjerui KH, Elgersma IH, Fretheim A. Increased COVID-19 Testing Rates Following Combined Door-to-Door and Mobile Testing Facility Campaigns in Oslo, Norway, a Difference-in-Difference Analysis. *International Journal of Environmental Research and Public Health*. 2021; 18(21):11078. <https://doi.org/10.3390/ijerph182111078>

Randomiserte studier fordrer tett samarbeid mellom besluttende myndighet, aktørene som skal iverksette virkemiddelet og forskere som skal gjennomføre evalueringen. Noen virkemidler egner seg i mindre grad for evaluering med RCTer grunnet juridiske, etiske eller praktiske utfordringer samt ressurs hensyn. Flere etater viser til at lovverket sjelden tillater unnlattelse av rettighetsfestede tjenester til kontrollgrupper, selv om gjeldende kunnskapsgrunnlag for disse tjenestene tilsier behov for grundige effektevalueringer. Etatene understreker at det til sist er en helhetsvurdering, der juridiske, etiske og praktiske utfordringer samt ressurs hensyn vurderes mot samfunnsnytt, som avgjør hvorvidt et virkemiddel er egnet for evaluering i en randomisert eller en kvasiexperimentell studie – eller både og.

Kunnskap om effektene av et tiltak er en forutsetning for å kunne gjennomføre kost-nytte vurderinger, som igjen er viktig for prioritering av virkemidler. RCTer og kvasiexperimentelle metoder bør anses som supplerende tilnærminger som kan understøtte kunnskapsbaserte beslutninger om virkemiddelbruk.

Boks 2. Eksempel Skatteetaten: Måle effekter av kontroll ved å utnytte regler fra maskinlæringsmodell¹

Skatteetaten benytter fradragsmodellen (maskinlæringsmodell) til å prioritere kontrollutplukk på føring av fradrag i skattemeldingen. Fradragsmodellen gir en risikoscore til alle skattepliktige som øker fradragene sine, og alle med risikoscore over en gitt scoregrense tas ut til kontroll. Skattepliktige som ligger lokalt rundt denne scoregrensen for kontroll (over og under) vil være tilnærmet like, slik at det også er tilnærmet tilfeldig hvem av disse som tas ut til kontroll. Dette har Skatteetaten utnyttet i flere analyser for å måle etterlevelseseffekten av kontroll gjennom bruk av metoden Regression Discontinuity Design (RD-design). Analysene viser at det er tydelige positive etterlevelseseffekter av kontroll, det vil si skattepliktige med kontroll fører færre uriktige fradrag i årene etter kontrollen, sammenliknet med skattepliktige uten kontroll.

¹Karto C, Øvrum A, Løyland K, Andersen K. Effekter av kontroll blant ulike grupper av skattepliktige. I *Skatteetatens Analysenytt* 2/2019.

3.2 Randomiserte og kvasiexperimentelle studier: metodiske styrker og svakheter

Etatene understreker at den store styrken ved RCTer er at den tilfeldige tilordningen av et virkemiddel til tiltaks- og kontrollgrupper minimerer systematiske forskjeller mellom gruppene. Dermed kan en f.eks. fastslå at forskjellen i skoleprestasjoner mellom barn som fordeles til å få, eller ikke få smågruppeundervisning skyldes undervisningsformen, og ikke andre forskjeller mellom barna i de to gruppene.⁵

RCTer kan gi resultater som vanskelig lar seg overføre til populasjoner som ikke er representert i studieutvalget, eller til andre miljøer. Deltakerne som fordeles til tiltaks- og kontrollgruppene er gjerne valgt ut basert på strenge inklusjonskriterier, og vil ofte mangle tilstrekkelig representasjon fra alle deler av befolkningen.⁶ I RCTer er virkemiddelet som evalueres ofte iverksatt under optimale betingelser, mens iverksetting i «det virkelige liv» kan være ganske annerledes. Det er også en svakhet ved mange RCTer at de går over en kort periode og dermed ikke fanger opp langtidseffekter, som endringer i helseutfall eller deltakelse i arbeidslivet. En strategi for å møte noen av disse utfordringene er å gjennomføre «pragmatiske» RCTer.⁷ Mange av studiene som er identifisert i denne kartleggingen kan betegnes som pragmatiske studier, dvs. at virkemiddelet er evaluert under så normale betingelser som mulig.

Randomisering av grupper av mennesker (klyngerandomisering) gjøres for å måle virkningene av tiltak som ikke er rettet mot enkeltpersoner. En kan f.eks. iverksette tiltak på noen utvalgte skoler, og ikke på andre. Det vil da være fare for at tiltaket også har en effekt på skolene som ikke mottar tiltaket («smitteeffekt»). Et annet eksempel på denne utfordringen kommer fra NAV: veiledere i NAV er vant til å dele erfaringer og praksis på tvers av NAV-kontorer. Under RCTen av helhetlig oppfølging av lavinntektsfamilier forelå

⁵Bonesrønning, Hans, et al. "Small-group instruction to improve student performance in mathematics in early grades: Results from a randomized field experiment." *Journal of Public Economics* 216 (2022): 104765.

⁶ Dette gjelder spesielt for individrettede RCTer.

⁷ Stensrud, Mats Julius. "La oss være pragmatiske." *Tidsskrift for Den norske legeforening* (2019). <https://tidsskriftet.no/2019/05/medisin-og-tall/la-oss-vaere-pragmatiske>

det risiko for at praksisen som ble testet ble spredt til kontrollkontorene.⁸ Smitteeffekter kan på tilsvarende måte forkludre resultater i kvasieksperimentelle studier.

Ettersom deltakerne i en kvasieksperimentell studie vanligvis mottar det aktuelle tiltaket av andre årsaker enn deltakelse i et forskningsprosjekt, er det mindre fare for at deltakerne utgjør et skeivt utvalg av målgruppa for tiltaket. I mange kvasieksperimentelle studier er det også mulig å følge deltakerne over lang tid ved hjelp av nasjonale registre. Den viktigste metodiske svakheten ved kvasieksperimentelle metoder er at det alltid vil hefte større eller mindre grad av usikkerhet om hvorvidt resultatene er påvirket av systematiske skeivheter (bias) pga. andre forskjeller mellom gruppene enn det aktuelle tiltaket.

Både RCTer og kvasieksperimentelle metoder fordrer høy presisjon på spørsmålet som en ønsker svar på, grundig forberedelse og konkretisering av antagelsene om hvorfor virkemiddelet forventes å ha en effekt, og tydelig definering av utfallsmålene. For RCTer er korrekt implementering av virkemiddelet som evalueres avgjørende for å motvirke implementeringsfeil som kan medføre at det feilaktig konkluderes at tiltaket er uten virkning. Etatene fremhever betydningen av å investere tid og kompetanse i solide implementeringsstudier før RCTer gjennomføres. F.eks., i forbindelse med en RCT av effekten av videreutdanningsstipend for fagarbeidere la Kunnskapsdepartementet opp til at studien skulle inkludere en modellutviklingsfase og en piloteringsfase før gjennomføringen av RCTen. Dette for å avklare om informasjonen om stipendet faktisk nådde målgruppen, undersøke hvorvidt systemene for å søke stipendet fungerte som normalt og informere studiens styrkeberegning. Slike implementeringsstudier har også nytteverdi for politikktutforming fordi det kan avdekke logiske brister i utformingen av virkemiddelet før det iverksettes.

3.3 Organisering, gjennomføring og formidling av effektevalueringer

Noen av forvaltningsorganene utformer og gjennomfører selv slike evalueringer ved hjelp av interne ressurser (f.eks. FHI, NAV, og Skatteetaten), mens andre hovedsakelig anskaffer ekstern ekspertise til å gjennomføre studiene (Bufdir, Udir, og Norad). I begge tilfeller er det ofte et utstrakt samarbeid mellom forvaltningen og forskningsmiljøer fra universitets-, høyskole- og instituttsektoren. Utfyllende informasjon om etatenes erfaringer med RCTer og kvasieksperimentelle studier er sammenfattet i vedlegg til denne rapporten (Vedlegg 2–6).

I tilfeller der studiene er anskaffet, har kunnskapsbehovet ofte blitt definert i tett dialog med departementet som er ansvarlig for sektoren. F.eks. anskaffer Udir forskning hovedsakelig på oppdrag fra Kunnskapsdepartementet. Videre utvikles opplæringssektorens kunnskapsgrunnlag gjennom investering i langsiktige programmer i Forskningsrådet, f.eks. LÆREEFFEKT som eksplisitt prioriterer randomiserte studier.⁹

Kompetansen som blir etterlyst for å gjennomføre flere effektevalueringer varierer mellom forvaltningsorganene, men to områder peker seg ut. For det første er det behov for mer kunnskap om forskningsmetodikk, både innen randomiserte og kvasieksperimentelle

⁸Helhetlig oppfølging av lavinntektsfamilier (HOLF-prosjektet). <https://www.oslomet.no/forskning/forskningsprosjekter/holf-prosjektet>

⁹LÆREEFFEKT – Lærertetthet og læringseffekt. https://www.forskingsradet.no/contentassets/965669a13e7c42e98bb879bdc2ddbdb/rammedokument_for_laereeffekt.pdf

tilnærminger. For det andre er mer juridisk kompetanse nødvendig for å forstå gjeldende regelverk, avklare rettslige rammer og utnytte handlingsrommet som regelverket gir.

Effektevalueringene blir som oftest publisert som rapporter på etatenes nettsider eller som artikler i fagfelleverderte tidsskrifter. F.eks. er effektstudier gjennomført på oppdrag av Udir søkbare og tilgjengelige på deres nettsider. Noen av etatene gjør rapporter tilgjengelig i Kudos–Kunnskapsdokumenter i offentlig sektor, som er en dokumentksamling og søkeløsning for offentlige kunnskaps- og styringsdokumenter (<https://kudos.dfo.no/>). Noen forvaltningsorganer, som Bufdir, har et nettbibliotek som er en søkbar database der alle bestilte effektevalueringer gjøres tilgjengelig.¹⁰ Skatteetaten publiserer et digitalt tidsskrift – Skatteetatens analysenytt – med aktuelle rapporter og analyser fra etatens analyseteam. Dette inkluderer også lenker til publikasjoner og formidling fra studiene, f.eks. protokoller, artikler og rapporter. Videre blir alle norske forskningspublikasjoner, inkludert forskningsprosjekter, registrert i CRISTIN-databasen som er søkbar (<https://www.cristin.no/om-cristin/>). De fleste forvaltningsorganene har som krav at RCTer registreres i internasjonale databaser før studiens oppstart.

3.4 Omfang og eksempler på evaluerte virkemidler

3.4.1 Randomiserte studier

De fleste forvaltningsorganene som har bidratt til kartleggingen har erfaringer med å gjennomføre eller anskaffe randomiserte studier. Hvor mange studier etatene anslo å ha gjennomført i løpet av de siste fem årene varierte fra 1-5 studier til over 15.

Forvaltningsorganene som har gjennomført over 15 er de som stort sett utformer og gjennomfører studiene selv.

Konkrete eksempler på gjennomførte eller pågående RCTer av virkemidler, grovt inndelt etter kategoriene presentert under kapittel 2, inkluderer:

- **Organisatoriske virkemidler:** (1) Vinmonopolets utvidede åpningstider (gjennomføres av FHI); (2) Et lag rundt eleven (anskaffet av Udir)
- **Pedagogiske virkemidler:** (1) Raskt i jobb for flyktninger (anskaffet av NAV); (2) Dulting i skattemeldingen (gjennomført av Skatteetaten, Boks 3); (3) Nurse family partnership (Sammen på vei) (pågående, anskaffet av Bufdir);
- **Økonomiske virkemidler:** (1) Resultatbasert finansiering av helsetjenester (flere studier finansiert av Norad)
- **Offentlig myndighetsutøvelse:** (1) Effekter av meldingskontroll blant personlig skattepliktige og effekter av meldingskontroll av nyregistrerte MVA-virksomheter (gjennomført av Skatteetaten)

¹⁰https://www2.bufdir.no/Bibliotek/?_gl=1*9gpm74*_ga*ODg5ODg5MDAxLjE2ODAwODQzMDg.*_ga_E0HBE1SMJD*MTY4MTQ1NTEyOS42LjAuMTY4MTQ1NTEzMy4wLjAuMA.

Boks 3. Eksempel fra Skatteetaten: Dulting i skattemeldingen hjelper folk til å rapportere riktig¹

Den nye skattemeldingen har et digitalt og brukervennlig grensesnitt som inneholder funksjonalitet for å veilede skattepliktige i sanntid. Det betyr at merknader kommer opp med beskjeder tilpasset den konkrete skattepliktiges forhold. Skatteetaten omtaler ofte denne veiledningen som «dulter». Skattemeldingen inneholder nå flere hundre ulike dulter. I 2022 har det blitt gjennomført A/B-testing (randomisert tilordning) av 17 av disse dultene. Effektmålinger viser at dultene fører til bedre etterlevelse ved at skattepliktige foretar endringer i egenfastsettingen av alminnelig inntekt. Bruttoeffekten er på ca. 3,1 milliarder kroner, mens nettoeffekten er på ca. 1,65 milliarder kroner. Rundt 300 000 personer leverte en mer korrekt og fullstendig skattemelding i 2022 som følge av disse 17 dultene.

¹Brevik A-L, Habbestad A, Nilsskog M. Dulting i skattemeldingen hjelper folk til å rapportere riktig. I Skatteetatens *Analysenytt* 2/2021.

3.4.2 Kvasiekperimentelle studier

Forvaltningsorganene som stort sett gjennomfører studiene selv rapporterer å ha gjennomført eller anskaffet over 15 kvasiekperimentelle studier, mens de som i hovedsak anskaffer studier i mindre grad anskaffer kvasiekperimentelle effektstudier.

Eksempler på kvasiekperimentelle studier beskrevet av etatene i forbindelse med kartleggingen er:

- **Organisatoriske virkemidler:** (1) Effekter av "Norway-India Partnership Initiative" på mødre- og barnehelse (finansiert av Norad); (2) Effekter av norm for lærertetthet (anskaffet av Udir)
- **Pedagogiske virkemidler:** (1) Effekter av NAV-veileder i videregående skoler (anskaffet av NAV)
- **Økonomiske virkemidler:** (1) Effekter av gratis skolefrukt på skoleprestasjoner (gjennomført av FHI); (2) Effekter av gratis kjernetid tid i barnehagen i Oslo (anskaffet av Udir)
- **Offentlig myndighetsutøvelse:** (1) Effekter av standardiserte tobakkspakninger på tobakksbruk (gjennomført av FHI); (2) Effekter av kortere maksimal varighet på AAP på overgang til arbeid og uføretrygd (anskaffet av NAV); (3) Effekter av ordinære, risikobaserte kontroller av skattemeldingen og effekter av kontroll og tillegg (gjennomført av Skatteetaten)

Kartleggingen har ikke hatt som ambisjon å gi et fullstendig bilde av bruken av RCTer og kvasiekperimentelle studier i den offentlige forvaltningen, men supplerende erfaringer og perspektiver ble innhentet fra flere statlige forvaltningsorganer, samt på kommunalt nivå (Boks 4).

3.5 Vitenskapelig kvalitet

Etatene tilstreber høy vitenskapelig kvalitet på effektevalueringer ved å bruke en eller flere av følgende strategier:

- Vurdering av vitenskapelig kvalitet av interne fagfolk samt eksterne ressurser med relevant forsker- og metodekompetanse
- Forskningsprotokoller som er utviklet, forhåndsregistrert og rapportert i tråd med internasjonale retningslinjer, bl.a. med pre-spesifisering av analyser
- Intern kompetanseheving
- Bygging av kultur med vekt på høy vitenskapelig standard
- Publisering i internasjonale tidsskrifter

Boks 4. Randomiserte og kvasiexperimentelle studier i andre deler av forvaltningen

Formålet med denne rapporten er å kartlegge erfaringer med bruk av randomiserte forsøk og kvasiexperimentelle metoder i deler av norsk forvaltning. Vi har ikke tatt mål av oss å gi et fullstendig bilde av i hvor stor grad slike metoder brukes, og vi har heller ikke forsøkt å dekke alle offentlige etater.

Helsedirektoratet (Hdir) og SSB bidro til kartleggingen gjennom korte muntlige samtaler og skriftlige innspill. Hdir rapporterte å ha vurdert muligheten for randomiserte studier, men at finansiering, tidsaspekter og virkemiddelets egnethet for randomisering ikke har gjort slike studier mulig til nå. Imidlertid rapporterer Hdir å ha gjennomført noen få kvasiexperimentelle studier med «før-etter» design. SSB bidrar til flere RCTer ledet av andre og leder også egne studier, som f.eks. en RCT om effekten av tilrettelagt matematikkundervisning for elever med svake resultater fra nasjonale prøver.³

Videre kjenner vi til flere eksempler fra andre deler av offentlig forvaltning enn de vi har med i denne rapporten, f.eks.:

- Vegdirektoratet har vært med på å finansiere et randomisert forsøk av å kurse yrkessjåfører i økonomisk kjøring, utført av Transportøkonomisk Institutt.¹
- Integrerings- og mangfoldsdirektoratet (IMDi) har finansiert et mindre randomisert forsøk av norskopplæring med flerspråklig støtte i regi av Arbeidsforskningsinstituttet (OsloMet) som et pilotprosjekt for å vurdere gjennomførbarhet av et tilsvarende forsøk i større skala.²
- Direktoratet for høyere utdanning og kompetanse (HK-dir) har, på oppdrag fra Kunnskapsdepartementet, anskaffet en RCT om effekten av å tilby et videreutdanningsstipend til fagarbeidere. Lånekassen og Statistisk Sentralbyrå (SSB) er også sentrale bidragsytere til denne studien. Denne studien er nærmere omtalt i boks 7.

IMDi har også utarbeidet en rapport om bruk av randomiserte forsøk og kvasiexperimentelle metoder med tanke på å «lære av disse erfaringene for å kunne tilrettelegge for flere utprøvinger av integreringstiltak».⁴

Vi er ikke kjent med at norske kommuner har tatt initiativ til randomiserte forsøk eller kvasiexperimentelle studier. Men det er flere eksempler på at slike metoder er brukt i forskerinitierte prosjekter som gjennomføres i norske kommuner, f.eks. IMPAKT-prosjektet i Bergen som omfatter et randomisert forsøk for å evaluere effekten av tiltak for kvalitetsforbedring i sykehjem.⁵

¹DR Pinchasik, IB Hovi, E Bø, CS Mjøsund. [Can active follow-ups and carrots make eco-driving stick? Findings from a controlled experiment among truck drivers in Norway](#). Energy Research & Social Science, Volume 75, 2021.

²<https://oda.oslomet.no/oda-xmlui/handle/11250/3054795>

³<https://www.ssb.no/en/forskning/discussion-papers/didactic-methods-and-small-group-instruction-for-low-performing-adolescents-in-mathematics-results-from-a-randomized-controlled-trial>

⁴<https://www.imdi.no/arkiv/arkiverte-publikasjoner/effektevalueringer-pa-integreringsfeltet/>

⁵<http://www.impakt.no/impakt-i-sykehjem/intervensjonen/index.html>

3.6 utfordringer ved gjennomføring av RCTer og kvasiekperimentelle studier

Etatene rapporterer mange av de samme utfordringene for gjennomføring av RCTer og kvasiekperimentelle studier, som utfordringer med datagrunnlag eller politisk aksept for gradvis utrulling og forskjellsbehandling når virkemidler iverksettes. I noen tilfeller kan det være spesifikke utfordringer knyttet til en av metodene. Det gjelder spesielt gjennomføring av RCTer, som ofte støter på juridiske, etiske og praktiske utfordringer samt ressurs hensyn. Som en konsekvens av dette er kvasiekperimentelle studier ofte mer gjennomførbare, noe som understreker viktigheten av å betrakte disse metodene som supplerende tilnærminger for å dekke ulike kunnskapsbehov.

3.6.1 Tilstrekkelig antall deltakere

For mange randomiserte studier er oppnåelse av tilstrekkelig statistisk styrke en vesentlig utfordring. Studier av virkemidler på grupper av individer (f.eks. NAV-kontorer, skoler, helsestasjoner eller kommuner) gjør det utfordrende å rekruttere tilstrekkelig antall grupper å randomisere. F.eks., da FHI foreslo å evaluere virkningene av å gjenåpne skoler i koronapandemiens første fase, ble det anslått at alle skoler i landet måtte inngå i randomiseringen for å få store nok tall til å kunne trekke pålitelige konklusjoner. Tilsvarende beskriver Bufdir at målgruppen for et tiltak, f.eks. enkelte brukergrupper av barnevernstjenester, er så begrenset at det kan være utfordrende å rekruttere mange nok deltakere til å oppnå tilstrekkelig statistisk styrke.

3.6.2 Datagrunnlag

Registerdata og spørreundersøkelser er de vanligst brukte datakildene i RCTer og kvasiekperimentelle studier. Store befolkningsundersøkelser, som Den norske mor, far og barn-undersøkelsen (*MoBa*), representerer datainfrastruktur som kan tilrettelegge for gjennomføring av RCTer og kvasiekperimentelle studier (Boks 5). Tilgang til registerdata og kobling av registre er imidlertid fremtredende utfordringer. F.eks. krever solide kvasiekperimentelle studier detaljerte data og omfattende datakoblinger. Mange slike studier forhindres av at det er tidkrevende og vanskelig å få tilgang til data, ofte på grunn av personvern hensyn og vurderinger av om hvorvidt data kan brukes uten samtykke. Når spørreundersøkelser blir benyttet er oppnåelse av tilstrekkelig høy svarprosent en stor utfordring.

Fravær av sammenlignbare og presise mål for effekt er en annen utfordring for datagrunnlaget. F. eks. innen opplæringssektoren utgjør manglende hyppighet og kontinuitet i målingene, at eksisterende datakilder bare måler enkelte deler av skolens arbeid, og at de ikke er utviklet for forskning sentrale utfordringer. Eksempelvis er Elevundersøkelsen og nasjonale prøver utformet for andre formål enn forskning og blir kritisert for å ikke å være sensitive nok til å fungere som gode måleinstrumenter for læringseffekter, læringstap, endringer i læringsmiljø eller psykisk helse i effektstudier. Det foreligger også mangel på individdata. I NOU 2019: 3, «Nye sjanser – bedre læring – Kjønnsforskjeller i skoleprestasjoner og utdanningsløp», ble det fremmet flere forslag for å bøte dette, bl.a. ved å opprette et nasjonalt register for personentydige forløpsdata for barnehage og grunnopplæring.¹¹

¹¹<https://www.regjeringen.no/no/dokumenter/nou-2019-3/id2627718/?ch=2>

Boks 5. Befolkningsundersøkelser: viktig datainfrastruktur for effektevalueringer

Store eksisterende befolkningsundersøkelser kan brukes til RCT-er for å besvare viktige spørsmål om befolkningsrettede informasjonstiltak, noe som kan redusere tidsbruken og kostnadene forbundet med datainnsamling og rekruttering. F.eks. vil totalt 20 000 menn og kvinner i Danish National Birth Cohort (DNBC) og MoBa i alderen 18 år eller eldre randomiseres til å motta eller ikke motta informasjon om alders- og livsstilsrelatert nedgang i fekunditet (evnen til å få barn) i en studie ledet av FHI. På forhånd vil FHI kartlegge hvordan deltakernes oppfatninger om fruktbarhet, både på populasjonsnivå og sin egen individuelle fruktbarhet. Hypotesen er at folk i gjennomsnitt er for optimistiske når det gjelder fruktbarhet, og undervurderer hvordan fruktbarheten faller. Deretter randomiseres deltakerne til å enten motta informasjon om ulike aspekter ved fekunditet og risikofaktorer for fallende fekunditet i form av et webinar, eller til å delta i en kontrollgruppe som ikke mottar informasjon. I etterkant vil det undersøkes om deltakerne husker informasjonen, og om den endrer kunnskapen deres om fekunditet. Forskerne vil deretter se på om denne informasjonen påvirker intensjoner om å få barn, og om hvorvidt deltakerne faktisk får barn, ved å følge deltakerne over tid i de to kohortene.

3.6.3 Juridiske utfordringer

En sentral juridisk utfordring for befolkningsrettede RCTer er kravet om informert samtykke, som gjør at mange studier er praktisk umulige å gjennomføre. For studier som faller under helseforskningsloven har det ufravikelige samtykkekravet vært et avgjørende hinder for gjennomføring av RCTer av smitteverntiltak (Boks 6). Det er mindre klart om og når samtykke er påkrevd for studier som er underlagt annen lovgivning og andre forskningsetiske retningslinjer. Flere etater rapporterer at samtykkekravet kommer an på virkemiddelet som evalueres, f.eks. hvor inngripende det er for deltakerne i studien.

Boks 6: Særskilte krav til «helseforskning»

Helseforskningsloven stiller en rekke krav til evalueringer som faller inn under lovens definisjon av «helseforskning». F.eks. regnes utprøving av luftrensere i klasserom som helseforskning dersom formålet er å redusere smittespredning. Dermed utløses bl.a. krav om samtykke fra alle elever/foresatte, både i skoler som randomiseres til å ha, og til å ikke ha luftrensere. Dersom luftrensere i stedet utplasseres for å undersøke om de fører til bedre læringsresultater for elevene, vil det neppe regnes som helseforskning, og dermed forsvinner det juridiske kravet om samtykke. Dersom rektorer installerer luftrensere på sin skole for å redusere smittespredning, uten å undersøke virkningen av luftrensingen, regnes det heller ikke som helseforskning, og dermed utløses heller ikke noe krav om samtykke fra elever/foresatte.

Gjennomføring av en randomisert studie, evt. tilrettelegging for en kvasieksperimentell studie, kan komme i konflikt med prinsippet om likebehandling og lovbestemmelser som beskytter enkeltpersoner grunnleggende rettigheter og tjenester. F.eks., i henhold til folketrygdloven kan ikke effektevalueringer av NAVs ytelse innebære innskrenkning av enkeltpersoners rettigheter eller påleggelse av større plikter enn det som følger av loven. I praksis betyr dette at personer som oppfyller kravene for å motta ytelse ikke kan forskjellsbehandles gjennom inndeling i tiltaks- og kontrollgrupper i en effektevaluering.

Et annet eksempel gjelder opplæringsloven med forskrifter, som inneholder en rekke minimumskrav, inkludert at alle elever skal få et likeverdig tilbud og et tilfredsstillende utbytte av opplæring i et trygt og godt skolemiljø. Dersom gjennomføringen av en effektstudie forutsetter avvik fra opplæringsloven med forskrifter, åpner loven for at Utdanningsdirektoratet godkjenner forsøk. Det er ikke åpnet for at det gjennomføres forsøk som berører elevenes mest grunnleggende rettigheter eller som medfører urimelige byrder for de deltagende elevene. Det tillates kun forsøk som med rimelig grunn antas å kunne føre til en forbedret opplæring for elevene totalt sett.

Under den tidlige fasen av pandemien foreslo FHI en randomisert gjenåpning av skoler for å kunne evaluere effekten på covid-19 smitte.¹² Barns lovfestede rett til skolegang i Norge og at myndighetene allerede hadde besluttet at skoler skulle gjenåpnes gitt den lave smitterisikoen ble av den Nasjonale forskningsetiske komité for medisin og helsefag (NEM) ført som argumenter mot gjennomføring av en randomisert studie.¹³ Det er imidlertid uklart hvorvidt denne tidsbegrensede innskrenkningen av rettigheter ble vurdert opp mot nytten av å få mer kunnskap om effekten av skolestengning som et smitteverntiltak, og potensielt unngå unødige skolestengninger senere.

Skatteetaten har erfart at loven legger klare begrensninger for evaluering av sanksjonering, et virkemiddel for å motivere riktig rapportering av skatt. Her forhindrer skatteforvaltningslovens regler om ileggelse av tilleggsskatt, som innebærer at alle som oppfyller kriteriene for tilleggsskatt *skal* ilegges slik skatt, en randomisert evaluering av ulike former for sanksjonering. Til sammenlikning har Skatteetaten erfart at randomisering av mykere forebyggende virkemidler som informasjonsbrev, SMS, telefonoppringning og liknende er enklere å få tillatelse til å gjennomføre. Imidlertid kan det også her oppstå konflikter med skatteforvaltningslovgivningen, siden innhenting av opplysninger om skattepliktige som trekkes tilfeldig til tiltaksgruppen kan utløse krav om kontroll, noe som vil gjøre at hensikten med mykere virkemidler faller bort. I dette tilfellet har det imidlertid blitt avklart innenfor skatteforvaltningslovgivningen at opplysninger kan innhentes, bl.a. til virkemidler som dulting, uten at de samme skattepliktige senere må kontrolleres.

Et annet eksempel der klargjøringen av det rettslige grunnlaget sto sentralt er en pågående randomisert studie av effekten av videreutdanningsstipend for fagarbeidere, initiert av Direktoratet for høyere utdanning og kompetanse (HK-dir) på oppdrag fra Kunnskapsdepartementet. Her ble det konkludert at en forskjellsbehandling mellom tiltaks- og kontrollgrupper ikke anses som en lovstridig innskrenkning av rettigheter; snarere ble det sannsynliggjort at utprøvingen vil være i tråd med utdanningsstøttelovens formål om like muligheter til utdanning for alle (Boks 7).

¹²School opening in Norway during the COVID-19 pandemic. <https://www.isrctn.com/ISRCTN44152751>

¹³Prosjektåpning under pandemien - evaluering (SÅPE). Saksnr. NEM 2020/106 REK Sør-øst A 136575 | [Klagesak]. <https://www.forskningsetikk.no/om-oss/komiteer-og-utvalg/nem/sok-i-nems-vedtak/prosjektapning-under-pandemien---evaluering-sape/>

Boks 7. Randomisert utprøving med videreutdanningsstipend for fagarbeidere: rettsgrunnlag

Økonomiske incentiver for å styrke rammebetingelsene for videreutdanning og kompetanseutvikling står sentralt i kompetansereformen.¹ For å få mer kunnskap om effektene av slike virkemidler ga Kunnskapsdepartementet HK-dir i oppdrag å gjennomføre en randomisert utprøving av effekten av videreutdanningsstipend for fagarbeideres deltagelse i videreutdanning og tilknytningen til arbeidsmarkedet. Forskningsmiljøene Fafo og Frisch-senteret bistår studien, Lånekassens utbetalingsløsning brukes for å fordele stipendene, og SSB er ansvarlig for å trekke et tilfeldig utvalg som tilbys stipendet.

Studien ble hjemlet i «Ny forskrift om midlertidige regler for utdanningsstøtte som følge av en ny forsøksordning med et eget stipend for fagarbeidere».² I forbindelse med utarbeidelsen av forskriften ble det reist spørsmål om forskriften og den ledsagende RCTen er i strid med utdanningsstøtteloven og annet rettsgrunnlag. RCTen innebærer at en begrenset gruppe fagarbeidere blir valgt ut ved loddtrekning og tilbys stipendet uten behovsprøving. Det var derfor nødvendig med en juridisk avklaring for å fastslå om RCTen innebar en innskrenkning av rettighetene til fagarbeiderne som ikke blir trukket ut, og om dette var i strid med formålet til utdanningsstøtteloven om like muligheter til utdanning for alle.

En uavhengig juridisk vurdering la tre punkter til grunn for å begrunne hvorfor forskriften ikke innebærer en lovstridig innskrenkning av rettigheter:³

- **Varigheten:** RCTen innebærer en midlertidig og forholdsvis kortvarig (to studieår) forskjellsbehandling
- **Omfanget:** RCTen omfatter et begrenset antall personer, og nesten alle andre fagarbeidere har fortsatt like muligheter til å få lån og stipend
- **Formål:** studiens mål om å fremskaffe pålitelig kunnskap om hvorvidt stipendordningen bør innføres permanent anses som et bidrag til å sikre like muligheter til utdanning, selv om RCTen midlertidig er i konflikt med dette formålet

¹<https://www.regjeringen.no/no/dokumenter/meld.-st.-14-20192020/id2698284/>

²<https://lovdata.no/dokument/LTI/forskrift/2021-02-22-511>

³Juridisk vurdering fra Vaar Advokat AS. Anskaffet av HK-dir (tidligere Kompetanse Norge)

Noen juridiske utfordringer er sektorspesifikke. Norad sin forvaltning av tilskudd reguleres av ordningsregelverk som setter rammene for hva midlene kan gå til. Effektevalueringer er ikke nevnt spesifikt, noe som har gjort det vanskelig å finansiere RCTer med slike midler. Imidlertid foreligger det nå en overordnet føring i statsbudsjettet (Prop. 1 S) om at tilskuddsmidler kan brukes til følgeforskning og effektevaluering, noe som forventes å skape et større handlingsrom for å finansiere effektevalueringer.¹⁴

Forvaltningsorganene beskriver at deres forskning kan være underlagt flere lovverk med ulike føringer og praksis for gjennomføring av RCTer. Flere av forvaltningsorganene, som Skatteetaten og NAV, har hatt eller har gjennomgang av det rettslige handlingsrommet for å gjennomføre effektevalueringer, inkludert vurdering av rettsgrunnlaget for tidsavgrenset forskjellsbehandling og behandling av personopplysninger under forsøk.

¹⁴[Prop. 1 S \(2022–2023\) - regjeringen.no](https://www.regjeringen.no/no/prop/1-s/2022-2023)

3.6.4 Etske hensyn

Kartleggingen peker på fire typer av etiske hensyn som bør vurderes når randomiserte studier planlegges og gjennomføres. Siden det sjeldent aktivt planlegges for kvasiexperimentelle studier gjelder disse betraktningene i mindre grad for slike studier. Samlet viser disse etiske hensynene behovet for konsistente forskningsetiske vurderinger av randomiserte studier, på tvers av sektorer.

Det første hensynet er at det kan være uforsvarlig å eksperimentere med et tiltak som forventes å ha positive virkninger for enkeltindivider og grupper. Det etisk «riktige» valget anses derfor å la alle dra nytte av tiltaket umiddelbart. F.eks. er en innvending mot RCTer i utdanningssektoren at noen elever får en fordel som ikke er tilgjengelig for kontrollgruppen. Et annet eksempel er fra internasjonal bistand, hvor det er argumentert for at det er etisk betenkelig å ha eksperimentelle kontrollgrupper der et «nyttig» tiltak holdes tilbake fra mennesker i fattigdom.

Det andre hensynet er at randomiserte studier er vanskelig å rettferdiggjøre dersom de medfører negative konsekvenser for de som er med i studien. F.eks. har dette argumentet blitt ført når RCTer medfører vesentlige forskjeller i skolehverdagen mellom barn som fordeles til tiltaks- og kontrollgruppene. Ett eksempel er den randomiserte studien «*Small group Instruction in Mathematics for Pupils Level 1-4: Effects of a randomized controlled trial intervention study*», som ble finansiert over LÆREEFFEKT-programmet.¹⁵ Den nasjonale forskningsetiske komité for samfunnsvitenskap og humaniora (NESH) belyste flere utfordringer som bør hensyntas i forbindelse med gjennomføringen av denne type studier, bl.a. behovet for barnefaglig kompetanse i forskerteamet og beredskap for å håndtere eventuelle negative følger av deltakelse i prosjektet.¹⁶

Det tredje hensynet gjelder i hvilken grad informert samtykke skal være et krav ved effektevalueringer av befolkningsrettede tiltak. Et argument for informert samtykke er at det er etisk betenkelig om deltakere ikke vet om de er en del av en RCT eller om de eksponeres for et virkemiddel uten å ha reell mulighet til å takke nei til deltakelse. På en annen side vil et absolutt samtykkekrav gjøre det umulig å gjennomføre effektevaluering av virkemidler som iverksettes på gruppenivå. For helseforskning er samtykkekravet absolutt, med noen få spesifikke unntak, f.eks. «I kliniske nødssituasjoner der pasienten ikke er i stand til å avgi samtykke, og der det er umulig å innhente samtykke fra vedkommendes nærmeste pårørende».¹⁷ Flere RCTer av smitteverntiltak i regi av FHI har ikke latt seg gjennomføre ettersom det måtte innhentes samtykke fra alle individer som skulle delta – en praktisk umulighet i studier der deltakerne er grupper av mennesker (f.eks. alle innbyggere i en kommune). Det ufravikelige kravet om samtykke fra deltakerne gjelder kun for forskning som kvalifiserer som «helseforskning» i helseforskningslovens forstand, dvs. forskning der formålet er «å skaffe til veie ny kunnskap om helse og sykdom».¹⁷ For forskning med andre formål er det ikke et absolutt krav om individuelt samtykke, og det gjøres en vurdering for hver enkelt studie om hvorvidt samtykke er påkrevd.

¹⁵Bonesrønning, Hans, et al. "Small-group instruction to improve student performance in mathematics in early grades: Results from a randomized field experiment." *Journal of Public Economics* 216 (2022): 104765.

¹⁶<https://www.forskningsetikk.no/om-oss/komiteer-og-utvalg/nesh/uttalelser/uttalelse-om-forskningsprosjektet-small-group-instruction-in-mathematics-for-pupils-level-1-4-effects-of-a-randomized-controlled-trial-intervention-study-2018331>

¹⁷https://lovdata.no/dokument/NL/lov/2008-06-20-44/KAPITTEL_4#%C2%A716

Det fjerde hensynet gjelder hvordan forskning på barn og andre sårbare grupper etisk skal begrunnes. Den etiske utfordringen er å nøye vurdere nytten og skaden det kan medføre for dem. Etiske vurderinger må derfor ta hensyn til barnas sårbarhet, behov, kompetanse, rettigheter og situasjon. For forskning på barn i sårbare situasjoner beskriver Bufdir samtykkekrav fra begge foreldrene og uklare prosesser for meldeplikt og oppfølging av deltakere dersom det oppstår alvorlige situasjoner, som spesielt utfordrende. Veiledningen fra den regionale etiske komiteen for denne type søknader har i enkelte tilfeller blitt oppfattet som mangelfull. F.eks. opplevde forskerne som jobbet med Nurse Family Partnership-studien (Boks 8), som ble anskaffet av Bufdir, en tidkrevende behandlingsprosess med å få godkjenning fra regional etisk komité (REK). Dette omhandlet ivaretagelse av etiske hensyn, inkludert hvordan kravet om samtykke fra begge foreldrene skulle balanseres mot den gravides rett til oppfølging.

3.6.5 Politiske utfordringer

Hvordan og i hvilken skala og hastighet virkemidler blir iverksatt er i stor grad politisk bestemt. Det finnes derfor et stort politisk handlingsrom for å tilrettelegge for bedre effektevalueringer. Kartleggingen peker på tre sentrale politiske utfordringer.

Den første er behovet for økt forståelse for at virkemidler som oftest er forbundet med usikkerhet om effekten og systematisk utprøving i form av randomiserte eller kvasiexperimentelle studier er viktige grep for å redusere denne usikkerheten. Dette kan eksempelvis skje gjennom gradvis – om mulig randomisert – utrulling av virkemidler eller mindre piloter ledsaget av evaluering fremfor rask fullskala implementering. For å oppnå dette kreves større politisk aksept og handlingsrom for tidsavgrenset, systematisk og tilfeldig forskjellsbehandling for å kunne evaluere effekten av virkemidler.

Den andre utfordringen er manglende samsvar mellom beslutningstagernes ønske om å få raske svar og utvise handlingskraft og gjeldende tidsvinduer for å gjennomføre effektevalueringer. Optimale effektevalueringer innebærer tid til konkretisering av problemet som skal løses, sammenfatning av kunnskapsgrunnlag, utforming av tiltak som kan vurderes for effektevaluering, forankring i tjenestene, implementeringsstudier for å utforske antagelsene som tiltakets forventede effekt bygger på, og etiske godkjenninger og datatilgang. Utprøvingen må også vare lenge nok til at relevante effekter kan måles. Mangelen på sanntidsdata utgjør en spesielt stor utfordring for kvasiexperimentelle studier og hindrer dermed beslutningstakere i få tilgang til pålitelig kunnskap om effekt i tide. Selv om det finnes et potensial for å gjennomføre effektevalueringer raskere, f.eks. med raskere prosesser for etisk godkjenning og datatilgang, så er tålmodighet blant beslutningstakere viktig for at relevante forskningsfunn kan informere beslutningsgrunnlaget.

Den tredje er at politiske prosesser som regjeringssendringer eller budsjettforlik kan endre konteksten virkemiddelet skal testes i og medføre at kontrollgruppen blir eksponert for virkemiddelet som utprøves. F.eks. måtte en randomisert studie om effekten av økt lærertetthet som Stortinget hadde bevilget midler til (LÆREFFEKT) avslutte utprøvingen før tiden grunnet innføringen av en ny norm for lærertetthet, som

innebar en endring av klassestørrelsen og påvirket sammenligningen mellom tiltaks- og kontrollgrupper.¹⁸

3.6.6 Andre utfordringer

En felles utfordring for RCTer og kvasiek eksperimenter er at det kan være skepsis mot gjennomføring av effektevalueringer dersom dette innebærer en forskjellsbehandling mellom de som får og de som ikke får tiltaket. Skepsisen kan også bero på risikoen for at studiene medfører en uforholdsmessig ulempe for de som berøres, som utdanningspersonell, saksbehandlere eller helsepersonell. RCTen av programmet Nurse Family Partnership har møtt både entusiasme og skepsis, sistnevnte grunnet motstand mot å gjennomføre effektstudier der gravide i kontrollgrupper ikke får tilgang til programmet. Forankring i kommunene og tjenestene sto sentralt for å gjøre studien gjennomførbar (Boks 8).

Boks 8. Forankring i kommunene: randomisert kontrollert studie om effektene av "Sammen på vei"

Programmet «Sammen på vei», også kjent som Nurse Family Partnership (NFP), retter seg mot førstegangsgravide med sammensatte utfordringer for å styrke deres kunnskap og utøvelse av eget foreldreskap. Programmet består av hyppige hjemmebesøk og støtte av spesialopplært sykepleier fra tidlig i svangerskapet og gjennom de to første årene i barnets liv. Programmet har en rekke utfallsmål knyttet både til den gravide og barnet helse, voldutsatthet og samspillet mellom dem.

Programmet har vært pilotert i Norge mellom 2016-2020 med ledsagende følgestudie utført av OsloMet AFI. Forskningen viste positive resultater og ut fra dette besluttet myndighetene at programmet skulle styrkes og utvides fra og med 2021. Utprøvingen løper videre til 2027. Som en del av denne fasen er en RCT om effekten av Sammen på vei planlagt, godkjent av REK, og registrert i databasen ClinicalTrials.gov. Planen er å begynne rekrutteringen av gravide i juni 2023. RCT-planene har vært møtt med både entusiasme og noe skepsis.

Skepsisen skyldes motstand mot å gjennomføre effektstudier der gravide i kontrollgrupper ikke får tilgang til programmet. Forberedelsen krevde derfor tid og ressurser til å motivere henvisende instanser til å rekruttere sårbare gravide deltakere til en RCT, selv om halvparten av de rekrutterte vil havne i kontrollgruppen. Videre var det sentralt for forskere, Regionsenter for barn og unges psykiske helse og Bufdir å være tilstede i kommunene hvor RCTen skal foregå og ta tak i skepsisen om studiens nytteverdi. Bufdir vektla at i den norske konteksten er det ikke slik at den gravide som havner i kontrollgruppen står uten noe tilbud fordi i Norge er det gode pre- og postnatale tilbud. Videre har lisenskravet til NFP, som krever at det gjennomføres en RCT for å vurdere om programmet er egnet for videre implementering i de aktuelle landene, gjort det enklere å overbevise om at studien må gjennomføres som en RCT.

Videre møter RCTer noen spesifikke utfordringer som gjør metoden krevende å gjennomføre. Disse utfordringene er mindre relevante for kvasiek eksperimentelle studier, fordi implementeringen av virkemiddelet som oftest skjer, eller har skjedd, som del av normal drift og praksis.

¹⁸<https://forskning.no/skole-og-utdanning-ny/ny-laerernorm-skaper-trobbel-for-forskere/304863>

I komplekse organisasjoner og tjenester kan RCTer være utfordrende og påvirke vanlig drift. Personalet ute i tjenestene har begrenset kapasitet, som kan gjøre det vanskelig å legge til rette for å gjennomføre RCTer. Det kan være formålstjenlig å anvende seg av tilgjengelig infrastruktur som gjør en studie administrativt og praktisk gjennomførbart. Et konkret eksempel er hvordan Lånekassens infrastruktur for utbetalinger av stipender har blitt brukt til den randomiserte studien av effekten av videreutdanningsstipend for fagarbeidere (Boks 7). Ved at Lånekassen administrerte utbetalingene vil virkemiddelet i stor grad tilsvare en faktisk stipendordning som innføres dersom ordningen blir permanent.

Randomiserte studier anses som kostnadskrevede. RCTer tar ofte opp en stor andel av den totale bevilgningen som er tilgjengelig for forskningsoppdrag- eller programmer, men kostnaden ved å gjennomføre en RCT varierer mye. Studiene som er identifisert og beskrevet i kartleggingen har kostet anslagsvis mellom 1 og 35 millioner kroner. Størrelsen påvirkes av om det er store omkostninger ved iverksetting av tiltaket, og om disse kostnadene medregnes. Dersom tiltaket koster lite, f.eks. en informasjonskampanje i sosiale medier for å oppfordre publikum til å la seg teste for Covid-19, kan totalkostnaden ved studien være liten. Kostnader kan også være fordelt over flere budsjetter, som kan innebære at spesifikke budsjetter ikke dekker alle kostnadene som er nødvendige for å gjennomføre et prosjekt eller en aktivitet. F.eks. kan FoU-midler være begrenset til å dekke forskningsrelaterte kostnader, men ikke kostnader knyttet til implementering eller opplæring av de som skal levere tjenesten som forskes på. Til sammenlikning kan mange kvasiekperimentelle studier gjennomføres både raskt og relativt billig dersom det nødvendige datagrunnlaget er tilgjengelig.

Randomiseringsmotvilje i befolkningen kan gjøre det vanskelig å iverksette studier som innebærer forskjellsbehandling. Dette kan ha vært bidragende til at forslaget om å randomisere gjenåpningen av skoler i Norge våren 2020 ikke ble noe av.^{19,20}

¹⁹Regjeringens pressekonferanse 7. mai 2020

²⁰Det er vanskelig å anslå betydningen og omfanget av randomiseringsmotstand i befolkningen, men resultater fra noen spørreskjemaundersøkelser tyder på en generell motvilje mot randomisering: Selv i en situasjon der tiltak B vurderes som dårligere enn tiltak A, foretrekker mange at B tilbys til alle, framfor randomisering til A og B [<https://www.medrxiv.org/content/10.1101/2023.04.05.23288189v1.full>]. Resultater fra andre studier tyder på at befolkningen ikke nødvendigvis er så negativt innstilt [<https://www.pnas.org/doi/10.1073/pnas.2217551120>]. Undersøkelsene er stort sett gjennomført i USA, og overførbareheten til norske forhold er også uvisst.]

3.7 Muligheter for gjennomføring av RCTer og kvasiekperimentelle studier

3.7.1 Forankring og politisk aksept for tidsavgrenset forskjellsbehandling

Tett dialog mellom fagetatene, departementene og forskere fra universitets-, høyskole- og instituttsektoren for å konkretisere kunnskapsbehov står sentralt for å motivere effektevalueringer. F.eks. sto tett dialog og åpne prosesser ledet av Kunnskapsdepartementet og Udir sentralt når loddrekning ble benyttet for å tildele statlige tilskudd til utstyr i fag- og yrkesopplæringen (Boks 9).

Boks 9. RCT på effekten av statlig tilskudd til utstyr i fag- og yrkesopplæringen: prosess og aksept

I forbindelse med statlige tilskudd til utstyr i fag- og yrkesopplæringen anbefalte forskningsmiljøer randomisert innvilgelse av søknader for å evaluere effekter av tildelingen. Beslutningsprosessen involverte en arbeidsgruppe sammensatt av representanter fra organisasjonene i opplæringssektoren.

Prosessen ble nøye utformet for å ivareta flere hensyn. For det første ønsket man å sikre en lik mulighet for alle skoler og skoleeiere å søke midler. For det andre tok man hensyn til ulik skolestørrelse og behov ved å tillate skoler med større behov å sende inn flere søknader. For det tredje ble en randomisert tildeling av midler gjennomført for å oppfylle målet om effektforskning. For å øke aksepten for fremgangsmåten ble søknadsprosessen gjort så enkel som mulig.

Det er imidlertid viktig å merke seg at tilskuddsordningens forutsetninger og prosedyrer var vanskelig å kommunisere. Bl.a. ble prosessen av NRK fremstilt som «statlig utstyrslotteri» hvor spillereglene ikke hadde blitt forstått av alle som søkte.¹ Representanter fra sektoren, som hadde deltatt i arbeidsgruppa, protesterte også mot fremgangsmåten, etter at den var blitt vedtatt. Det vil være viktig å lære av denne erfaringen og kommunisere tydeligere om forutsetningene og prosedyrene for randomiserte tilskuddsordninger for å unngå misforståelser og kritikk.

¹<https://www.nrk.no/vestland/vidaregaande-skular-fekk-77-millionar-i-statsstotte-ved-loddtrekking-1.15815655>

Et annet eksempel på hvordan gode prosesser med departement og aktørene som iverksetter virkemiddelet kan tilrettelegge for en RCT er effektstudien om utvidelse av Vinmonopolets åpningstider (Boks 10).

Boks 10. Aksept for tilfeldig og gradvis implementering av Vinmonopolets utvidelse av åpningstid

Stortinget vedtok juni 2020 å tillate en times lengre åpningstid for Vinmonopolet på lørdager. Ett år i forveien hadde Helse- og omsorgsdepartementet (HOD) gitt FHI i oppdrag å evaluere effekten av utvidet åpningstid dersom Stortinget vedtok lovendringen. FHIs forskningsteam anbefalte en klyngerandomisert studie (klynge-RCT) som en foretrukket metode i denne settingen. Tett dialog, samt å ta bekymringer på alvor og å forankre metoden med aktørene, herunder Vinmonopolet og dets butikker, var viktige forutsetninger for at Vinmonopolet kunne akseptere det foreslåtte forskningsdesignet.¹

Vinmonopolets sentrale bekymringer var risikoen for misnøye blant kunder i kontrollområdene som ville få utvidet åpningstid på et senere tidspunkt, og bekymring for omdømme og omsetningen. FHIs forskningsteam hadde flere møter med Vinmonopolets arbeidsgruppe for å forklare mulige RCT-design samt forstå Vinmonopolets bekymringer. En klyngerandomisert RCT ble til slutt foreslått som en måte å møte disse bekymringene på. Vinmonopolet delte detaljerte salgsdata som ble brukt for styrkeberegning og informere andre deler av designet, herunder inndelingen av pol i klynger.

Behovet for å finne en løsning for RCT-studien ble akutt i april 2020, da det ble klart at lovforslaget sannsynligvis ville bli vedtatt og at utvidelsen av åpningstiden ville bli iverksatt kort tid etterpå. Under nye møter presenterte FHIs forskningsteam en meny med fem alternativer for hvordan implementeringen av utvidede åpningstider kunne gjøres. De presenterte også fordeler og ulemper fra både Vinmonopolets perspektiv og et forskningsperspektiv. Det ble oppnådd enighet om å velge alternativet som innebar klyngerandomisering av butikker på handelsdistriktnivå til en av tre faser, samt å ekskludere Oslo og tre andre store byer fra randomiseringen.

Vinmonopolet selv har formidlet seks grunner til at de til slutt ble enige om å iverksette utvidede åpningstider gradvis og randomisert:

- Vinmonopolets ønske om å bidra til forskningsbasert alkoholpolitikk
- Ønske om å imøtekomme helsemyndighetenes ønske om effektevaluering
- Opplevelse av ha god dialog med forskningsteamet og at Vinmonopolets bekymringer ble møtt med forståelse og fikk innflytelse på forskningsdesignet, bl.a. ved at de største byene fra fire forskjellige regioner ble utelukket fra randomiseringen
- Den foreslåtte blokkrandomiseringsstrategien fremsto praktisk gjennomførbart og gjorde det lettere for Vinmonopolet å kommunisere endringene
- Vinmonopolets egen arbeidsgruppe var bredt sammensatt og hadde et tydelig mandat om å finne en praktisk og gjennomførbart løsning
- Arbeidsgruppens forslag ble støttet av topplederne og fagforeningene

¹Rossow, Ingeborg, and Maja Weemes Grøtting. "Evaluation of an alcohol policy measure employing a randomised controlled trial design: Why was it possible?" *Drug and Alcohol Review* 40.3 (2021): 468-471.

Videre kan informasjons- og holdningsarbeid fremme større tverrpolitisk aksept for tidsavgrenset, systematisk og gradvis (og om mulig randomisert) utrulling av reformer og virkemidler.

3.7.2 Bygge kultur og heve kompetanse

Kartleggingen fremhever behovet for å bygge kultur for robuste effektevalueringer. Dette innebærer samarbeid på tvers i fagetatene, spesielt mellom de som jobber med utviklingen av tjenester og virkemidler og de som besitter analyse- og metodekompetanse.

Institusjonelle grep og kompetansehevingstiltak kan bidra til flere og bedre effektevalueringer. F.eks. har Norad gjennomført «forskningsinkubator» om effektevalueringer, med deltakelse fra sivilsamfunnsorganisasjoner som er aktive i bistandsfeltet.²¹ Skatteetaten besluttet i 2010 å bygge opp et analysemiljø, som i dag består av mer enn 80 analytikere med kompetanse innen sentrale fagområder for gjennomføring av randomiserte og kvasiexperimentelle studier. Videre har et Program for effektmåling blitt etablert, med mål om å øke Skatteetatens kunnskap om sammenhengen mellom virkemidler og effekter.

Kompetanse kan bidra til å øke bevisstheten om mulighetene og utfordringene knyttet til effektevalueringer. Dette kan legge til rette for grundige evalueringer av virkemiddelbruk i dialog med sektoransvarlig departement. Videre er kompetanse om forskningsmetoder og studietyper avgjørende for å kunne utforme og gjennomføre anskaffelser.

Både RCTer og kvasiexperimentelle metoder anvendes på tvers av fagområder som medisin, psykologi, økonomi, statsvitenskap og sosiologi. Disse fagfeltene kan ha noe ulike tradisjoner og tilnærminger til disse metodene, og det kan derfor være verdifullt med tverrfaglig sammensetning av kompetanse i etatene når slike metoder anvendes.

3.7.3 Tydelig juridisk og etisk rammeverk

De juridiske og etiske utfordringene skissert i avsnitt 4.5 peker på behovet for et tydelig lovverk og klare, konsistente prinsipper for å avklare rettsgrunnlag for vurderinger av forsvarligheten ved tidsavgrenset forskjellsbehandling. For øyeblikket er vurderingen av relativt likeartede virkemidler hjemlet i ulike lovbestemmelser, bl.a. avhengig av om den planlagte studien er å regne som helseforskning (Se eksempel i Boks 6: Særskilte krav til «helseforskning»).

Juridiske rammer som lover, forskrifter og rundskriv, samt enhetlig tolkning av disse, er nødvendig for å støtte forsøksvirksomhet i regi av forvaltningen. Kartleggingen viser at mange forvaltningsorganer står overfor et uavklart rettsgrunnlag når det gjelder gjennomføring av RCTer og kvasiexperimentelle studier, inkludert hjemmel til å registrere og behandle personopplysninger om deltakere i studiene. Videre er den juridiske tolkningen av handlingsrommet ikke alltid konsekvent på tvers av sektorer. Denne usikkerhet kan forhindre planlegging og gjennomføring av grundige effektevalueringer. Flere tolkningsuttalelser og veiledninger, samt kjennelser og dommer som etablerer praksis, kan være nyttig for å avhjelpe denne situasjonen.

²¹Invitation to Impact-Evaluation incubator, 5-7 December 2022. <https://www.Norad.no/en/front/about-Norad/news/2022/invitation-to-impact-evaluation-incubator-5-7-december-2022/>

3.7.4 *Bedre utnyttning av eksisterende data og utvikling av nye datakilder*

Enklere tilgang til nasjonale registre og tilrettelegging for kopling av forskjellige datakilder kan senke terskelen betydelig for gjennomføring av RCTer og kvasiekperimentelle studier i forvaltningen. Det ligger et betydelig potensial for dette i eksisterende datakilder, og i utvikling av nye registre for dette formålet.

3.7.5 *God medvirkning av berørte parter*

Virkemidlene som myndighetene iverksetter berører en rekke systemer og tjenester. God medvirkning av de som berøres er vesentlig for å forstå kunnskapsbehovet som skal møtes, for å utvikle intervensjonen og skape aksept for randomisering, evt. kvasiekperimentelle studier, som en tilnærming for å fremskaffe kunnskap. Solid medvirkning er også viktig for å sikre at en studie ikke medfører en uforholdsmessig byrde for tjenestene som berøres. En viktig lærdom fra den randomiserte studien av LOG-modellen, et systematisk opplegg for ledelse, organisering og utviklingsarbeid i skolen som ble prøvd ut i 19 barneskoler i fire kommuner, var betydningen av å forberede, forankre og tilpasse implementering til skolehverdagen i de ulike kontekstene.²²

3.7.6 *Kunnskapsorientert mål- og resultatstyring i forvaltningen*

Forvaltningsorganene beskriver et stort fokus på prosesser og aktiviteter som kan gi synlige resultater på kort sikt, med mindre prioritet viet til innsats som kan frembringe kunnskap på lengre sikt. F.eks. beskriver Skatteetaten at skatte- og avgiftsprosesser har en årlig syklus med sterke krav til aktiviteter og årlig resultatoppnåelse. Det er behov for å tilpasse mål- og resultatstyringen slik at beslutningsprosesser i større grad baserer seg på informasjon om effekter av virkemidlene som vurderes, og at det vies tilstrekkelig tid til å produsere denne kunnskapen.

3.7.7 *Finansiering og langsiktige forskningsprogram*

Randomiserte studier krever som oftest bevilgninger over flere år. Dersom hele prosessen fra beslutning til tilgjengelig effektmål går over flere stortingsperioder, vil en langsiktig finansiering ofte kreve tydelig tverrpolitisk enighet, enten midlene kommer fra direktorat, Forskningsrådet eller andre. Når politisk støtte foreligger kan større satsninger, f.eks. i form av langsiktige forskningsprogram i regi av Forskningsrådet, spille en viktig rolle i å tilrettelegge for effektevalueringer. Forskningssatsningen LÆREEFFEKT er et slikt eksempel.²³

Videre kan randomiserte studier integreres i programmer som skal følge opp myndighetenes politikk og reformer. F.eks. skal Kompetanseprogrammet, under HK-dir, samordne og styrke virkemidler i kompetansereformen lansert i forbindelse med stortingsmelding «Kompetansereformen – Lære hele livet». Et av programområdene handler om utprøvinger av insentivordninger for livslang læring og den tidligere nevnte RCTen av videreutdanningsstipend for fagarbeidere inngår i dette programmet (Boks 7).

²²<https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/et-lag-rundt-eleven-en-klyngerandomisert-effektevaluering-av-log-modellen/#:~:text=Et%20lag%20rundt%20eleven%3A%20LOG-modellen%20LOG-modellen%20er%20et,av%20nye%20og%20bedre%20m%C3%A5ter%20C3%A5%20samarbeide%20p%C3%A5.>

²³LÆREEFFEKT – Lærertetthet og læringseffekt. Rammedokument for forskningssatsing på effekter av flere lærere i skolen. https://www.forskningsradet.no/contentassets/965669a13e7c42e98bb879bdc2ddbdb/rammedokument_for_laereeffekt.pdf

3.7.8 Veiledning

Sentrale forutsetninger for vellykkede RCTer er et systematisert kunnskapsgrunnlag for å vurdere alternativer og informere utformingen av virkemiddelet, implementeringsstudier, metodekompetanse, og god medvirkning fra parter og tid til forankring i systemer og tjenester som berøres. For å ivareta disse komponentene, spesielt dersom forvaltningsorganer som har mindre erfaring med effektevalueringer ønsker å gjennomføre dem, kan en veileder/veiviser være et nyttig verktøy.

4 Konklusjon

Myndighetenes iverksettelse av virkemidler er ofte forbundet med usikkerhet om hva som er mest effektivt og hvem som vil få mest nytte av tiltak som innføres. For å imøtekomme denne usikkerheten og måle effekten av et virkemiddel, kan RCTer og kvasieksperimentelle studier være nyttig. Solide effektevalueringer fremmer også befolkningens rettigheter og velferd ved å sikre at de ikke utsettes for virkemidler som er unyttige eller skadelige.

Kartleggingen har identifisert flere eksempler på hvordan tett dialog mellom fagetater, departement og politisk ledelse har tilrettelagt for bedre effektevalueringer, som det kan trekkes lærdom fra. Flere forvaltningsorganer erfarer imidlertid at det er enklere å rulle ut programmer med usikker effekt i stor skala enn å begynne med gradvise endringer som forsøkes på med randomiserte eller kvasieksperimentelle metoder. Større politisk aksept for systematisk, gradvis – og om mulig, randomisert – utrulling av reformer vil kunne tilrettelegge for bedring læring om virkemidlenes effekt. Dette vil kunne føre til en mer effektiv og målrettet politikktutforming og øke sjansene for at effektive virkemidler skaleres opp, eller at uvirksomme virkemidler prioriteres vekk.

Randomiserte studier kan møte mange juridiske, etiske, politiske og praktiske utfordringer, og de kan være ressurskrevende. RCTene identifisert i denne kartleggingen har kostet mellom 1 og 35 millioner kroner. Gitt ressursinnsatsen er det derfor nødvendig å ha gode prioriteringsprosesser for hvilke virkemidler som velges for evaluering, kanskje særlig for RCTer.

For mange virkemidler er det mer gjennomførbart å gjennomføre en kvasieksperimentell evaluering enn en RCT, og disse metodene bør anses som supplerende tilnærminger. Vanligvis blir kvasieksperimentelle studier gjennomført *etter* at forskere har identifisert variasjoner i hvem som har blitt påvirket av virkemiddelet, f.eks. ved at et virkemiddel ble iverksatt til forskjellige tidspunkt ved forskjellige skoler, fastlegekontorer, NAV-kontorer, kommuner eller andre enheter. Det er imidlertid mulig å tilrettelegge for evaluering *på forhånd*, ved at virkemiddelet iverksettes på en måte som gjør det mulig å studere effektene av det. Dette fordrer tett samarbeid mellom beslutningstakere og forskere med relevant metodekompetanse.

Flere av utfordringene som denne kartleggingen reiser har blitt belyst i rapporten «*Juridiske og etiske problemstillinger knyttet til innsamling, tilgjengeliggjøring, deling og bruk av data, samt bruk av randomiserte forsøk, i kriser*», hvor det også skisseres mulige løsninger.²⁴ Konkrete eksempler inkluderer fastsettelse av en særlig lovhjemmel, f. eks. i forvaltningsloven, som avklarer de juridiske sidene ved gjennomføring av forsøk og utprøvinger i regi av forvaltningen; styrke rettslige vurderinger for bruk og deling av data for å redusere usikkerhet og unødig forsiktighet; og endre helseforskningsloven slik at REK kan gjøre unntak fra samtykkekravet når særskilte vilkår er tilfredsstilt.

Kartleggingen viser også at forskningsetiske vurderinger av ett og samme virkemiddel blir for øyeblikket vurdert ulikt, avhengig av hvilket lovverk som gjelder, f.eks. med hensyn til

²⁴Juridiske og etiske problemstillinger knyttet til innsamling, tilgjengeliggjøring, deling og bruk av data, samt bruk av randomiserte forsøk, i kriser. <https://www.regjeringen.no/contentassets/436542ebbc77473b96e4c45c5668b07a/juridiske-og-etiske-problemstillinger-knyttet-til-innsamling-tilgjengeliggjoringdeling-og-bruk-av-data-samt-bruk-av-randomiserte-forsok-i-kriser.pdf>

kravet om informert samtykke. En risikobasert vurdering av forholdsmessigheten mellom fordeler og ulemper av en studie og behovet for individuelt samtykke når virkemiddelet er befolkningsrettet, kan bidra til en mer enhetlig tilnærming når den etiske forsvarligheten til studier vurderes.

Inspirasjon kan hentes fra utviklingen internasjonalt (Boks 11). F.eks. etablerte britiske myndigheter i 2013, som del av sin satsning på å fremme kunnskapsbasert praksis og politikkkutforming, «What works initiative».²⁵ Dette omfatter et nettverk av flere forskningssentre som skal understøtte beslutningstakere ved å fremskaffe kunnskap om effekter av virkemidler innen bl.a. helse, utdanning og kriminalomsorg. Som ledd i satsningen har embetsverket blitt kurset i metoder for effektevaluering. Videre har et «Trial Advice Panel» blitt etablert, som embetsverket kan benytte for å få ekspertråd om hvilke metoder som egner seg når virkemidler skal evalueres. Siden 2015 har panelet gitt råd til over 170 prosjekter på tvers av departementer, på områder som utdanning, transport, utdanning, kriminalitet og mer.²⁶ En liknende mekanisme, i form av et kompetansemiljø som kan veilede forvaltningen i bruken av RCTer og kvasiexperimentelle metoder, ble i 2022 foreslått av ekspertgruppen nedsatt av kjernegruppen.²²

Kartleggingen viser et stort potensial for å dele kunnskap på tvers av etater og fremme effektevalueringer som en tverrsektoriell prioritet. Etatene har delt erfaringer som understreker at kunnskapsbasert politikkkutforming prioriteres høyt i forvaltningen, og at kompetansen for å levere grundige effektevalueringer har økt over tid. Faktorer som detaljerte nasjonale registre, godt samarbeid mellom forvaltningen og kompetente fagmiljøer, og befolkningens tillit til offentlige myndigheter og forvaltning tyder på at Norge har et solid utgangspunkt for flere og bedre effektevalueringer av virkemidler foreslått av politisk ledelse og forvaltningen.

²⁵https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/677478/6.4154_What_works_report_Final.pdf

²⁶<https://www.gov.uk/government/publications/cross-government-trial-advice-panel-role-and-membership>

Boks 11: Effektevalueringer som grunnlag for politikktutforming – et internasjonalt perspektiv

De siste 20 årene har det vært økende oppmerksomhet omkring bruken av effektevalueringer for å informere politikktutforming, spesielt innen helse- og utviklingsforskning. Internasjonale fagmiljøer, som f.eks. Abdul Latif Jameel Poverty Action Lab (J-PAL) som inkluderer Nobelprisvinnerne Esther Duflo og Abhijit Banerjee, har vært førende i utviklingen. De har bl.a. gjennomført flere RCTer på effektene av utdannings- og helsetiltak, i samarbeid med offentlige myndigheter og forskningsmiljøer i lav- og mellominntektsland. J-PAL var sentrale da en RCT av effekten av gratis utdeling og informasjonstiltak for å fremme bruken av munnbind ble gjennomført i Bangladesh. Dette er for øvrig en studie som etter gjeldende lovverk ikke ville latt seg gjennomføre i Norge.¹

Et annet eksempel er International Initiative for Impact Evaluation (3ie), som har spilt en ledende rolle i å fremme bruken av effektevalueringer ved å finansiere studier, utvikle retningslinjer for god praksis, og bidra til å bygge kapasitet innen effektevaluering i lav- og mellominntektsland. Evidensportalen til 3ie inneholder i dag over 11 000 effektevalueringer innen utvikling og bistand. Portalen vokser raskt: det ble i 2020 publisert rundt 4 ganger flere effektevalueringer enn i 2010.

Etatene som er inkludert i denne kartleggingen har også bidratt til effektevalueringer i lav- og mellominntektsland. Norad har finansiert flere RCTer av resultatbasert finansiering av helsetjenester, og planlegger for øyeblikket en RCT av lokalprodusert skolemat i Malawi, i samarbeid med Verdens matvareprogram. FHI har samarbeidet med fagmiljøer og offentlige myndigheter om flere RCTer. I Uganda ble effekten av utdanningstiltak på kritisk vurdering av helsepåstander evaluert.² I Palestina ble effekten av innføring av elektroniske pasientjournaler på kvaliteten på svangerskapsomsorg evaluert.³ Flere RCTer av liknende tiltak er pågående i Uganda, Kenya og Bangladesh.

¹<https://www.aftenposten.no/meninger/debatt/i/jaK78n/munnbindstudie-ville-vaert-ulovlig-i-norge>

²Nsangi, Allen, et al. "Effects of the Informed Health Choices primary school intervention on the ability of children in Uganda to assess the reliability of claims about treatment effects: a cluster-randomised controlled trial." *The Lancet* 390.10092 (2017): 374-388

³Venkateswaran, Mahima, et al. "A digital health registry with clinical decision support for improving quality of antenatal care in Palestine (eRegQual): a pragmatic, cluster-randomised, controlled, superiority trial." *The Lancet Digital Health* 4.2 (2022): e126-e136

Vedlegg 1: Spørreskjema

Del 1: Introduksjon og definisjoner

Gjennom tildelingsbrevet har Folkehelseinstituttet fått ansvar for å koordinere en kartlegging av hvordan relevante statlige etater bruker **randomiserte** og **kvasiekperimentelle** studier for å gjøre virkemiddelbruken mer kunnskapsbasert. Vi er interessert i erfaringer fra studier som etaten har forsøkt å gjennomføre selv, studier etaten har anskaffet gjennom oppdrag til forskere eller studier som etaten på andre måter har vært en aktiv part i.

Kartleggingens målsetning er å presentere etatenes erfaringer med slike metoder, styrkene, svakhetene og begrensningene ved å bruke slike metoder, og hva som hindrer bruken av dem. Oppdragsteksten i tildelingsbrevet beskrev FHI, Bufdir, NAV, Udir, Skatteetaten og Norad å være blant de mest relevante etatene for et slikt oppdrag, men at også andre statlige og kommunale etater kan bli invitert. FHI er tildelt ansvaret for å koordinere gjennomføringen av denne kartleggingen og presentere resultatene i en rapport innen 1. mai 2023.

Sentrale definisjoner

1. Virkemiddel

Statlige, kommunale eller fylkeskommunale etater eller virksomheter kan bruke en rekke virkemidler²⁷ for å iverksette politikk. Slike virkemidler inkluderer:

- juridiske virkemidler, for eksempel nye lover og forskrifter
- økonomiske virkemidler, for eksempler endringer i skatter og avgifter
- organisatoriske virkemiddel, for eksempel omorganisering, sentralisering og desentralisering av statlige oppgaver
- pedagogiske virkemidler, for eksempel veiledere, retningslinjer, kompetanseutviklingstiltak eller holdningskampanjer

Disse virkemidlene, i større eller mindre skala, kan være utgangspunkt for effektevaluering med randomiserte eller kvasiekperimentelle studier.

2. Randomiserte kontrollerte studier

I randomiserte studier måles effektene av et virkemiddel ved å sammenligne deltakere som fordeles tilfeldig, ved loddtrekning eller tilsvarende, til å bli eksponert eller ikke bli eksponert for virkemiddelet. Randomisering kan også foregå gjennom gradvis og tilfeldig innfasing av et virkemiddel. Deltakerne kan være individer eller grupper av individer, f.eks. skoleklasser, skoler eller kommuner. Sistnevnte kalles vanligvis en klyngerandomisert studie. Randomisering av tilstrekkelig antall deltakere/klynger sannsynliggjør at intervensjonsgruppen og kontrollgruppen er like mht. kjennetegn som kan påvirke utfallet, slik at eventuelle forskjeller i utfallet kan tilskrives virkemiddelet som iverksettes.

3. Kvasiekperimentelle studier

I kvasiekperimentelle studier har det ikke skjedd en tilfeldig fordeling til intervensjons- eller kontrollgruppen. Årsakssammenhenger lar seg likevel avdekke fordi fordelingen av individer

²⁷ Vi har brukt klassifiseringen benyttet av Direktoratet for forvaltning og økonomistyring: <https://www02.dfo.no/fagomrader/statsforvaltningen/virkemidler>

eller klynger til et virkemiddel har skjedd på en slik måte at gruppene er sammenlignbare. Graden av sammenlignbarhet vil variere mellom ulike typer kvasiekperimentelle studier.²⁸

Del 2: Bruk av randomiserte studier om effekter av virkemidler

Med spørsmålene i del 2 ønsker vi å kartlegge erfaringer med **randomiserte** studier etaten har gjennomført selv, har anskaffet gjennom oppdrag til forskere eller på andre måter fremmet gjennomføringen av (f. eks. som en aktiv part i et samarbeid med forskere).

Omfang

1. Anslagsvis, hvor mange randomiserte studier har etaten gjennomført—enten med egen kapasitet, gjennom anskaffelser og/eller gjennom å være aktiv part i et samarbeid med forskere—de siste fem årene?

▪ Ingen	▪ 1-5	▪ 6-15	▪ >15	▪ Vet ikke
---------	-------	--------	-------	------------

2. Anslagsvis, hvor høy **andel** av disse er gjennomført av andre (e.g., forskere fra instituttsektoren, universitetet, høyskoler) gjennom anskaffelser? Angi i prosent.
3. Har dere eksempler på virkemidler som har vært evaluert med randomiserte studier? Beskriv
 - ✓ Målgruppe som var fokus for forsøket
 - ✓ Virkemiddelet/tiltaket som testes
 - ✓ Ønskede/uønskede resultater
 - ✓ Overordnet lærdom fra studien og prosessen
4. Hvor mange randomiserte studier har vært forsøkt planlagt, men **stoppet** grunnet ulike typer utfordringer? (f.eks., etiske, juridiske, politiske, annet). Gi gjerne eksempler på konkrete studier som har blitt stoppet og utfordringene som ble møtt.
5. Hva anser dere som **fordeler** ved bruk av randomiserte studier?
6. Hva anser dere som **metodologiske utfordringer** ved bruk av randomiserte studier?
7. På hvilke måter har etaten forsøkt å **tilrettelegge** for at et virkemiddel lar seg evaluere med randomiserte studier?

Kvalitet og datagrunnlag

8. Hva gjør etaten for å sikre at randomiserte studier utføres i tråd med vitenskapelige prinsipper og kan gi **troverdige** svar? *Eksempler kan være: sikre at studiene har tilstrekkelig statistisk styrke, optimal randomiseringsteknikk er valgt osv.*
9. Hvilke **datakilder (e.g. helseregistre eller spørreundersøkelser)** har vært brukt for randomiserte studier gjennomført med egen kapasitet og/eller gjennom anskaffelser?

²⁸Et eksempel på en kvasiekperimentell studie er denne fra Kinge og Grytten som utnytter forskjeller i legetetthet mellom kommuner, som følge av jobb i noen kommuner innebar sletting av studielån, og bruker *difference-in-difference metode* for å avdekke at forskjellene i legetetthet har effekt på spedbarnshelse: <https://onlinelibrary.wiley.com/doi/full/10.1002/hec.4426>

10. Hvilke **utfordringer** har etaten eller forskere som etaten har engasjert møtt i forbindelse med bruk av slike datakilder?

Utfordringer og muligheter

11. Hvilke **lovhjemler og forskrifter** bruker etaten for å forsvare gjennomføring av randomiserte studier?
For eksempel åpner §12 i arbeidsmarkedsloven opp for forsøk: "Departementet kan gi forskrifter om forsøk med arbeidsmarkedstiltak."
12. Hvilke **juridiske utfordringer** har lovhomeiler og forskrifter skapt for gjennomføring av randomiserte studier?
13. Hvilke **etiske utfordringer** har etaten eller forskere etaten samarbeider med møtt i forbindelse med planlegging og gjennomføring av randomiserte studier?
14. Hvilke **politiske utfordringer** har etaten eller forskere etaten samarbeider med møtt i forbindelse med planlegging og gjennomføring av randomiserte studier? *Et eksempel på en politisk utfordring er at tidsvinduet for å iverksette og rulle ut et virkemiddel i stor skala ikke nødvendigvis passer med tiden som trengs for å gjennomføre en grundig effektevaluering.*
15. Utover etiske, juridiske og politiske utfordringer, hvilke **andre faktorer** hindrer utstrakt bruk av randomiserte studier for å evaluere tiltak anbefalt og iverksatt av etaten?
16. Hva kan **tilrettelegge** for økt bruk av randomiserte studier for å evaluere etatens virkemiddelbruk?
17. Hvilke virkemidler under etatens sektoransvar **egner** seg for evaluering med randomiserte studier?
18. Hvilke virkemidler under etatens sektoransvar **egner seg ikke** for evaluering med randomiserte studier?

Del 3: Bruk av kvasieksperimentelle studier om virkemidler

Med spørsmålene i del 3 ønsker vi å kartlegge erfaringer med kvasieksperimentelle studier etaten har gjennomført selv, har anskaffet gjennom oppdrag til forskere eller på andre måter fremmet gjennomføringen av.

1. Anslagsvis, hvor mange kvasieksperimentelle studier har etaten gjennomført—enten med egen kapasitet, gjennom anskaffelser og/eller gjennom å være aktiv part i et samarbeid med forskere—de siste fem årene?

▪ Ingen	▪ 1-5	▪ 6-15	▪ >15	▪ Vet ikke
---------	-------	--------	-------	------------

2. Anslagsvis, hvor høy **andel** av disse er gjennomført av andre (e.g., forskere fra instituttsektoren, universitetet, høyskoler) gjennom anskaffelser? Angi i prosent.

3. Kan dere gjengi eksempler på virkemidler som har vært evaluert med kvasieksperimentelle studier?
4. Har dere eksempler på kvasieksperimentelle studier som har vært forsøkt planlagt, men **ikke ferdigstilt** grunnet ulike typer utfordringer? (f.eks., etiske, juridiske, datamangel, virkemiddelet ikke iverksatt som planlagt). *Gi gjerne eksempler på konkrete studier som har blitt stoppet og utfordringene som ble møtt.*
5. På hvilke måter har etaten forsøkt å tilrettelegge for at et virkemiddel lar seg evaluere med kvasieksperimentelle studier?
6. Hva anser etaten som **fordeler** med å gjennomføre kvasieksperimentelle studier?
7. Hva anser etaten som **metodologiske utfordringer** med å gjennomføre kvasieksperimentelle studier?

Kvalitet og datagrunnlag

8. Hva gjør etaten for å sikre at kvasieksperimentelle studier utføres i tråd med vitenskapelige prinsipper og kan gi **troverdige** svar? *Eksempler kan være: oppmuntre praksis som forhåndsregistrering; kurs og andre tiltak for kompetanseheving.*
9. Hvilke **datakilder (e.g. helseregistre eller spørreundersøkelser)** har vært brukt for kvasieksperimentelle studier gjennomført med egen kapasitet og/eller gjennom anskaffelser?
10. Hvilke **utfordringer** har etaten eller forskere som etaten har engasjert møtt i forbindelse med bruk av slike datakilder?

Utfordringer og muligheter

11. Hvilke **juridiske utfordringer** har lovhjemler og forskrifter skapt for gjennomføring av kvasieksperimentelle studier?
12. Hvilke **etiske utfordringer** har etaten eller forskere etaten samarbeider med møtt i forbindelse med planlegging og gjennomføring av kvasieksperimentelle studier?
13. Hvilke **politiske utfordringer** har etaten eller forskere etaten samarbeider med møtt i forbindelse med planlegging og gjennomføring av kvasieksperimentelle kontrollerte studier? *Et eksempel på en politisk utfordring er at tidsvinduet for å iverksette og rulle ut et virkemiddel i stor skala ikke nødvendigvis passer med tiden som trengs for å gjennomføre en grundig effektevaluering.*
14. Utover etiske, juridiske og politiske utfordringer, hvilke **andre faktorer** hindrer utstrakt bruk av kvasieksperimentelle studier for å evaluere virkemidler anbefalt og iverksatt av etaten?
15. Beskriv hvilke faktorer som kan **tilrettelegge** for økt bruk av kvasieksperimentelle studier for å evaluere etatens virkemiddelbruk?

16. Hvilke virkemidler under etatens sektoransvar **egner seg** for evaluering med kvasiekperimentelle studier?
17. Hvilke virkemidler under etatens sektoransvar **egner seg** ikke for evaluering med kvasiekperimentelle studier?

Del 4: Generelle spørsmål om kapasitet og forutsetninger for å gjennomføre effektevalueringer

I denne siste delen stiller vi generelle spørsmål om etatens forutsetninger for å fremme bruken av randomiserte studier og kvasiekperimentelle studier—herfra betegnet som **effektevalueringer**.

18. Hvor mange jobber med å gjennomføre og/eller anskaffe effektevalueringer?
- Ingen
 - 1-5
 - 6-15
 - >15
19. Hvilken **kompetanse** trenger etaten mer av for å kunne gjennomføre og/eller samhandle med eksterne forskere og oppnå flere effektevalueringer?
20. Kan dere beskrive **ansvarsfordeling** og **organisering** mellom de som utvikler, anbefaler og iverksetter virkemidler og de som har ansvar for å fremskaffe effektevalueringer?
21. Hvilke **publiseringskanaler** (e.g., interne rapporter, fagfelleverderte tidsskrift, eksternt tilgjengelige rapporter, annet) brukes for å formidle gjennomførte effektevalueringer? Del gjerne lenker til publiserte rapporter og artikler.
22. Er det **andre erfaringer** etaten ønsker å tilføye, som ikke ble dekt av spørsmålene i del 2-4?

Vedlegg 2: Barne-, ungdoms- og familiedirektoratets (Bufdir) erfaringer med randomiserte og kvasieksperimentelle studier

Organisering, publiseringskanaler og kvalitetssikring

Barne-, ungdoms- og familiedirektoratet (Bufdir) er underlagt Barne- og familiedepartementet, og er et fagorgan på områdene barnevern, barn, ungdom og oppvekst, adopsjon, familievern, likestilling og ikke-diskriminering og vold og overgrep i nære relasjoner. I Bufdir er arbeidet med forskningsanskaffelser og metodekunnskap organisert i en egen avdeling i Kunnskapsdivisjonen. Avdelingen samarbeider med fagansvarlige avdelinger som har ansvar for å utvikle, iverksette eller anbefale tiltak, metoder og program. Bufdir arbeider aktivt med å få til forskning som er godt informert av brukere og praksisfeltet. Bufdir bruker en kunnskapsbasert tilnærming der innhenting og bruk av kunnskapsoppsummeringer står sentralt.

For å sikre høy vitenskapelig kvalitet på studiene har Bufdir siden 2022 krevd at anskaffelsene skal ha forhåndsregistrering av forskningsprotokoller og rapportering i samsvar med internasjonale retningslinjer. Oppnåelse av nødvendig statistisk styrke er også et prioritert fokus. Videre er Bufdir involvert i valg av utfallsmål og bidrar til å tilrettelegge for en så god gjennomføring som mulig.

Rapporter til Bufdir.no publiseres gratis gjennom Nettbiblioteket (<https://www2.bufdir.no/Bibliotek/Dokumentside/?docId=BUF00005409>). Forskere står også fritt til å skrive artikler, konferansebidrag, arrangere lanseringsseminar og lignende.

Omfang og eksempler på evaluerte virkemidler

Bufdir er opptatt av å ha solid og pålitelig kunnskapsgrunnlag når det gjelder valg av virkemidler som kan bidra til å forbedre situasjonen for barn, ungdom og familier. Tre randomiserte studier er anskaffet de siste fem årene. ICDP er et foreldreveiledningsprogram som ble testet i en klynge-randomisert studie. Målgruppen var alle foreldre med barn og utfallsmål var foreldreferdigheter og barns livskvalitet. Studien hadde imidlertid kort oppfølgingstid og andre metodologiske svakheter ved gjennomføringen slik som manglende blinding. Resultatene viste ingen forskjell i utfallsmål mellom ICDP og venteliste. Dette var overraskende fordi i forkant var det høye forventninger til programmets positive effekt.

Gravide med sammensatte utfordringer er målgruppen for programmet «Sammen på vei», også kjent som Nurse Family Partnership (NFP). Arbeidet med denne RCTen begynte i 2022 og vil starte rekruttering i 2023. Studien vil fortsette frem til og med 2027. Programmet tilbys som tillegg til vanlig svangerskap- og barselsoppfølging og sammenliknes med vanlig praksis. Programmet har en lang rekke utfallsmål knyttet til den gravide, barnet, helse, vold og samspill. Denne studien møtte flere forskningsetiske og praktiske utfordringer.

«Jeg vet» er et undervisningsopplegg som er rettet mot barneskolen. Arbeidet med å planlegge denne RCTen startet i 2023. «Jeg vet» programmet vil testes ut som en skole-RCT i 2024. Formålet med programmet er å øke kunnskapen om mobbing, vold og seksuelle overgrep.

Bufdir har ikke anskaffet kvasieksperimentelle studier for å evaluere sin virkemiddelbruk de siste årene.

Utfordringer

Metodologisk er rekruttering av tilstrekkelig antall deltakere for å oppnå statistisk styrke en stor utfordring, spesielt for tjenester der brukergruppene kan være av begrenset størrelse.

Forskningsetisk er det utfordrende når virkemidlene retter seg mot sårbare grupper og situasjoner, eksempelvis vold og seksuelle overgrep. For å innhente data om barn kreves samtykke fra begge foreldrene, noe som kan være en utfordring. Spesielt i situasjoner der foreldrene er i konflikt i tillegg er det uklart hvordan meldeplikt ved alvorlige situasjoner og oppfølging av deltakerne skal håndteres, noe forskere opplever som spesielt utfordrende. En annen etisk innvending, blant annet fra ansatte i tjenestene, er skepsis mot randomiserte studier fordi kontrollgruppen ikke får det «beste» tilbudet. Videre har Bufdir observert at forskere har hatt utfordringer med å få veiledning og hjelp fra offentlige instanser, samt at behandling av søknader for eksempel til regional etisk komite har vært tildels svært tidkrevende.

Politisk kan det være et press for å produsere resultater som kan brukes innenfor gjeldende stortingsperiode, selv om dette ikke alltid er forenlig med tiden som er nødvendig for å gjennomføre en RCT. Så langt har imidlertid dette ikke vært et problem knyttet til RCTene som er anskaffet, planlagt og gjennomført i perioden.

Datagrunnlag: For å gjennomføre effektevalueringer legges det til rette for bruk av eget barnevernsregister, samt spørreskjemaer, nasjonale helseregister og SSB-register. Tilstrekkelig svarprosent på spørreskjemaer og tilgang til og kobling av registerdata er sentrale utfordringer.

Andre utfordringer inkluderer

- **Oppfølging:** Å følge opp RCTer er krevende med tanke på planlegging, kontraktsoppfølging og forståelse for forskningen.
- **Kostnader og flerårige bevilgninger:** RCTer kan være kostnadskrevende og krever ofte bevilgninger over flere år, noe som kan skape utfordringer med hensyn til forutsigbarhet. I noen tilfeller kan budsjettene også mangle stikkordsfullmakter for overføringer fra ett år til det neste, noe som kan gjøre det vanskelig å håndtere forsinkelser i forskningsprosjekter.
- **Kompleks setting:** Komplekse organisasjoner og tjenester gjør det utfordrende å gjennomføre RCTer i uten at det blir en byrde for vanlig drift.

Muligheter

Bufdir har erfart at innsats på følgende områder kan tilrettelegge for effektevalueringer

- **Styrke kompetanse:** Det er behov for ulike typer kompetanse – jus og metode – for å oppnå flere effektevalueringer. Mer kompetanse om forskningsmetodikk, spesielt RCTer, vil være spesielt fordelaktig for organisasjonen, da det er begrenset erfaring med RCT grunnet de få studiene som er gjennomført til nå.
- **Forankring tverrsektorielt og på ledernivå i forvaltningen:** Forankring på ledernivå i departementer og direktoratet, og på tvers av sektorer som helse og utdanning, kan bidra til økt bruk av slike studier der det er relevant.

- **Investere tid i medvirkning:** ansatte i tjenestene som blir fordelt til tiltaks- og kontrollgrupper bør få muligheter til å bli hørt og gi innspill til planleggingen og gjennomføringen av studier. Bufdir opplevde at det var viktig å besøke tjenestene der RCTen skulle foregå for å lytte til både begeistring og skepsis og diskutere innvendingene åpent.
- **Engasjere og finansierte kompetente miljøer:** Utlysning av forskningsoppdrag og bevilgning av midler til forsknings- og kompetansemiljøer som er rustet til å gjennomføre effektevalueringer
- **Langsiktige forskningssatsninger:** Sikring av bevilgninger over budsjettet over flere år kan være utfordrende, og forskningsmidler som disponeres gjennom langsiktige forskningssatsninger er nødvendig

Vedlegg 3: Folkehelseinstituttets erfaringer med randomiserte og kvasiexperimentelle studier

Omfang og eksempler på evaluerte virkemidler

Folkehelseinstituttet (FHI) har de siste fem årene ledet og bidratt til en rekke randomiserte (>15) og kvasiexperimentelle studier (>15). Mange studier er ledet av FHIs egne forskere. FHI-forskere er også bidragsytende til studier ledet av andre fagmiljøer.

Flere randomiserte studier (RCTer) har vært gjennomført i lav- og mellominntektsland som del av FHIs globale helseforskning. Eksempler inkluderer RCTer på effekten av ekstra helseyskepleierressurs i skolene, implementering av elektroniske helseregistre, munnbind, Facebook-kampanje for å øke koronatesting blant innvandrergupper, og utvidelsen av Vinmonopolets åpningstider.

En rekke kvasiexperimentelle studier—med bruk av metoder som forskjell-i-forskjeller, avbrutte tidsserier og regression discontinuity design—har også blitt gjennomført, f.eks. for å evaluere virkningene av å tilby nye vaksiner.^{29,30} En fullstendig oversikt lar seg ikke oppdrive. Mange studier er forskerinitiert, som for eksempel evalueringen av effekter av gratis skolefrukt på skoleprestasjoner og effekter av standardiserte tobakkspakninger på tobakksbruk. I noen tilfeller blir studier gjennomført på oppdrag fra Helse- og omsorgsdepartementet (HOD). Under covid-19 pandemien ble en rekke kvasiexperimentelle evalueringer av smitteverntiltak gjennomført bl.a. for å kunne understøtte FHIs rolle som rådgiver, inkludert effekter av stengte skoler eller stengte barer på smittespredning og effekter av dør-til-dør aksjoner og mobile teststasjoner på omfanget av koronatesting.

Organisering, publiseringskanaler og kvalitetssikring

Ansvarsfordeling og organisering mellom de som er ansvarlig for rådgivning og de som er ansvarlig for forskning varierer innad i FHI. Noen deler av instituttet driver mye forskning og lite rådgivning, eller omvendt, mens andre deler er tungt engasjert i begge aktiviteter. For eksempel spiller mange av medarbeiderne på smittevernfeltet en viktig rolle som rådgivere overfor myndigheter og publikum, samtidig som de er engasjert i forskning. Enkelte miljøer ved instituttet driver kun med rådgivning og enkelte kun med forskning.

Effektevalueringer gjøres tilgjengelig hovedsakelig i vitenskapelige tidsskrift, men også som egne rapporter på instituttets hjemmesider. Strategier for å oppnå effektevalueringer av høy vitenskapelig kvalitet og troverdighet inkluderer: (1) Bygging av intern kultur med vekt på høy vitenskapelig standard (f.eks. pre-registrering og forhåndsspesifisering av antagelser som kan påvirke modelleringen av effektestimater); (2) Tilstrebe publisering av resultater i vitenskapelige tidsskrift og; (3) Styrke metodekompetanse gjennom erfaringsutveksling mellom personer med ulik faglig bakgrunn (e.g., økonomi, medisin,

²⁹ Bruun T. et al. Impact of the Rotavirus Vaccination Program in Norway After Four Years With High Coverage. *The Pediatric Infectious Disease Journal* 40(4):p 368-374, April 2021.

³⁰ Feiring B et al. Substantial Decline in Prevalence of Vaccine-Type and Nonvaccine-Type Human Papillomavirus (HPV) in Vaccinated and Unvaccinated Girls 5 Years After Implementing HPV Vaccine in Norway. *J Infect Dis.* 2018 Nov 5;218(12):1900-1910

sosiologi, psykologi) som kan ha forskjellige tilnærminger til gjennomføringen av kvasiexperimentelle studier.

Utfordringer

Noen av disse utfordringene har forhindret gjennomføringen av randomiserte studier – en problemstilling som ble spesielt fremtredende under covid-19 pandemien:

- **Overordnet** er det en svak tradisjon for å evaluere effekter av forebyggende helsetiltak, som smitteverntiltak eller befolkningsrettede tiltak for å fremme sunnere kosthold eller fysisk aktivitet, med randomiserte studier
- **Metodologisk** er oppnåelse av tilstrekkelig statistisk styrke og smitteeffekter mellom intervensjons- og kontrollgruppene en utfordring, spesielt for randomiserte studier.
Eksempelvis ble en RCT av munnbind i militærleirer forhindret av en antagelse om at tilstrekkelig statistisk styrke ikke kunne oppnås
- **Juridisk** er Helseforskningslovens krav om individuelt samtykke den mest fremtredende utfordringen, som i praksis gjør det umulig å gjennomføre RCTer av tiltak rettet mot grupper av i befolkningen. F.eks. har kravet forhindret RCTer av luftrensere eller hyppig utlufting av klasserom i skoler. Lovverket legger dessuten begrensninger på tilgang til og kobling av sanntidsdata – det store nyten av tilgang til og kobling av registerdata ble tydelig vist under koronapandemien gjennom bruken av beredskapsregisteret (BEREDT-C19).
- **Politiske** utfordringer inkluderer skepsis til å forfordel deler av befolkningen; manglende forståelse for verdien av randomisering; og manglende politisk vilje og tålmodighet til å gradvis iverksette virkemidler for å tilrettelegge for evaluering. F.eks. har RCTer av massetesting i skolen eller effekten av 4. dose koronavaksine ikke vært mulig grunnet manglende politisk støtte.
- **Forskningsetikk** omfatter bl.a. innvendinger mot å frata kontrollgruppen et nyttig tiltak eller utsette den for økt risiko; skepsis mot RCTer som involverer barn; og innvendinger mot at informert samtykke skal unntas for befolkningsrettede studier. F.eks. ble en studie av effekten av hurtigtesting av deltakere ved større arrangement ikke mulig fordi lokale myndigheter var bekymret for risikoen for smitteutbrudd i kontrollgruppen
- **For datagrunnlaget** gjør mangel på data, mangel på relevante opplysninger i tilgjengelige registre eller tidkrevende prosesser for å kople registre det vanskelig å gjennomføre studier. F.eks. har kvasiexperimentelle studier av effekter av helsepersonells arbeidsbelastning og sykefravær på pasienters helseutfall (e.g., liggetider, sykkelighet, død) ikke vært mulig grunnet mangel på data på helsepersonell. Videre er rekruttering av deltakere og lav svarandel en tilbakevendende utfordring.
- **Praktisk** interagerer mange virkemidler med andre tiltak og prosesser, noe som kan gjøre effektmåling vanskelig, også i en RCT. Noen typer RCTer, som f.eks. utplassering av luftrensere i barer, kan være praktisk/logistisk vanskelig å gjennomføre.

Muligheter

Internt har FHI lagt til rette for bedre effektevalueringer gjennom følgende innsatsområder

- Vektlegging av mer og bedre effektevalueringer fra ledelsen og i måldokumenter

- Tett dialog med departement, kommuner og bydeler samt implementerende aktører (som f.eks. Vinmonopolet³¹) for å fremme randomisering/gradvis utrulling av tiltak
- Årlige undersøkelser/datainnsamling som kan brukes til kvasiexperimentelle senere
- Ansettelse av forskere med relevant metodekompetanse
- Dokumentasjon av når og hvor virkemidler har blitt innført for å tilrettelegge for kvasiexperimentelle evalueringer: f.eks. har FHI utviklet en pandemidatabase med strukturert oversikt over alle lokale og nasjonale forskrifter med smitteverntiltak

Eksternt kan følgende tilrettelegge for bedre effektevalueringer

- Økt kunnskapsnivå og større forståelse blant politikere, byråkrater og publikum om verdien av randomisering eller annen tidsavgrenset forskjellsbehandling for å oppnå bedre effektevalueringer
- Etablere rutiner for å prioritere effektevaluering av store, inngripende og kostbare tiltak med høy risiko, jf. Finansdepartementets veileder til gjennomføring av evalueringer
- Investering i å utvikle datakilder som er tilgjengelige for bruk i effektevaluering, og som er mulige å koble med hverandre
- Finansiering og instruksjoner, f. eks. i tildelingsbrev, om å gjennomføre effektevalueringer
- Innføring av virkemidler på en måte som gjør det mulig å etablere troverdige sammenligninger og bedre oversikt over når og hvor ulike virkemidler som blir innført
- Fjerne Helseforskningslovens absolutte krav om individuelt samtykke ved gjennomføring av studier i befolkningen, dvs. evaluering av tiltak rettet mot grupper
- Forskningsutlysninger som prioriterer gjennomføring av randomiserte eller kvasiexperimentelle evalueringer

³¹ Rossow, Ingeborg, and Maja Weemes Grøtting. "Evaluation of an alcohol policy measure employing a randomised controlled trial design: Why was it possible?." *Drug and Alcohol Review* 40.3 (2021): 468-471.

Vedlegg 4: NAVs erfaringer med randomiserte og kvasiekperimentelle studier

Omfang og eksempler på evaluerte virkemidler

NAV har de siste fem årene selv gjennomført eller anskaffet en rekke randomiserte (>15) og kvasiekperimentelle studier (>15). RCTer har evaluert virkemidler som Individuell jobbstøtte, Ny medisinsk vurdering, og Raskt i jobb for flyktninger. I sistnevnte prosjekt var målgruppen flyktninger som deltok på introduksjonsprogrammet i Bergen kommune. Tiltaksgruppen fikk Raskt i jobb i tillegg til den ordinære oppfølgingen i introduksjonsprogrammet, mens kontrollgruppen mottok vanlig oppfølging gjennom introduksjonsprogrammet som er obligatorisk for alle flyktninger i Norge. Oppsummert viste resultatene fra effektevalueringen at over halvparten av deltakerne i tiltaksgruppen kom i ordinært arbeid i løpet av oppfølgingsperioden på 12 måneder. Den tilsvarende andelen i kontrollgruppen var i underkant av 8 prosentpoeng lavere. Forskjellen mellom gruppene på arbeidsdeltakelse og antall måneder i arbeid var imidlertid ikke statistisk signifikant.

NAV har også anskaffet kvasiekperimentelle studier eller gjennomført slike med intern kompetanse. Konkrete eksempler inkluderer effektstudier av NAV-veileder i videregående skoler, forsterket ungdomsinnsats i NAV og kortere maksimal varighet på AAP.

Organisering, publiseringskanaler og kvalitetssikring

Gjennomføring av forsøk driftes hovedsakelig av NAVs fagavdelinger. Videre har NAV en FoU-seksjon som arbeider med å anskaffe og/eller finansiere evalueringen av forsøket og som kan støtte fagavdelinger med råd om forskningsdesign. Etter vurdering av søknader blir prosjekter anskaffet eller midler tildelt. Når søknader vurderes brukes interne ressurser med relevante metodekompetanse for å vurdere forskningsdesignet, i tillegg til kompetanse og dokumentert erfaring med effektevalueringer blir vurdert. Effektevalueringene formidles med rapporter fra prosjektene og i sosiale medier som Facebook-siden NAV kunnskap.

Utfordringer

Metodologisk er oppnåelse av statistisk styrke, smitteeffekter og tid før relevante utfallsmål påvirkes sentrale utfordringer for randomiserte studier

- **Statistisk styrke:** mange virkemidler i NAV-sammenheng er best egnet på kontornivå, som skaper behov for mange «klynger», noe som gjør oppnåelse av tilstrekkelig statistisk styrke ressurskrevende
- **Smitteeffekter:** NAV-veiledere er vant til å dele på tvers, noe som medfører risiko for at praksisen som testes spres til NAV-kontorer i kontrollgruppen
- **Tid:** utfallene som myndighetene er opptatt av tar tid før de materialiseres

Juridisk har usikkerhet om lovligheten av forskjellsbehandling sannsynligvis hindret igangsetting av RCTer og forhindret kvasiekperimentelle evalueringer. For eksempel har det ikke vært mulig å evaluere effekter av AAP og sanksjoner som følge av dette. Videre er mange tjenester og ytelser rettighetsfestede, og det er ikke mulig å evaluere dem med eksperimentelle metoder på grunn av manglende mulighet for å ha kontrollgrupper som ikke mottar tjenester eller ytelser.

Etisk har NAV eller forskere NAV samarbeider med erfart innvendinger mot at kontrollgruppen fratas et tiltak som fagpersoner tror er nyttig. Til sammenlikning er tanken om at brukerne potensielt utsettes for ineffektive eller skadelige tiltak ved å holde fast ved status quo, langt mindre utbredt.

Politisk utfordringer inkluderer manglende forståelse for usikkerhet forbundet med effekten av et virkemiddel; manglende vilje til gradvis utrulling; manglende forståelse for nødvendigheten av flerårige forsøk for å måle effekt; og manglende vilje til å vente på forskningsresultater før et virkemiddel iverksettes i stor skala.

Datautfordringer: For RCTer kan tidkrevende prosesser for å skaffe tilgang til data føre til at studier ikke blir gjennomført eller at resultatene kommer for sent til å kunne påvirke politiske beslutninger. For kvasiexperimentelle studier kan mangel på data hindre at viktige evalueringer blir gjennomført. Videre kan kvaliteten på registerdata være en utfordring, både feilregistreringer eller manglende verdier.

Praktiske utfordringer inkluderer at RCTer er tid- og ressurskrevende. For kvasiexperimentelle studier er en utfordring at innføringen av virkemidler, med henblikk på hvor, når og hvem målgruppen er, ikke alltid dokumenteres godt nok til at kvasiexperimentelle evalueringer kan gjennomføres i ettertid.

Muligheter

Interne faktorer som kan følge til rette for bedre effektevalueringer

- **Gjøre beslutningstakerne oppmerksomme** på mulighetene for gradvis utrulling og skapt aksept for systematisk forskjellsbehandling i en avgrenset periode
- Heve intern **juridisk kompetanse** for å styrke forståelse av regelverket og utnytte handlingsrommet som regelverket gir
- Heve **metodekompetanse** for å unngå sentrale fallgruver ved RCT og forstå forholdene som begrenser mulighetene for grundige effektevalueringer

Eksterne faktorer som kan legge til rette for bedre effektevalueringer

- **Informasjons- og holdningsarbeid** som skaper økt politisk aksept for usikkerheten forbundet med effekten av et virkemiddel, og vilje til å tillate tidsavgrenset forskjellsbehandling, enten gjennom RCTer eller gradvis utrulling eller pilotering av virkemidler
- **Bedre dokumentasjon** av hvordan virkemidler iverksettes og definerte kriterier/terskler for hvem som får dra nytte av virkemiddelet. For eksempel finnes det potensiale for å gjennomføre kvasiexperimentelle evalueringer med regression discontinuity design dersom det er klart definerte kriterier eller terskler for å velge ut de som vil motta virkemiddelet, slik at det er mulig å sammenligne effekten av å motta virkemiddelet med effekten av å ikke motta det, basert på hvordan utfallet endres i nærheten av terskelen.
- **Bygge systemer** for gradvis innføring av virkemidler for å tillate læring om effekter
- **Tålmodighet** med å vente på permanent, nasjonal innføring av virkemidler før kunnskap om effekter foreligger
- **Veiledning** i hvordan man kan planlegge, forankre og gjennomføre en RCT. Aspekter som kan inkluderes i dette er informasjon om
 - Lovhjemmel og juridisk grunnlag for å kunne gjøre forsøk.

- Råd til hvilke etiske hensyn man må være bevisst på, og hvordan de kan ivaretas i forsøket
- Hvordan man legitimerer forskjellsbehandling overfor sentrale beslutningstakere og brukere
- Praktiske anbefalinger til planlegging av en RCT, inkludert fallgruver
- Forhold som avgjør om et eksperiment bør stoppes

Vedlegg 5: Norads erfaringer med randomiserte og kvasieksperimentelle studier

Organisering, publiseringskanaler og kvalitetssikring

Det arbeides på tvers av seksjoner i Norad for å motivere bruk av RCT-er og kvasieksperimentelle studier (heretter omtalt samlet som «effektevalueringer») i utlysningstekster på temaer der kunnskapshull er identifisert og hvor det foreligger økonomisk handlingsrom for å gjennomføre effektevalueringer. Videre oppfordres aktører som mottar tilskudd fra Norad til hensiktsmessig bruk av effektevalueringer. Norad har en Avdeling for kunnskap der enkelte medarbeidere fungerer som metoderådgivere for spesifikke utlysninger hvor effektevalueringer kan være aktuelt.

I utgangspunktet har Norad begrenset innflytelse på den vitenskapelige utformingen av en studie. Imidlertid kan Norad påvirke gjennom godkjenning av terms of reference før partnere anskaffer konsulentbistand i forbindelse med en effektevaluering, samt tett dialog med partnere for å oppnå høy kvalitet på finansierte studier.

Effektevalueringer blir sammenfattet som eksternt tilgjengelige rapporter. Noen ganger publiseres de på Norad.no, mens som oftest blir de publisert hos partneren som har gjennomført evalueringen. Det er sannsynlig at noen publiseres i fagfelleverderte tidsskrifter, men en fullstendig oversikt lar seg ikke oppdrive.

Omfang og eksempler på evaluerte virkemidler

Norad har de siste fem årene anslagsvis finansiert 1-5 randomiserte studier og 1-5 kvasieksperimentelle studier. Disse inkluderer evalueringer der Norad har stilt krav om effektevaluering, eller der Norad har tatt initiativ eller vært aktiv i prosessen, samt evalueringer som har blitt gjennomført i forbindelse med en bilateral tilskuddsavtale.

Norad har medfinansiert flere randomiserte studier av resultatbasert finansiering i helsetjenesten. Videre har Norad finansiert kvasieksperimentelle studier, som effektevaluering av Norway India Partnership Initiative for bedre mor-barn helse ved matching av tiltaks- og kontrollområder, og en pågående studie om avskaffelse av skadelige skikker og kjønnslemlestelse i Etiopia. Det planlegges for øyeblikket for å finansiere en RCT i regi av Verdens matvareprogram om effekter av lokalprodusert skolemat i Malawi.

Utfordringer

Studier som Norad finansierer foregår i lav- og mellominntektsland, og dette innebærer spesielle hensyn til **ressursutfordringer, etikk og datagrunnlag**.

Metodologisk kan det være spesielt utfordrende å sikre at randomiseringen foregår korrekt i lavinntektssettinger. Videre er det vanskelig å forhindre lekkasje mellom tiltaks- og kontrollgrupper, noe som gjør det utfordrende å estimere den kausale effekten av tiltaket som evalueres.

For Norad er en sentral **juridisk** utfordring ordningsregelverkene som regulerer forvaltningen av tilskudd. Ordningsregelverkene er UDs tolkning av økonomiregelverket for spesifikke tilskuddsordninger, og er strengt tatt ikke juridiske dokumenter. De definerer målgruppe, prosjektkriterier, tilskuddsmottakere og evalueringskrav, men nevner ikke eksplisitt effektevalueringer eller randomiserte studier. Dette har ført til

vanskeligheter med å finansiere slike studier med tilskuddsmidler. Prop. 1 S (2022–2023) gir imidlertid en overordnet føring om at tilskuddsmidler kan brukes til effektevaluering og følgeforskning³²: «For å bidra til mer kunnskapsbasert gjennomføring av bistandsprogrammer, kan effektevaluering og følgeforskning integreres i aktuelle programmer.»

Forskningsetisk fremmes det iblant argumenter mot å gjennomføre effektevalueringer eller annen forskjellsbehandling som tillater læring om effekt, fordi man da «eksperimenterer» med folks livsforhold. I forlengelsen av dette vurderer noen det uetisk å ha en kontrollgruppe som ikke får dra nytte av tiltaket som testes, spesielt dersom studien rekrutterer sårbare grupper som personer i fattigdom eller personer som har vært utsatt for konflikt og vold.

Politisk har det tidligere vært mindre støtte for å bruke penger på effektevalueringer, og spesielt RCTer, i bistanden, men dette har endret seg. Evalueringer kan bli ansett å innebære bruk av konsulenter og «administrasjonskostnader», som kan være politisk upopulært. I mottakerland kan det, som i Norge og i bistandsforvaltningen, være en utfordring at ikke alltid resultatene av en effektevaluering faktisk blir brukt til å informere videre opp- eller nedskalering av virkemiddelet. Videre kan det være tilbakeholdenhet overfor å gjennomføre piloter med effektevaluering fremfor fullskala implementering, foreligge politiske preferanser for hvem som bør motta tiltaket, og være politisk utfordrende å ikke la alle i et gitt område motta et tiltak.

Datagrunnlaget er som regel spørreundersøkelser hvor dataene samles inn i forbindelse med studiene, noe som kan være kostbart, innebære variabel datakvalitet og ha høy risiko for frafall. Eksisterende surveydata (som DHS) anvendes for å planlegge og identifisere målgrupper. Sammenlignet med Norge er registerdata i Norads partnerland ofte mangelfulle eller ikke-eksisterende, noe som nødvendiggjør at forskere samler inn data selv.

Andre utfordringer inkluderer:

- Fokus på **kortsiktige aktivitetsmål**, som antall skolebøker distribuert, antall skolebarn som fikk skolemat eller antall vaksiner, kan ta oppmerksomheten vekk fra effektevalueringer som kan bidra til å bedre bistandsresultater men over tid
- Innvendinger mot den **alternative kostnaden** av å bruke midler på effektevalueringer som ellers kunne blitt brukt til å nå flere med tiltaket
- Mangelfull **kapasitet** og **kompetanse i mottakerland**, eller kjennskap til den (når den eksisterer), kan gjøre det vanskelig å forankre studiene og sikre lokalt eierskap til dem
- Manglende **kapasitet og kompetanse blant bistandspartnere**, deriblant norske sivilsamfunnsorganisasjoner, om gjennomføring av effektevalueringer, samt tidkrevende prosesser for etiske godkjenninger, kan gjøre det vanskelig å gjennomføre slike studier
- Manglende **insentiver** for tilskuddsmottakere til å gjennomføre effektevalueringer på grunn av frykt for konsekvenser av negative funn
- **Budsjett- og programsykluser** som gjør det vanskelig å garantere finansiering for langsiktige studier

³² <https://www.regjeringen.no/no/dokumenter/prop.-1-s-20222023/id2931090>

- **Usikkerhet rundt finansieringen** av hele effektevalueringen kan være avskrekkende, da alle avtaler inkluderer et forbehold om midler fra statsbudsjettet

Muligheter

Norad har forsøkt å tilrettelegge for randomiserte og kvasiekperimentelle studier gjennom blant annet

- **Øke regelverkets handlingsrom:** Norad, i samarbeid med Utenriksdepartementet, jobber for sikre at regelverk og føringer åpner for at effektevalueringer kan bakes inn i tilskuddsavtaler der det er hensiktsmessig, og på den måten sørge for at effektevaluering er en naturlig del av planlegging og implementering av virkemidler.
- **Styrket datainnsamling:** Norad jobber blant annet aktivt med å fremme innsamling og bruk av geodata blant sivilt samfunn, for å tilrettelegge for studier som utnytter geografisk nedslagsfelt for å studere effekter på geolokaliserede utfallsvariabler («geospatial impact evaluations»), f.eks. satellittbilder av regnskog.
- **Påvirkningsarbeid:** Partnere oppfordres til å prioritere effektevalueringer av tiltak hvor kunnskapsgrunnlaget er mangelfullt. For eksempel har Norad en pågående prosess med Verdens matvareprogram om gjennomføring av en RCT for å tette kunnskapshull om lokalprodusert skolemat i Malawi.
- **Kompetanseheving gjennom forskningsinkubatorer:** Norad gjennomførte i desember 2022 en forskningsinkubator, i samarbeid med Development Learning Lab på Christian Michelsens Institutt og 3ie, om planlegging og design av effektevalueringer, med deltakelse fra sivilsamfunnsorganisasjoner som er aktive i bistandsfeltet.³³ Flere slike inkubatorer planlegges fremover.

Sivilt samfunn har under en innspillsrunde pekt på at det ville vært nyttig om Norad kunne ha tilbudt veiledning og bistand fra erfarne forskere; fasilitert samarbeid mellom Norads partnere og forskningsinstitusjoner, inkludert utvikling av standardiserte maler; tilbudt opplæring og støtte i land der Norads partnere er tilstede; og etablert egne finansieringsmekanismer for effektevalueringer.

³³Invitation to Impact-Evaluation incubator, 5-7 December 2022. <https://www.norad.no/en/front/about-norad/news/2022/invitation-to-impact-evaluation-incubator-5-7-december-2022/>

Vedlegg 6: Skatteetatens erfaringer med randomiserte og kvasiekperimentelle studier

Organisering, publiseringskanaler og kvalitetssikring

I Skatteetaten har kjernedivisjonene faglig ansvar for virkemiddelbruken, mens analyseteamet bistår kjernedivisjonen med den metodiske tilnærmingen til gjennomføring av analyser, herunder effektmålinger. Analyseteamet består av analysemiljøer i kjernedivisjonene og en fellesfunksjon som har det koordinerende ansvaret.

Analyseteamets disponering av ressurser styres av en sentral analyseplan, som er del av etatsstyringen. Etatsledelsen beslutter strategiske føringer for analyseplanen, som bidrar til god koordinering og prioritering av analyseressursene. Nytte, kostnader, metodisk gjennomførbarhet og juridiske problemstillinger utgjør grunnlaget for å prioritere analyseressursene. Over tid har samarbeidet mellom analysemiljøet og kjernedivisjonene endret seg fra en bestiller-utfører-modell til mer kontinuerlig og tverrfaglig samarbeid. Dette har fremmet flere analyser som kan omsettes i endret praksis og mer effekt virkemiddelbruk.

Analyser formidles gjennom interne rapporter, presentasjoner, vitenskapelige artikler og Skatteetatens analysenytt³⁴

Omfang og eksempler på evaluerte virkemidler

Skatteetaten har de siste fem årene gjennomført en rekke randomiserte (>15) og kvasiekperimentelle (anslagsvis 6-15) studier. Skatteetaten evaluerer virkemidler som skal øke måloppnåelsen slik at skatte- og opplysningspliktige etterlever reglene og har tillit til etaten. Disse kan grovt inndeles i proaktive og reaktive virkemidler. Eksempler på proaktive virkemidler er brev, ringerobot, dulting i den nye skattemeldingen, og medieoppslag.

- RCTer av meldingskontroll blant personlig skattepliktige har vist tydelig effekt i form av beløp som avdekkes, men også på skattepliktiges atferd i påfølgende perioder. Videre ferdigstilles en studie om effekter av meldingskontroller blant nyregistrerte MVA-virksomheter.
- Dulting brukes i veiledning mens skattemeldingen fylles ut, som er tilpasset den konkrete skattepliktiges forhold, og som trigges i sanntid basert på informasjonen den skattepliktige fyller inn. Randomisert tilordning av slike dulter viser at dultene fører til bedre etterlevelse ved at skattepliktige foretar endringer i egenfastsettingen av alminnelig inntekt.
- Veiledning med brev er tiltaket som har gjennomgått flest effektmålinger, og hvor det med størst sikkerhet kan fastslås at bruken av virkemiddelet har en kausal effekt på etterlevelsen.
- For ringerobot til purring mot skattepliktige er det identifisert positiver effekt på innlevering av skattemelding blant enkeltpersonforetak og AS, men få effektmålinger er gjennomført.
-

³⁴ <https://www.skatteetaten.no/om-skatteetaten/analyse-og-rapporter/analysenytt/>.

- For medieoppslag har en effektmåling på rapportering av kryptovaluta blitt gjennomført, som viste sprikende resultater og understreket behovet for flere effektmålinger.

Skatteetaten har også anvendt flere kvasieksperimentelle tilnærminger. For eksempel har effekten av forhåndsutfyllt informasjon i selvangivelsen vært forsøkt evaluert ved å sammenligne andelen skattepliktige som oppgir riktig gevinst og tap på selvangivelsen i årene før forhåndsutfylling, med andelen som oppgir riktig beløp i 2011. Et annet virkemiddel, bruk av risikoscore fra en fradragmodell for å plukke skattepliktige ut til kontroll, har vært evaluert med metoden Regression Discontinuity Design. Her utnyttet det faktum at skattepliktige som ligger over og under scoregrensen for kontroll vil være tilnærmet like, slik at det også er tilnærmet tilfeldig hvem av disse som tas ut til kontroll. Samme metode har vært anvendt for å måle effekten av kontroll og tilhørende tilleggsskatt hos skattepliktige som har redusert på forhåndsutfyllt lønn ved levering av skattemeldingen, siden lønnsk kontrollene var satt opp med bestemte regler og beløpsgrenser.

Videre har difference-in-difference-analyser blitt kombinert med ulike matching-teknikker for å evaluere effekter av ordinære, risikobaserte kontroller. Selv om matching bidrar til at tiltaks- og referansegruppen blir tilnærmet like på observerbare kjennetegn, så anses konklusjonene som usikre blant annet fordi konklusjoner fra matching-studier og randomiserte studier har trukket i ulike retninger.

Utfordringer

En **metodologisk utfordring** er statistisk utsagnskraft. For eksempel kan begrensede kontrollressurser redusere muligheten til å trekke pålitelige konklusjoner av effektevalueringer av kontroller. Videre kan **seleksjonseffekter** være en utfordring, selv etter tilfeldig tilordning. Kontrolluttrekk av skattepliktige er normalt risikobasert, ved at kandidater for kontroll først plukkes ut gjennom definerte risikofiltre og deretter velges basert på saksbehandlernes egne erfaringer med kontrollarbeidet. Selv når kandidater trekkes tilfeldig i første trinn, kan uobserverbare forskjeller i saksbehandlernes gjennomføring (e.g., grunnet ulik motivasjon eller kompetanse) av kontroller bidra til seleksjonseffekter.

Juridisk har Skatteetaten erfart at lovverket legger begrensninger for evaluering av sanksjonering—et virkemiddel for å motivere riktig rapportering av skatt. Skatteforvaltningslovens regler om illeggelse av tilleggsskatt innebærer at alle som oppfyller kriteriene skal ilegges en slik skatt. Under gjeldende lovverk kan derfor en randomisert evaluering av ulike former for sanksjonering ikke la seg gjennomføre. Til sammenlikning er randomisering av mykere forebyggende virkemidler, som informasjonsbrev, SMS, telefonoppringning og liknende, enklere å få tillatelse til. En utfordring er imidlertid at innhenting av opplysninger om skattepliktige som trekkes tilfeldig til tiltaksgruppen kan utløse krav om kontroll, noe som vil gjøre at hensikten med iverksetting av mykere virkemidler faller bort. Det er imidlertid avklart innenfor skatteforvaltningslovgivningen at opplysninger kan innhentes, f. eks. for tiltak som dulting, uten at de samme skattepliktige behøver å kontrolleres senere. Videre står prinsippet om likebehandling sterkt for skattekontorets saksbehandling, noe som gjør forskjellsbehandling under effektevalueringer vanskelig å gjennomføre. Etaten har et pågående arbeid for å tydeliggjøre det rettslige handlingsrommet for forskjellsbehandling.

Forskningsetisk er har etatens studier berørt flere aspekter knyttet til informert samtykke. Det kan være utfordrende for skattebetaler å la være å samtykke til studier gjennomført av Skatteetaten ut av frykt for konsekvenser.

En sentral **politisk** utfordring er at mål- og resultatstyringen har en årlig syklus og ikke er godt tilrettelagt for å styres av informasjon om effekter, som for noen virkemidler kan ta flere år før de blir synlige. I tillegg er det betydelig politisk fokus og krav til aktiviteter og årlig resultatoppnåelse, noe som kan begrense handlingsrommet for å samle informasjon om effekter over en lengre tidsperiode.

Datagrunnlag: Dataene for Skatteetatens studier stammer som regel fra etatens fagsystemer og registre. Ved enkelte studier har ikke fagsystemene tilstrekkelig fleksibilitet for at saksbehandlere kan registrere all informasjon som er relevant for studien og dette nødvendiggjør supplerende løsninger for registrering av informasjon utenfor fagsystem og datavarehus. En utfordring med spørreundersøkelser er at Skatteetaten som avsender kan påvirke deltakere til å gi bestemte svar. Videre er frafall og seleksjon av de som svarer på spørreundersøkelser også en utfordring.

Muligheter

Skatteetaten har erfart at følgende kan tilrettelegge for effektevalueringer

- Økt **prioritering** av effektevalueringer internt: I Skatteetaten har dette blant annet involvert (1) **Oppbygning av metode- og analysemiljø:** det nasjonale analyseteamet, styrket siden 2010, består i skrivende stund av mer enn 80 analytikere og prosjektledere fra de ulike divisjonenes analysemiljø, med kompetanse innen dataanalyse og maskinlæring, adferdsteori, økonometriske metoder og kausale analysedesign; (2) **Etablering av Program for effektmåling:** Lansert av etatsledelsen i 2020 for å øke Skatteetatens kunnskap om sammenhengen mellom virkemidler og effekter, og sikre felles prioriteringer og forsvarlig faglig metodebruk.
- Bygging av **kultur** for effektevalueringer
- Integrere utvikling av **systemstøtte** for effektevalueringer i moderniseringsløp: Kostnaden ved å utvikle og kjøpe funksjonalitet mens moderniseringsprosjektene pågår er antatt lavere enn å tilpasse dette i ettertid.

Vedlegg 7: Utdanningsdirektoratets erfaringer med randomiserte og kvasiekperimentelle studier

Organisering, publiseringskanaler og kvalitetssikring

Utdanningsdirektoratet (Udir) er Kunnskapsdepartementets (KD) utøvende organ for barnehage, skole og fagopplæring og skal sette opplæringspolitikken ut i praksis. Det årlige tildelingsbrevet inneholder mål, prioriteringer og økonomiske rammer. KD sender også Udir oppdragsbrev om utlysning av forskningsoppdrag, inkludert effektevalueringer. Udir evaluerer ulike virkemidler og tiltak som skal bidra til økt kvalitet og utvikling av barnehage, grunnskole og videregående opplæring, for at alle barn og unge skal lære, mestre og trives. Udir har ansvar for å anskaffe og følge opp disse evalueringene. Udir erfarer at det er viktig å tydelig beskrive kunnskapsbehovet for å få et hensiktsmessig forskningsdesign. Vitenskapelige kvalitet er blant tildelingskriteriene i konkurransene og tilbudene vurderes av interne arbeidsgrupper i Udir, samt av en ekstern ekspert med relevant forsknings- og metodekompetanse. Udir formidler fra effektevalueringer på ulike måter: egne og oppdragstakers nettsider, forskningsrapporter, vitenskapelige artikler, populærvitenskapelige artikler, webinarer og podcaster.

Bruken av randomiserte (RCT) og kvasiekperimentelle studier er i tråd med idealet om skrittvis reform i opplæringssektoren, som innebærer at et tiltak utprøves i liten skala før de eventuelt oppskaleres og blir virkemidler i opplæringspolitikken. Viktige steg er systematisering av kunnskapsgrunnlaget og utredning og utvikling av lovende modeller, før disse så prøves ut og forskes på i en randomisert studie med ledsagende implementeringsstudie, eller gjennom andre tilnærminger til å evaluere effekt.

Konteksten for effektstudier i opplæringssektoren er viktig for å forstå i hvilken grad ulike typer effektevalueringer kan la seg gjennomføre. Det rapporteres om prosjekttrøtthet i sektoren som følge av mange og ikke alltid koordinerte tiltak, krav og satsninger relatert til opplæring, oppvekst og barn og unge. Opplevelsen av en hektisk hverdag uten tilstrekkelige ressurser, forsterkes av ulike utfordringer og ettervirkninger av pandemi, kommunesammenslåinger og streiker. Det blir derfor spesielt viktig å tilrettelegge for god medvirkning fra relevante aktører i praksisfeltet i utforming og gjennomføring av effektevalueringer. Det kan bidra til å sikre at tiltakene som prøves ut blir hensiktsmessige og mulig å implementere, og at man skaper økt forståelse og motivasjon for deltagelse i forskningen.

Omfang og eksempler på evaluerte virkemidler

Udir har de siste 5 årene anskaffet 3 randomiserte (RCT) og 1-5 kvasiekperimentelle evalueringer. Forskningen anskaffer Udir hovedsakelig på oppdrag fra KD, ofte etter dialog om deres kunnskapsbehov. Innen opplæringssektoren utvikles det også kunnskapsgrunnlag gjennom langsiktige programmer finansiert gjennom Forskningsrådet. Her er programmet Lærertetthet og læringseffekt (*LÆREEFFEKT*) et eksempel, som er en forskningssatsing på effekter av flere lærere i skolen. I programmet er RCTer og kvasiekperimentelle studier prioritert.

Eksempler på virkemidler utprøvd med RCT anskaffet av Udir inkluderer to modeller for hensiktsmessig anvendelse av flerfaglig kompetanse i skolen (effekten av en ekstra helsepsykepleierressurs og utprøving av LOG-modellen, som er en modell for ledelse, organisering og gjennomføring av et systematisk utviklingsarbeid og tverrfaglig

samarbeid); modeller for fysisk aktivitet for elever i ungdomsskolen; og tilskuddsmodeller for utstyr i fag- og yrkesopplæringen.

Eksempler på virkemidler evaluert med ulike kvasieksperimentelle metoder (forskjell-i-forskjell metoden) inkluderer evaluering av gratis kjernetid i barnehagen i Oslo, evaluering av norm for lærertetthet og utprøving av modeller for mer inkluderende praksis.

Utfordringer

Metodologiske utfordringer for RCTer inkluderer manglende **statistisk styrke** siden virkemidler som oftest må randomiseres på skolenivå og **smitteeffekter** som følge av at praksisen som testes tas i bruk i kontrollskolene. En sentral utfordring for kvasieksperimentelle studier er valget av kontrollgruppen, som skal være så lik tiltaksgruppen som mulig. Videre kan kvaliteten på **implementering** av tiltaket påvirke effekt målet. Gjennomførte effektevalueringer, som f.eks. av LOG-modellen³⁵, understreker behovet for gode implementeringsstudier, som for eksempel kan avdekke forskjeller mellom kommuner og skoler i implementeringen av virkemiddelet som utprøves, graden av forankring og medvirkning og i hvilken grad skoler tilpasset utviklingsarbeidet til egne behov og forutsetninger.

Juridisk setter opplæringsloven med forskrifter minimumskrav til opplæringen og beskytter retten til et likeverdig tilbud. En forskjellsbehandling, enten i regi av en RCT eller andre måter å tilrettelegge for læring om effekt, kan vanskelig rettferdiggjøres dersom det medfører store fordeler/ulempes for noen grupper av elever. Dersom gjennomføringen av slike studier forutsetter avvik fra opplæringsloven med forskrifter har Udir (delegert myndighet) anledning til å godkjenne forsøket. Loven gir ikke åpning for forsøk som berører de mest grunnleggende rettighetene eleven har, og det åpnes bare for forsøk som det er rimelig grunn til å anta at totalt sett vil kunne føre til en opplæring som er bedre for elevene.

Etisk kan det være en utfordring å rettferdiggjøre at kontrollgruppen i en RCT ikke skal motta det tiltaket som prøves ut, når man antar at det vil ha sterk positiv effekt f.eks. på barn og unges trivsel eller læringsutbytte, eller at tiltaket kan ha potensielle negative konsekvenser for barn og ungdom i tiltaksgruppen. Det er en etisk utfordring å vurdere nytte og skade ved barn og unges deltakelse i forskning, ikke bare når det gjelder mulige konsekvenser, men også med tanke på involvering og tilpasning til deres sårbarhet, behov, situasjon, kompetanse og rettigheter.

Politiske utfordringer inkluderer **endrede politiske føringer som endrer forutsetningene for effektevalueringer og ønske om raske svar og manglende vilje til å vente på forskningsresultater før reformer blir iverksatt**. Et eksempel på dette er resultatet av regjeringens budsjettforhandlinger i 2017, da det ble enighet om at en norm for lærertetthet skulle innføres fra høsten 2018. Dette brøt med tanken om skrittvis reform, da regjeringen ikke ventet på resultatene fra det igangsatte *LÆREEFFEKT-programmet som undersøkte effektene av flere lærere i skolen*. I tillegg gjorde vedtaket at det var vanskelig å gjennomføre forskningen det hadde blitt bevilget penger til to år tidligere. En av de to pågående RCTene, Two teachers-prosjektet, måtte avslutte

1 ³⁵Et lag rundt eleven: LOG-modellen. <https://www.udir.no/tall-og-forskning/finn-forskning/rapporter/et-lag-rundt-eleven-en-klyngerandomisert-effektevaluering-av-log-modellen/>

utprøvingen tidligere enn planlagt, siden studien var avhengig av at tiltaksskolene hadde stabile klasser.

Andre utfordringer inkluderer

- **Ressurser:** RCTer er kostnadskreven og kan ta opp store deler av budsjettet for styrking av kunnskapsgrunnlaget gjennom for forskningsoppdrag.
- **Krevende kontekst i opplæringssektoren:** Prosjekttrøtthet, opplevd mangel på ressurser og økte utfordringer i sektoren kan gjøre det utfordrende å gjennomføre effektevalueringer. Det blir avgjørende med koordinerte satsninger, nok ressurser og god medvirkning.
- **Kunnskap og motivasjon til deltagelse:** effektevalueringer krever at aktørene i sektoren bidrar til utprøving av tiltak og ulike former for undersøkelser, både i tiltaks- og kontrollgruppen. I en kontekst med ressursmangel og prosjekttrøtthet kan det være vanskeligere å motivere til å delta i datainnsamling over tid, særlig for de som er i kontrollgruppen og ikke får nyttiggjøre seg tiltaket som prøves ut. Dette kan f.eks. føre til økt frafall, lavere svarprosent på spørreundersøkelser og/eller vanskeliggjøre samarbeid om øvrig datainnsamling. I tillegg kan manglende forståelse for hvorfor og hvordan et forsøk gjennomføres, være et hinder for deltakelse i datainnsamling.
- **Datagrunnlag:** opplæringssektoren mangler et bredt tilfang av nasjonale mål for elevenes læring og trivsel, og eksisterende undersøkelser er utformet for andre formål enn forskning. For eksempel er det få/ingen nasjonale mål for elevenes læring og trivsel på de laveste trinnene og først på 5., 8. og 9. trinn finnes nasjonale, sammenlignbare mål på elevenes læring i grunnleggende ferdigheter i lesing, regning og engelsk. Det er også et generelt behov for individdata til forskning i opplæringssektoren.

Muligheter

Udir har erfart at følgende kan tilrettelegge for effektevalueringer

- **Sikre at riktig kompetanse blir inkludert i arbeidet med effektforskning:** dette inkluderer god metodekompetanse, god kompetanse på virkemidlene som skal evalueres, kompetanse på sektoren som evalueres og juridisk kompetanse slik at lovverket blir overholdt.
- **Tett dialog med departement og organisasjoner i sektoren om tilnærminger for å møte kunnskapsbehov:** for å evaluere tilskuddsmodeller for utstyr i fag- og yrkesopplæringen la Udir i dialog med KD til rette for at loddtrekning kunne anvendes for å tildele tilskuddene. En sentral utfordring var å ivareta randomiseringsbehovet for effektforskningen, samtidig som det ble opprettholdt like muligheter for å motta tilskuddet. Dette krevde forankring og dialog med organisasjonene i sektor for å skape forståelse for RCTens behov for loddtrekning fremfor behovsprøvde vurderinger av søknadene. Skolene ble informert om at tildeling ville skje ved hjelp av loddtrekning i forbindelse med utarbeiding av søknaden. En forenklet søknadsprosedyre skulle også bidra til større aksept for framgangsmåten. Imidlertid var det vanskelig å kommunisere forutsetningene og prosedyre for tilskuddet.³⁶

³⁶Statleg utstyrslosteri for yrkesfag: – Litt som «Joker Nord». <https://www.nrk.no/vestland/vidaregaande-skular-fekk-77-millionar-i-statsstotte-ved-loddtrekking-1.15815655>

- **Langsiktige forskningssatsninger:** f.eks. prioriterer programmet LÆREEFFEKT randomiserte og kvasiekperimentelle studier.³⁷

³⁷LÆREEFFEKT – Lærertetthet og læringseffekt.

https://www.forskningsradet.no/contentassets/965669a13e7c42e98bb879bdc2ddbdb/rammedokument_for_laereeffekt.pdf

Utgitt av Folkehelseinstituttet

August 2023

Postboks 4404 Nydalen

NO-0403 Oslo

Telefon: 21 07 70 00

Rapporten kan lastes ned gratis fra

Folkehelseinstituttets nettsider

www.fhi.no