



# Måling med mening

*Om Norads langtidsstrategi og bruk av robuste målemetoder*

**Jacob Mæhle og Skage Lem**

**Veileder: Kjetil Bjorvatn**

Masterutredning i økonomi og administrasjon

Hovedprofil: Strategi og Ledelse

NORGES HANDELSHØYSKOLE

Dette selvstendige arbeidet er gjennomført som ledd i masterstudiet i økonomi- og administrasjon ved Norges Handelshøyskole og godkjent som sådan. Godkjenningen innebærer ikke at Høyskolen eller sensorer inntår for de metoder som er anvendt, resultater som er fremkommet eller konklusjoner som er trukket i arbeidet.



# Forord

*Norge er et land i verden.*

- Lars Korvald

Lars Korvald er ingen mann å kimse av, og med denne oppgaven trer vi altså ut av Norges Handelshøyskole og inn i verden. Å avslutte fem år ved NHH med en oppgave som illustrerer viktigheten av økonomifaget i den virkelige verden, gir oss stor glede. Å få bruke skolegangens siste halvår på et dypdykk i norsk bistandspolitikk, bærekraft og økonomisk metode har vært både meningsfullt og givende. Vårt håp er at oppgaven kan fungere som et konstruktivt bidrag i debatten om hvordan vi kan gi mest mulig, best mulig.

Vi ønsker å rette en stor takk til vår veileder Kjetil Bjorvatn. Hans engasjement, ideer, innsats og motivasjon har ikke bare gjort oppgaven betraktelig bedre, men også langt mer inspirerende å skrive.

Videre ønsker vi å takke Norad for stor samarbeidsvilje og åpenhet. Sammen med ansatte i UD, CMI, UIB og NHH har de bidratt til å stake ut kursen for oppgaven og gitt verdifulle innspill underveis. Vi vil også takke Petter Krogh Nilsen, Peter Syse og Syver Hanken for deres hjelp og tålmodighet. Hverken oppgaven eller livet for øvrig ville vært det samme uten dem.

Aller sist vil vi takke Roger Valhammer, Bergens utelivsbransje og Kjelleren ved NHH for lesero og humane leggetider. Aldri før har det vært enklere å sitte på hjemmekontor med maks to stykker. Bistand kommer i mang en form!

Norges Handelshøyskole

Bergen, mai 2021



Jacob Mæhle



Skage Lem

## Sammendrag

Hvert år setter Norge av flere milliarder til bistand, og nærmere halvparten forvaltes av direktoratet Norad. Disse midlene benyttes til å finansiere og evaluere arbeidet til tredjepartsorganisasjoners bistandsprosjekter. Med sitt overordnede mål om å *sette penger i arbeid for å oppnå resultater* har Norad likevel over mange år motatt kritikk for lite robuste metoder og følgelig upålitelige resultater. I senere år har direktoratet tatt grep, og i 2021 ble en langsiktig strategi lansert av Norad for arbeidet deres frem mot 2030. Strategien bygger på Bærekraftsmålene og en kunnskapsbasert tilnærming der *fakta har makt*. Denne oppgaven foretar dermed en gjennomgang av dagens metoder for å undersøke hvorvidt Norad i tilstrekkelig grad har tatt høyde for tidligere kritikk i deres langsiktige strategi og virke.

I gjennomgangen av dagens metoder fremkommer det at Norad i stor grad benytter seg av kvalitative metoder. De kvantitative metodene er fortsatt lite robuste og dagens kriterier for måloppnåelse gjør at Norad ikke er i stand til å si hvorvidt bistandsmidler har effekt eller settes til god bruk. Kvantitative metoder som RCT eller Difference-in-Difference vil kunne skape mer pålitelige resultater og dette illustreres i oppgaven gjennom eksempler hvor det tydeliggjøres hvordan Norads nåværende måling kan slå feil.

Opgaven argumenterer videre for hvordan robust resultatmåling kan sees på som en investering fremfor en kostnad. Ved rigorøs måling kan Norad oppnå pålitelige resultater som kan styrke direktoratets legitimitet og tillit. Slike oppnåelser vil være gunstig for fortsatt budsjettvekst i takt med BNI. Norad burde også implementere robuste målemetoder i deres langsiktige strategi og formidle dette til samtlige berørte parter og partnere.

Ved krav om robuste målemetoder bør Norad prioritere eksperimentell og presedensdannende bistand, samt de store og veletablerte organisasjonene. Oppgaven konkluderer med at Norad fortsatt ikke har grunnlag for å stole på dagens resultater og at det derfor er avgjørende med en ytterligere omstilling både i deres strategi og virke. Dette begrunnes i overnevnte kritikk, dagens metoder og verdi av robuste metoder for Norad.

**Nøkkelord** – Norad, Bistand, NHH, Masteroppgave, Resultatmåling, Difference-in-Difference, RCT, Strategikart

# Innhold

<b>1</b>	<b>Innledning</b>	<b>1</b>
1.1	Motivasjon bak oppgaven . . . . .	3
1.2	Problemstilling . . . . .	4
1.3	Oppgavens struktur . . . . .	4
1.4	Forutsetninger og avgrensninger . . . . .	5
<b>2</b>	<b>Bakgrunn</b>	<b>7</b>
2.1	Bistandsproblematikken i dag . . . . .	7
2.1.1	Uendelig behov - begrensede midler . . . . .	8
2.2	Norads virksomhet . . . . .	9
2.3	Dagens målemetode for norsk bistand . . . . .	10
2.3.1	SOS Barnebyers prosjekt for å forbedre skoletilbud for barn utsatt for omsorgssvikt . . . . .	13
2.4	Strategikart . . . . .	14
<b>3</b>	<b>Litteratur og forskning</b>	<b>16</b>
3.1	Resultatmåling og den nye gullstandarden . . . . .	16
3.1.1	Hvorfor måler vi? . . . . .	16
3.1.2	RCT . . . . .	17
3.2	RCT-bevegelsen . . . . .	18
3.3	Difference-in-Difference . . . . .	20
<b>4</b>	<b>Tidligere evalueringer og nåværende strategi</b>	<b>22</b>
4.1	Kritikk av Norads metode . . . . .	22
4.2	Norads strategi . . . . .	24
4.3	Oppsummering . . . . .	25
<b>5</b>	<b>Analyse av ungdomsklubber i Uganda</b>	<b>27</b>
5.1	Innledning om prosjektet . . . . .	27
5.2	Metodikk . . . . .	29
5.2.1	Empowerment - tabell og grafisk fremstilling . . . . .	29
5.2.2	Bruk av kondom - tabell og grafisk fremstilling . . . . .	32
5.3	Funn og resultater . . . . .	34
5.3.1	Empowerment - regresjon . . . . .	34
5.3.2	Bruk av kondom - regresjon . . . . .	36
5.4	Begrensninger for inferens fra prosjektet . . . . .	37
<b>6</b>	<b>Diskusjon</b>	<b>39</b>
6.1	Implikasjoner for Norads resultater . . . . .	39
6.2	Kvalitativt eller kvantitativt - hvilken metode bør benyttes? . . . . .	41
6.3	Oppsummering og anbefalinger . . . . .	43
<b>7</b>	<b>Implikasjoner og implementering i Norads strategi</b>	<b>44</b>
7.1	Strategikart . . . . .	44
7.1.1	Betydning for Norad og sirkularitet . . . . .	45
7.2	Implementering av robuste metoder i Norads strategi og virke . . . . .	47

7.3	Når bør rigorøs kvantitativ metode anvendes? . . . . .	49
7.4	Kan norske bedrifter lære noe av våre funn? . . . . .	52
7.5	Oppsummering og delkonklusjon . . . . .	53
<b>8</b>	<b>Konklusjon</b>	<b>54</b>
8.1	Robusthetsanalyse . . . . .	56
	<b>Referanser</b>	<b>58</b>
	<b>Appendiks</b>	<b>63</b>

## Figurliste

1.1	Millenium Villages Project Kenya: Andel som eier mobiltelefon, per husholdning . . . . .	2
2.1	Resultatkjede - Results Management in Norwegian Development Cooperation (Norad, 2008) . . . . .	12
2.2	Resultrammeverk (Norad, 2020b) . . . . .	13
2.3	Resultrammeverk for SOS Barnebyer i Malawi (Norad, 2020a) . . . . .	14
3.1	Publisert bruk av RCT (Cameron et al., 2015) . . . . .	18
3.2	Utvikling i 6 og 8 distrikt (Angrist og Pischke, 2014) . . . . .	21
5.1	Barplot av empowerment i kontrollgruppe og behandlingsgruppe . . . . .	30
5.2	Empowerment indeks - grafisk fremstilling . . . . .	31
5.3	Barplot av prosent som alltid bruker kondom i kontrollgruppe og behandlingsgruppe . . . . .	32
5.4	Kondom - grafisk fremstilling . . . . .	33
7.1	Sirkulært strategikart . . . . .	46
A0.1	Results Monitoring Framework Template . . . . .	63
A0.2	SOS Barnebyer - utvidet resultatrammeverk . . . . .	63

## Tabelliste

5.1	Regresjon av empowerment index . . . . .	35
5.2	Regresjon av prosent som alltid bruker kondom . . . . .	37

## Forkortelser

AEJ	American Economic Journal
BNI	Brutto Nasjonalinntekt
BRAC	Bangladesh Rural Advancement Committee
CMI	Chr. Michelsen Institutt
DAC	Development Assistance Committee
DiD	Difference-in-Difference
ELA	Empowerment and Livelihood for Adolescents
ESG	Environmental, Social and Governance
FN	Forente Nasjoner
KOM	Klima- og Miljødepartementet
KPI	Key Performance Indicator
MVP	Millenium Villages Project
ODA	Official Development Assistance
OECD	Organisation for Economic Co-operation and Development
RAM	Resource Allocation Model
RCT	Randomized Controlled Trial
UD	Utenriksdepartementet
UNDP	United Nations Development Programme



# 1 Innledning

*The most useful thing a development economist can do in this environment is stand up for hard evidence.*

- Abhijit Banerjee (Banerjee, 2005)

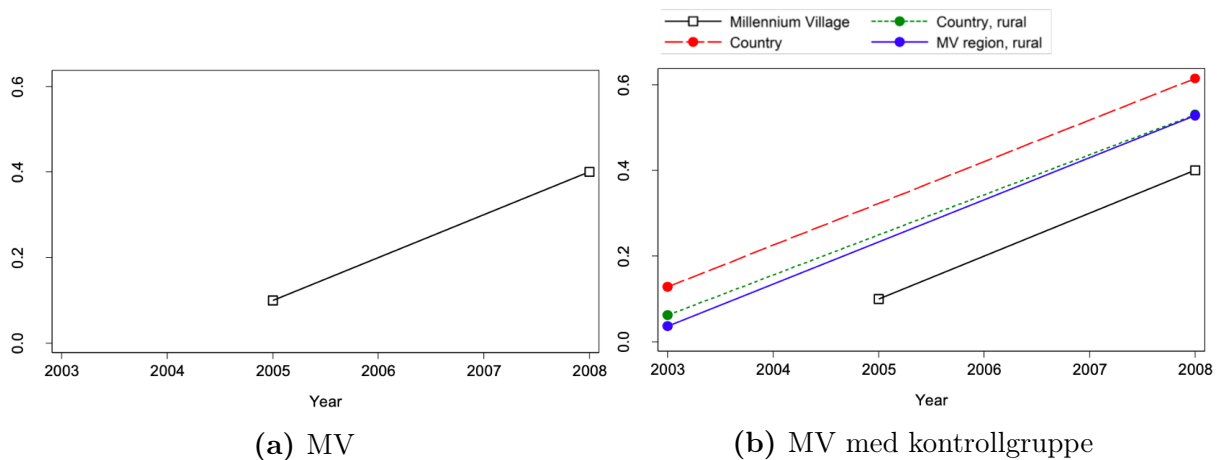
FNs Tusenårs mål ble lansert i år 2000 som et sett med ambisiøse mål for global utvikling (United Nations, 2015a). Vanskelige oppgaver krever ofte nye løsninger, og i 2005 ble Millennium Villages Project, heretter MVP, lansert som et svært omfattende og eksperimentelt prosjekt (Devex, 2020). Arkitekten bak prosjektet, Jeffrey Sachs, hadde allerede gjort seg kjent som en forkjemper for såkalte “big push” (Sachs, 1999) - forsøk på å løfte hele land, områder eller grupper ut av det som omtales som “fattigdomsfellen” og inn i bærekraftig vekst ved å innrette intensiv bistand fra flere kanter samtidig. Prosjektet skulle inneholde alt fra infrastruktur og helsetilbud til gjødsel og utdanning, og Jeffrey Sachs samarbeid med UNDP (United Nations Development Programme) ble fulgt nøye av store deler av verden.

Allerede før noen form for evaluering ble publisert, ble prosjektet utvidet til flere regioner. Dette tyder på at initiativtakerne tidlig regnet prosjektet som en suksess. Samtidig møtte det kritikk for å ligne for mye på tidligere prosjekter som har mislyktes i å skape vedvarende utvikling (Clemens og Demombynes, 2013).

Tankegangen rundt “big push” har også vært gjenstand for kritikk både før og etter MVP ble iverksatt (Easterly og Sachs, 2006). Like fullt var det ikke selve bistandsprosjektet som ble prosjektets akilleshæl, men heller den metodiske fremferden og evalueringen av arbeidet. Etter hvert som prosjektets første evalueringer kom ut, ble det møtt med kritikk for manglende forutsetninger for å trekke robuste konklusjoner fra store deler av arbeidet (Clemens og Demombynes, 2013). Årsakene var blant annet manglende baseline-data og at områdene som ble valgt ut som deltakerområder ble nettopp dét – valgt ut bevisst fremfor ved randomisert metode. På enkelte områder ble konsekvensene av dette særlig synlige. Et eksempel på dette var rapporter om redusert dødelighet hos barn under 5 år, som viste seg å ikke stemme (Bump et al., 2012). Den statistiske usikkerheten illustreres

likevel kanskje aller best ved et enkelt eksempel om kontrollgrupper og mobiltelefonbruk. Et av resultatene som ble løftet frem fra MVP-landsbyene, viste solid økning i eierskap av mobiltelefoner - et tegn på både bedre råd og infrastruktur. Det viste seg imidlertid, som man kan se i Figur 1.1, at dersom man så økningen relativt til kontrollgrupper økte mobiltelefonbruk i samme takt i MVP-landsbyene som i landet for øvrig – ikke raskere, som det først ble hevdet (Clemens og Demombynes, 2013).

**Figur 1.1:** Millenium Villages Project  
Kenya: Andel som eier mobiltelefon, per husholdning



MVP har i etterkant blitt stående som et symbol på betydningen av rigorøs måling og riktig metodebruk for evaluering av bistand. Jeffrey Sachs og Millennium Villages kan ha skapt større suksess med prosjektet sitt enn de har blitt kreditert for, men det er vanskelig å si det helt sikkert grunnet målemetodene deres. Det kan nemlig like fort vise seg at tilskuddene ikke har gitt den lovede effekten i det hele tatt. Et prosjekt med ambisjoner om å endre den globale bistanden, endte i stedet som et statistisk skrekkeeksempel. I dag tjener hele prosjektet kanskje aller best som eksempel på hvor viktig det er å benytte seg av gode målemetoder. Alternativet til god måling kan nemlig bli tvil om hvorvidt man i det hele tatt har gjort noe godt. Dermed danner denne historien et godt bakteppe for oppgavens tematikk og mål.

## 1.1 Motivasjon bak oppgaven

Norsk bistand og bistand generelt er komplekst arbeid, sammensatt av en rekke aktører, ledd, partnere og prosjekter, ofte i flere ulike land samtidig. Bistand må videre ofte gis i vanskelige omgivelser og krever dermed også gjerne varsomme fremgangsmåter og særskilte hensyn for å bistå mennesker i allerede sårbare situasjoner. Det er med andre ord mange ulike hensyn som må tas ved utformingen av et bistandsprosjekt. En forutsetning for at bistandsmidler bør innvilges til et spesifikt prosjekt, er likevel at prosjektet har en positiv effekt. Å måle effekt av et slikt prosjekt kan være vanskelig. At det er vanskelig, gjør det imidlertid ikke mindre viktig. Ved fravær av robust måling risikeres ineffektiv bruk av bistandsmidler og liknende eksempler som nevnt innledningsvis. Gjennomgang av Norads arbeid indikerer imidlertid at vi ikke alltid vet hvorvidt norske bistandsmidler har effekt. Årsaken er at man per i dag ikke er gode nok på måling før, underveis og etter bistandsprosjektet utføres. Slik manglende kvantitativ måling gjør det vanskelig å bevise kausale effekter av midler, og man kan følgelig gå glipp av avgjørende effekter som burde ha blitt fanget opp. En nærliggende hypotese vil være at manglende dokumenterbarhet kan utgjøre flere trusler for norsk utviklingshjelp. Dersom man ikke vet om bistanden virker, kan det i det hele tatt fastslås at den ikke virker negativt? Er det forsvarlig å følge et prosentmål om bistand - slik Norge gjør - dersom man ikke vet at midlene har effekt? Hva vil det på sikt kunne bety for den allmenne legitimiteten og oppslutningen rundt bevilgninger til bistand dersom myndighetene ikke kan legge frem dokumenterbar effekt? Vårt formål er ikke primært å drøfte disse spørsmålene, men de melder seg naturlig som argumenter for hvorfor man stadig burde være opptatt av hvordan man kan finne best mulig bistandsprosjekter eller -metoder og hvordan man kan dokumentere effekten av disse. Norad har selv et bevisst forhold til problematikken og fremhever i sin tiårsstrategi for 2030 (Norad, 2021b) flere satsninger for utviklingsarbeidet, inkludert å "sette penger i arbeid på en mer strategisk måte" og å "styrke og systematisere utvikling, deling og bruk av kunnskap i Norad". Strategien opererer med et overordnet fokus på å bidra til FNs Bærekraftsmål (FN, 2021) med en kunnskapsmessig tilnærming der "*fakta har makta*". Dette vitner om at Norad ønsker å imøtekomme kritikken. Like fullt vil det være sentralt å sette spørsmålsteget ved hvorvidt de er i stand til dette med dagens metoder og strategi. Dersom Norad har et underliggende problem bør ikke strategien kun fungere som en

diagnose av sykdommen, men også som en kur for veien videre. Derfor er det viktig med en strategi som i tilstrekkelig grad peker på Norads svakheter, samtidig som den skisserer en konkret løsning for fremtiden. Norads målemetode i dag og deres plan for videre utvikling står dermed sentralt i denne oppgaven. Til slutt må det også nevnes at vi oppfatter rigorøst bistandsarbeid som særdeles viktig både for Norads legitimitet og for å gjennomføre effektivt bistandsarbeid som hjelper flest mulig, best mulig. Vi ønsker derfor ikke å argumentere mot norsk bistand i denne oppgaven, men snarere bidra til bedre bistandsarbeid gjennom mer effektive metoder bedre resultater på lang sikt.

## 1.2 Problemstilling

Som vist i eksempelet innledningsvis er det mulig, selv for svært anerkjente forskere, å begå metodiske feil med store konsekvenser. Bistandssektoren har gjennomgående fått mye kritikk i senere år for hvorvidt midler kommer frem, blir satt til god bruk og faktisk medfører langsiktig effekt (Easterly og Pfutze, 2008). Norad har heller ikke vært skjermnet for denne kritikken, men direktoratet har i senere tid tatt flere grep. Senest i år forankret de kunnskapsbasert arbeid i sin nylig fremlagte langtidsstrategi. Vi ønsker derfor å undersøke hvordan norsk bistand fungerer, hvilke metoder som legges til grunn for arbeidet og om det finnes et presserende behov for mer robuste målemetoder. Videre ønsker vi å undersøke hvorvidt det er rom og behov for ytterligere integrasjon av robuste målemetoder i Norads langsiktige strategi. Hypotesen vi arbeider med sier at Norad med dagens målemetoder risikerer å gå glipp av avgjørende data for å bevise effekt av prosjektene de finansierer. Vår problemstilling blir dermed som følger:

*Hvilke alternative målemetoder finnes til metodene Norad benytter i dag, og burde direktoratet vurdere å implementere andre målemetoder i sin langtidsstrategi og virke?*

## 1.3 Oppgavens struktur

Ettersom oppgaven dreier seg både om kvantitative målemetoder og Norads langtidsstrategi, er den strukturert for å gi hver tematikk tid og plass til grundig gjennomgang.

I Kapittel 2 blir Norads strategi frem mot 2030 presentert, og deretter blir målemetodene direktoratet benytter i dag grundig gjennomgått. I tillegg blir Strategikartet (Lien et al., 2016) gjennomgått, som sammen med målemetoder er sentrale for oppgavens tematikk. I Kapittel 3 gjennomgås Randomized Controlled Trials og Difference-in-Difference som alternative kvantitative målemetoder Norad kan benytte seg av i sine rammeverk for resultatmåling. I Kapittel 4 presenteres en rekke evalueringer av Norad som er gjennomført av eksterne fagfolk. Her dokumenteres tidligere kritikk av Norads til enhver tid gjeldende krav til metode. I lys av dette blir også Norads strategi gjennomgått ytterligere for å se på hvorvidt den svarer på kritikken som har blitt presentert. I Kapittel 5 illustreres forskjellene mellom Norads målemetoder og mer robuste metoder ved bruk av et faktisk forskningsprosjekt. Ved å bruke ekte tall fra prosjektet, av Bandiera et al. (2020), illustreres reelle utfall som kan følge av svake kvantitative målemetoder. I Kapittel 6 blir funnene gjort i det foregående kapittelet anvendt for å diskutere implikasjoner for Norad. Dette kapittelet fungerer som en diskusjon og gjennomgang av oppgavens målemetodiske poenger. Her drøftes og demonstreres potensielle feilkilder som kan oppstå ved fravær av kontrollgrupper, og deretter følger en diskusjon om hvilke målemetoder som bør benyttes i Norads virke. I Kapittel 7 følger en lengre diskusjon om implementering i Norads strategi, hvordan direktoratet bør tilnærme seg oppgavens funn og hvordan metodebruk inngår i Strategikartet som strategisk rammeverk. Til slutt blir det, i Kapittel 8, foretatt en oppsummering og konklusjon av oppgaven samt gjennomført en robusthetsanalyse for funnene og påstandene fremmet i oppgaven.

## 1.4 Forutsetninger og avgrensninger

I denne oppgaven tar vi for oss kvaliteten på målemetodene brukt for å evaluere effekten av norsk bistand. Med norsk bistand blir det i denne oppgaven gjennomgående referert til Norads andel av norsk bistandsarbeid. Oppgaven tar dermed ikke for seg effekten av midlene som eksempelvis settes i arbeid gjennom giverlandskonferanser, nødhjelp eller andre former for bistand som ikke administreres av Norad. I sammenligninger og bruk av tall fra andre land, bruker vi i denne oppgaven også utelukkende tall fra OECD-landene og deres ODA-system som ligger til grunn for hva som godkjennes som bistand i offisielle statistikker i dag (Norad, 2012). ODA står for Official Development Assistance og inneholder et offisielt sett med definisjoner, gjeldende for alle OECD-land, for hva som

regnes som bistand (OECD, 2021).

Funnene fra oppgaven baserer seg på kilder som ligger åpent tilgjengelig. Brorparten av kildene vedrørende norsk bistand kommer fra rapporter som Norad selv har publisert, enten forfattet av Norad selv eller av tredjeparter på oppdrag fra direktoratet. I tillegg benytter vi oss av relevant forskning og informasjon som ligger åpent tilgjengelig enten via internett eller som fag- og lærebøker. Kritikken som fremkommer av Norad i kapittel 4 er dermed ikke vår egen, men en gjengivelse av kritikk Norad selv har publisert på sine egne nettsider. I kapittel 5 hvor målemetoder illustreres gjennom et eget prosjekt, har vi benyttet oss av et prosjekt og datagrunnlag hentet fra American Economic Journal (Bandiera et al., 2020). Vi har valgt dette fordi AEJ er et seriøst tidsskrift hvor anvendt data ligger åpent tilgjengelig, slik at vi skal være sikre på at vi benytter oss av troverdig underlag for påstandene som fremmes.

Oppgaven tar også i bruk både kvantitativ og kvalitativ metode. For å vurdere utviklingseffekten av bistand har vi benyttet oss av kvantitativ metode og veid mer rigorøs kvantitativ metode opp mot den eksisterende bruken av kvantitativ metode i Norad. I oppgavens strategirettede deler har vi basert oss på kvalitativ metode. Dette er gjort etter våre egne vurderinger av hensiktsmessighet.

Samlet sett viser oppgaven at Norad har et betydelig forbedringspotensial når det kommer til krav til kvantitativ metode som benyttes i prosjektene de selv finansierer. Gjennom både analyse og litteraturgjennomgang viser vi at Norad ikke alltid kan fastslå den nøyaktige effekten, og trolig heller ikke om det *er* effekt, av norske bistandsmidler. Vi konkluderer med at det er helt nødvendig for Norad å ta grep knyttet til metodebruken som kreves av søkerorganisasjonene og de eksterne evalueringene av prosjekter.

Norad lanserte nylig, og som tidligere nevnt, en langsiktig strategi for eget arbeid under slagordet “Fakta har makta” (Norad, 2021b). Denne rapporten peker overordnet i samme retning som våre anbefalinger, og vi oppfatter dermed at Norad deler vårt syn på hvilken retning arbeidet med målemetoder bør gå. Det betyr derimot ikke at vi tar for gitt at Norad deler vårt syn på hvor *langt* man bør gå for å sikre evidens, og vi kjenner enda ikke til detaljene i hva denne strategien vil si for Norads utvikling på dette feltet. Like fullt mener vi at oppgaven står seg som et relevant innspill i en debatt som er særs viktig ikke bare for Norad, men for internasjonal bistand generelt.

## 2 Bakgrunn

### 2.1 Bistandsproblematikken i dag

Norsk utviklingspolitikk har som eksplisitt mål å fremme menneskerettigheter og å redusere fattigdom verden over (Norad, 2011a). Det overbærende målet gir mange muligheter for hva norske bistandsmidler skal brukes til, men i nyere tid har nærmere føringer særlig blitt gitt av sittende regjeringer (Solberg, 2014)(NTB, 2007) og gjennom FNs mål for bærekraft og fattigdomsreduksjon. Sistnevnte kjenner man først fra Tusenårsmålene, som ble vedtatt av medlemslandene i FN ved tusenårsskiftet. Målene skulle vare til 2015, og har i etterkant blitt revidert og oppdatert til det vi i dag kjenner som Bærekraftsmålene (FN, 2021). Disse 17 målene med de tilhørende 169 delmålene, danner et verdensomspennende fundament for hvordan bistand bør og skal fokuseres frem mot 2030. Målene er satt med formål om å “utrydde fattigdom, bekjempe ulikhet og stanse klimaendringene” (FN, 2021) – en noe omformulert versjon av den allerede eksisterende norske bistandsprioriteringen.

Samtlige 193 FN-land tilsluttet seg Bærekraftsmålene i 2015 (United Nations, 2015b). Legitimiteten knyttet til et slikt globalt samarbeid, tilsier at målene også får stor oppmerksomhet. Likevel er veien lang for å nå flere av målene. Mens fattigdom er mer enn halvert internasjonalt siden Tusenårsmålene ble definert (The World Bank, 2017), er reduksjonstakten for sult langt langsommere (Cooper et al., 2020). For analfabetisme er det allerede flere som slår fast at vi ligger for langt bak til å nå målene (Coughlan, 2019).

Det har hele veien vært klart at både Tusenårsmålene og Bærekraftsmålene kom til å bli vanskelige å nå. Bak mange av målene skjuler det seg formidable utfordringer, og innsatsen som skal til for å nå dem er opplagt stor. Nesten 700 millioner mennesker – i underkant av 10 prosent av verden – lever i ekstrem fattigdom (The World Bank, 2020). I 2014 var det fortsatt over en kvart milliard barn som ikke gikk på skolen (Our World in Data, 2019), og så sent som i 2016 levde over 800 millioner mennesker med kronisk underernæring (Hunger Notes, 2018). Flere tungvektene innenfor utvikling og bærekraft påpeker at det er utenkelig å nå målene innen 2030 dersom innsatsen ikke trappes opp kraftig. I følge The Social Progress Imperative, vil vi ikke nå målene før 2092 med dagens innsatsnivå (Palladium, 2020).

### 2.1.1 Uendelig behov - begrensede midler

I boken *The end of poverty* (Sachs, 2005) fra 2005, hevder Jeffrey Sachs at fattigdom kunne vært utryddet dersom verdens rike land bidro med 195 milliarder dollar hvert år fra 2005 til 2025. Den verdenskjente økonomens påstand er omstridt. Andre mener at bistand er nytteløst, mens atter andre mener det må mye mer til for å løse verdens fattigdomsproblemer. For å sette Sachs tall i perspektiv ble det av OECD-landene gitt rundt 161 milliarder dollar til bistand i 2020 (Bredvei, 2021), og i de foregående årene tilbake mot 2005 har bidragene som regel vært betydelig lavere. Behovet for bistand er altså i beste fall stort, og i verste fall er det uendelig. Samtidig er midlene begrenset. Å sette av midler til bistand er en politisk prioritering som både går på bekostning av andre tiltak og som krever politisk konsensus.

Statistikker for utviklingstrekk kan sammen med størrelsen på de beregnede kostnadene gi inntrykk av at kampen mot fattigdom er umulig å vinne (UN, 2020). Likevel gis det stadig mer til kampen mot utryddelsen av de nevnte verdensproblemene. Med andre ord finnes det stadig tro på at bistand fungerer og at det er mulig å løse noen av verdens største utfordringer. Denne oppfatningen deler også vi. Samtidig ønsker vi å peke på enkelte mekanismer som er avgjørende for at man skal lykkes i arbeidet, og hvor vi oppfatter at dagens norske bistand hemmes.

Den norske budsjettammen for bistand og utviklingshjelp er totalt på rundt 39,5 milliarder kroner (Regjeringen, 2021), og Norad forvalter rundt 16,5 av disse (Utenriksdepartementet, 2021). I OECDs tabeller for giverland, stikker de nordiske landene seg ut sammen med Luxemburg og Storbritannia. Disse fire landene er de eneste som oppfyller det vedtatte målet om at OECDs medlemsland skal bidra med 0.7 prosent av BNI til bistand. Mens Norge overoppfyller dette målet og i 2019 bevilget 1.02 prosent (OECD, 2019), var snittet blant de 30 medlemslandene i 2019 på 0.3 prosent av BNI (OECD, 2020). Norge er med andre ord blant de få OECD-landene som overoppfyller sine internasjonale bistandsforpliktelser. I tillegg er det verdt å merke seg at til tross for å være et lite land, ligger Norge på en 10. plass blant de 29 landene i absolutte ODA-bidrag – like bak Canada og foran land som Australia, Korea og Danmark. Dette bidragsnivået er det politisk konsensus i Norge om, og det ligger i dag til grunn for norsk bistand et mål om at vi skal bidra med 1 prosent av Brutto Nasjonalinntekt til bistand og utviklingsarbeid.



Til tross for stor politisk oppslutning rundt norsk bistandsaktivitet her hjemme, ser man jevnlig diskusjon om hvor mye Norge bør bidra med internasjonalt - særlig i tiden rundt fremleggelsen av statsbudsjettet. I tillegg til gjenvendende debatter om hvorvidt Norge bør bruke større eller mindre summer, ser man også debatt om hvorvidt det ovenfornevnte 1-prosentsmålet i det hele tatt bør være førende for norsk bistand. Kritikken mot dette målet bunner ut i at man mener at resultater bør være viktigere enn summen man bruker (Hagen, 2018).

Debatten om prosentmålet preges av to ulike syn. Forkjempernes syn oppsummeres godt av utviklingsminister Dag Inge Ulsteins svar når han blir møtt med spørsmål om hvorvidt målet egentlig er god politikk: (...) *det er ikke tvil om at når vi ser de resultatene vi har oppnådd, er det viktig å opprettholde nivået. (Det er) veldig sterk sammenheng mellom det vi bidrar med og utviklingen* (Frafjord, 2020). På den andre siden finner vi dem som mener at prosentmålet ikke er et fornuftig mål, fordi det er et utgiftsmål og ikke et resultatmål. En av disse er økonomiprofessor Rune Jansen Hagen ved Universitetet i Bergen. Han peker på at veksten i prosentbaserte budsjettposter har vært enorm på kort tid i Norge, da statsbudsjettet er nærmere tredoblet over en 20-årsperiode. Én måte å lese ham på, er å si at behovet bør dokumenteres før midlene innvilges, slik omtrent alle andre poster på statsbudsjettet settes opp (Hagen, 2018).

Det er med andre ord sprikende meninger om den videre retningen for norsk bistandspolitikk, men til tross for uenigheten erkjenner de aller fleste at det er et reelt behov for norske bistandsmidler. Og uansett hvilke av synene som blir førende for bistandsbudsjettet, vil det til en hver tid være større globale behov enn hva Norge stiller til disposisjon over statsbudsjettet. Utgangspunktet for denne oppgaven er at vi må gjøre mest mulig med de begrensede midlene vi har og her har vi kommet til oppgavens kjerne. Mye tyder nemlig på at mens Norge er blant de beste i klassen på å sette av midler, er det ikke nødvendigvis slik at vi får mest mulig ut av midlene.

## 2.2 Norads virksomhet

Når bistand diskuteres i en norsk kontekst, er det som regel snakk om midlene som forvaltes av Norad. Norad - direktoratet for utviklingssamarbeid, er ansvarlig for utviklingsarbeid og forvaltning av norske bistandsmidler. I følge direktoratets egne nettsider er deres

hovedoppgave å “sette penger i arbeid for å oppnå resultater” (Norad, 2015b) og dette gjøres gjennom kunnskapsdeling, tildeling av støtte til humanitære prosjekter og oppfølging av disse. Driften finansieres over statsbudsjettets avsetninger til bistand, og Norad mottar eksterne føringer for driften gjennom instruksjer og føringer fra Utenriksdepartementet og Klima- og Miljødepartementet (Norad, 2011b). I 2021 er Norad gjennom tildelingsbrev fra Stortinget satt til å forvalte 16 540 059 000 kroner (Utenriksdepartementet, 2021) og tildelingene har i de siste årene ligget forholdsvis stabilt opp mot det som ofte kalles 1-prosentsmålet – at Norge skal gi én prosent av brutto nasjonalinntekt til bistand.

Sett fra Norge kan det ofte være vanskelig å se de direkte effektene av bistandsmidlene og om de har gitt resultater. I tillegg har det, som vi skal komme tilbake til i Kapittel 4, blitt fremlagt rapporter som sier at det i det hele tatt kan være vanskelig å fastslå effekt av midlene som brukes i dag. Ett av svarene Norad har presentert for å imøtegå dette er deres egen strategi mot 2030 (Norad, 2021b). Denne forfekter et kunnskapsbasert syn på hvordan bistandsmidlene anvendes og gir sterke implikasjoner for hvordan Norad skal arbeide mer strategisk med finansiering av bistandsprosjekter.

Den overbærende tematikken i Norads strategi mot 2030 er mottoet *Fakta har makta*. I lys av strategien, vil vi først redegjøre for hvordan tilstanden er i Norad i dag. Å kjenne til hvilke krav de per nå stiller til resultatrapportering, målemetoder og informasjon fra prosjektene de finansierer er sentralt for å diskutere direktoratets videre utvikling i årene som kommer.

## 2.3 Dagens målemetode for norsk bistand

Norads hjemmesider inneholder en rekke ressurser for å lære mer om hvordan de arbeider. I en video om deres virke og bistandsarbeid sier de følgende om hvordan man måler et resultat: *“For å vite om disse tiltakene etter en stund har effekt, må man måle endringen de har bidratt til. Det gjør man ved å sammenlikne situasjonen før og etter tiltaket.”* (Norad et al., 2021). Avslutningsvis fremkommer det også i videoen hvordan det kan være flere faktorer som bidrar til effekt. Videre blir det sagt at man ved ytterligere måling i større grad kan fastslå effekten av hvert enkelte tiltak. Som vi skal komme tilbake til i oppgaven, stilles det imidlertid ikke krav om ytterligere måling og bruk av kontrollgrupper når organisasjoner skal søke om finansiering av bistandsprosjektmidler (Norad, 2020b).

Dette er heller ikke dekket i retningslinjer for ansatte i Norad (Norad, 2008). Direktoratet opererer i dag med en rekke retningslinjer, regelverk og manualer for tildeling av støtte, men kravene som fremkommer av disse sier lite om kvantitativ resultatmåling. Dette delkapittelet tar for seg hvilke krav som stilles til søkerorganisasjoner fra Norad, samt et relevant eksempel fra virkeligheten for å demonstrere kravene i praksis.

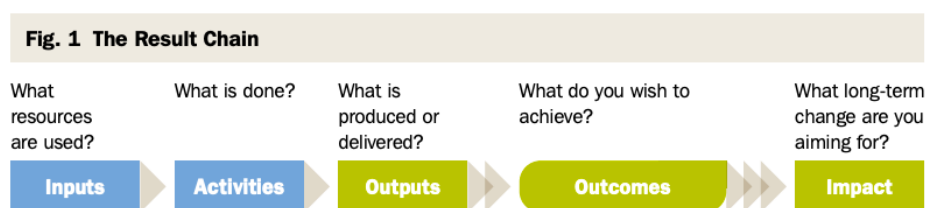
Ved gjennomgang av søknader benytter Norad det som omtales som et RAM (Resource Allocation Model)-system hvor et sett med kategorier og medfølgende kriterier legges til grunn for en samlet score (Norad, 2021a). Overordnet omfatter disse kategoriene (1) vurdering av søkerorganisasjon og hvorvidt de har kompetanse til å gjennomføre prosjektet, (2) vurdering av tidligere oppnådde resultater, (3) vurdering av tiltakets relevans, (4) vurdering av resultatstyring i tiltaket, (5) vurdering av risiko og tverrgående hensyn, (6) vurdering av tiltakets bærekraft, lokalt eierskap og exit-strategi, og (7) vurdering av tiltakets budsjett (Norad, 2021a). Under krav om resultatstyring i tiltaket, som er ytterst relevant for denne oppgaven, fremkommer det særlig et fokus på tydelige “outputs”, med konkrete utfall fra prosjektet, “outcomes”, med hvilke endringer dette vil medføre, og videre hvilket “impact” dette vil ha, det vil si hvilken samfunnseffekt dette vil bidra til. Videre legges det vekt på at søker må fremvise kunnskapsbasert endringsteori og/eller begrunnelse for hvorfor dette vil ha en effekt, samt sannsynlighet for dette. Resultatrammeverket skal også være tilpasset organisasjonens størrelse, ha tydelige indikatorer og kunne benyttes for oppfølging og eventuelle kursendringer.

Norad opererer også med en rekke regelverk som er spesifikke for støtte til ulike sektorer og avdelinger. Disse spesifiserer eller virker som tillegg til tidligere nevnte kriterier fra RAM-systemet. Videre omhandler de gjerne resultatstyring og -måling og spesifiseres som egne regelverk, underlagt ulike avdelinger, og for spesifikke målområder.

Særlig relevant for denne oppgaven, foreligger det også krav om at sivilsamfunnsorganisasjoner som søker om prosjektmidler også må dokumentere baseline-data. Som spesifisering av hva de definerer som baseline-data, presiserer Norad blant annet omfattende problembeskrivelse, samt kvantitative og/eller kvalitative data som dokumenterer den aktuelle tilstanden i området som skal behandles i det omsøkte prosjektet (Norad, 2015a). På Norads nettsider finnes det til slutt også flere utfyllende manualer som kan benyttes for Norad-ansatte, så vel som søkerorganisasjoner, sånn

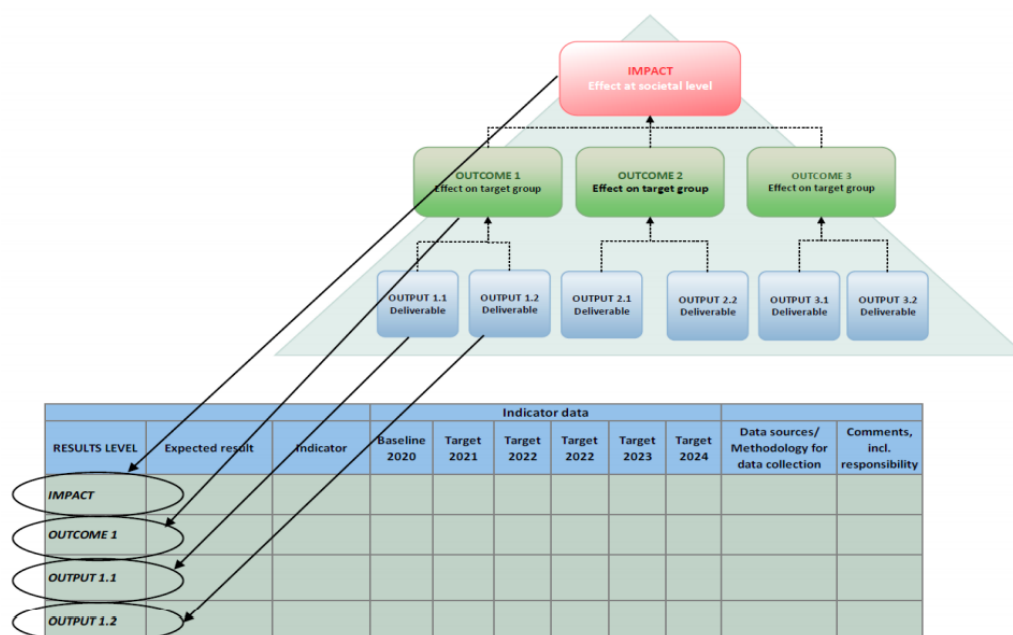
som *Results Management in Norwegian Development Cooperation - a practical guide og Guidance note on risks and results management for applicants* (Norad, 2008)(Norad, 2020b). Disse manualene ligger åpent tilgjengelig på Norads nettsider og vektlegger baseline, targets, indikatorer, outputs og outcomes, samt risikostyring og rapportering. I førstnevnte manual illustreres kjeden av resultater som vist under.

**Figur 2.1:** Resultatkjede - Results Management in Norwegian Development Cooperation (Norad, 2008)



Manualen erkjenner at effekter kan komme som resultat av et mangfold av faktorer som ligger utenfor omsøkte programs kontroll og rekkevidde. Like fullt er det i rammeverket for måloppnåelse kun krav om måling fra baseline til endline. Kun i ett tilfelle i manualen nevnes bruk av kontrollgrupper som tema. Dette finnes i manualens del om forberedelsesfasen hvor søkeren stilles spørsmålet “*Er det behov for en separat baseline study?*”. Dette adresseres ikke videre i foreslått rammeverk for resultatmåling. Det foreslåtte rammeverket for resultatmåling tar utgangspunkt i resultatkjeden ved å definere ønsket *impact* og *outcome*, for så å identifisere nødvendige *outputs*, *activities* og *inputs* som kreves for å oppnå disse. Slike rammeverk for resultatkjeder er også ofte brukt i andre analyser. I denne oppgaven vil også et tilnærmet likt rammeverk - Strategikartet gjengitt fra Lien, Knudsen og Baardsen bli brukt (Lien et al., 2016). Rammeverket er likevel kvalitativt og særlig nyttig i strategiutvikling, men mangelfult som resultatrammeverk om det ikke komplementeres med robuste målemetoder og mer aktiv resultatstyring.

*Guidance note on risks and results management for applicants* er en annen manual som kan hentes fra Norads nettsider og benyttes i utforming av prosjekter. Denne manualen bygger på mye av foregående nevnte punkter som baseline values, targets, indikatorer, outputs, outcomes og impact (Norad, 2020b). Manualen legger videre frem et tentativt resultatrammeverk som kan benyttes av søkerorganisasjoner. Dette rammeverket er vist under.

**Figur 2.2:** Resultrammeverk (Norad, 2020b)

På samme måte som foregående manualer og retningslinjer vektlegger dette rammeverket benyttelse av baseline og oppfølging av endringer i disse verdiene gjennom mål og indikatorer. Heller ikke her forekommer det spesifiserte krav til bruk av kontrollgrupper. Norad har i tillegg utarbeidet en mal for resultatrammeverk for sine søkere og denne ligger tilgjengelig på deres utlysningssider. Denne malen bygger hovedsakelig på rammeverket fra Figur 2.2 og en fullstendig versjon er vedlagt i appendiks.

### 2.3.1 SOS Barnebyers prosjekt for å forbedre skoletilbud for barn utsatt for omsorgssvikt

På Nordads egne hjemmesider publiseres det jevnlig artikler om gjennomførte bistandsprosjekter. Et av prosjektene som løftes frem som en suksesshistorie handler om prosjektf finansiering av SOS Barnebyer og hvordan de i perioden 2016-2019 bidro til at 7200 barn utsatt for omsorgssvikt fikk skolegang. Prosjektet skal også ha bidratt til bedre levevilkår og økonomi i barnas hjem, samt påvirket myndigheter til å styrke lovverket for barn utsatt for omsorgssvikt (Norad, 2020a). Det er liten grunn til å betvile at prosjektet har ført mye godt med seg. Like fullt er prosjektet tuftet på det overnevnte rammeverket med tilhørende resultatmåling.

I Figur 2.3 er et utdrag fra prosjektets resultatrammeverk hentet frem som illustrasjon. Likheten mellom det benyttede rammeverket og Norads krav og regelverk er åpenbar. Rammeverket følger Norads krav til baseline, tilhørende indikatorer og targets som fører til output, som igjen fører til outcome. Under output 3.1.1 i figuren, kan vi eksempelvis se nærmere på indikatorene og hva disse forteller. Antall barn som går på skole har gått fra 82 til 94%, med et mål på 95%. Dropout har gått opp fra 3 til 5%, prestasjonsnivå blant elevene har gått fra 55 til 85% for gutter og 86% for jenter og overgang til videregående har gått opp fra 56% for gutter og 47% for jenter, til 74% for gutter og 72% for jenter (Norad, 2020a). Disse tallene forteller helt klart en historie om positiv utvikling. Like fullt: Hvorvidt hele fortjenesten for resultatene burde tilskrives prosjektets gjennomføring og innsats er strengt tatt ikke annet enn en kvalifisert gjetning. Uten å se på relativ utvikling i liknende områder kan vi nemlig ikke med sikkerhet fastslå hvorvidt endringen er korrekt, høyere eller lavere og som følge av prosjektets innføring eller andre underliggende årsaker.

**Figur 2.3:** Resultrammeverk for SOS Barnebyer i Malawi (Norad, 2020a)

RESULTS	Indicators	2015 Baseline	2019 Achieved	2019 Targets
<b>Outcome 3:</b> Increased access to and benefit from quality education or vocational training for target children and youth in programme areas	# local schools providing quality education to target children	20 schools	54 schools	55 schools
<b>Output 3.1.1:</b> Improved school statistics for target children in the programme	% enrolled, attending school	82% boys 83% girls	94% boys 94% girls	95% boys 95% girls
	% Dropout	3% boys 3% girls	5% boys 5% girls	2% boys 2% girls
	% of educational performance	55%	85% boys 86% girls	80% boys 80% girls
	% Transition to secondary school	56% boys 47% girls	74% boys 72% girls	75% boys 75% girls

## 2.4 Strategikart

Gjennomgangen av hvordan Norad stiller krav til og måler resultater av prosjekter er særlig relevant i lys av deres nylig fremlagte langtidsstrategi. Når de her fremhever at de skal sette bistandsmidler i arbeid på en mer strategisk måte, anser vi direktoratets målemetoder som en viktig ressurs. Årsaken til at målemetoder betraktes som en ressurs er først og fremst at de er viktige for å identifisere gode prosjekter, fremskaffe troverdige resultater og for å fastslå hvorvidt midlene har effekt.

For å drøfte måling som ressurs videre i denne oppgaven, vil vi som tidligere nevnt analysere Norad og deres målemetoder med utgangspunkt i Strategikartet (Lien et al.,

2016).

En strategi kan beskrives som et samlet sett med valg og aktiviteter som legges til grunn for å oppnå mål og resultater (Lien et al., 2016). En strategi blir således en teori om hvordan en bedrift burde operere for å oppnå ønskede resultater. Det er her viktig å påpeke at strategier fungerer som teorier og hviler på antakelser og usikkerhet, noe som viktiggjør behovet for veloverveide antakelser og scenarioer for usikkerhet. En robust strategi med klare føringer for ulike scenarioer, vil gjøre det lettere å identifisere når strategien avviker fra dagens hendelser og hvordan håndtere disse endringene.

Strategikartet er et rammeverk utledet av Lien, Knudsen og Baardsen og benyttes for å analysere hvordan en bedrift eller organisasjon kan og burde operere med utgangspunkt i ressurser, omstendigheter og mål. Strategikartet omhandler fire deler; **ressurser, aktiviteter, kunder og resultater/mål**.

Hvilke *ressurser* en organisasjon har påvirker hvilke *aktiviteter* de burde gjennomføre, som igjen påvirker hvilke *kunder* de burde rette seg mot, som til slutt påvirker hvilke *resultater/mål* de kan oppnå. Verktøyet kan også benyttes i motsatt retning ved å ta utgangspunkt i hvilke *resultater/mål* en organisasjon ønsker å oppnå, som legger føringer på hvilke *kunder* de må nå ut til, som legger føringer på hvilke *aktiviteter* de må gjennomføre, som til slutt kan fortelle noe om hvilke *ressurser* de må anskaffe. I vårt tilfelle vil det være sistnevnte metode som vil være hensiktsmessig å benytte, med utgangspunkt i Norads tiårsplan som deres resultat/mål.

## 3 Litteratur og forskning

### 3.1 Resultatmåling og den nye gullstandarden

#### 3.1.1 Hvorfor måler vi?

Robuste målemetoder er fundamentalt for å kunne stole på resultater og påstander (Woolridge, 2012). All empirisk forskning hviler derfor på robuste målemetoder og resultater for å i det hele tatt ha verdi og mulighet for anvendelse. Dersom en styresmakt innfører en endring er de avhengige av gode målemetoder for å trekke inferens og med sikkerhet si hvilke ringvirkninger politikken deres har medført. Ved fravær av robuste målemetoder kan endringen ha medført både gode, dårlige og ikke-eksisterende effekter, men å trekke konklusjoner vil være fåfengt og uten særlig kredibilitet. Det samme gjelder bistand. Dersom Norad ikke evner å måle prosjektene sine på tilstrekkelig vis, vil de ikke kunne trekke konklusjoner av eget arbeid, noe som svekker deres kredibilitet. Dette delkapittelet vil forsøke å gi en grunnleggende innføring i ulike målemetoder og hvorfor Norad vil dra fordel av å benytte disse.

Resultatmåling baserer seg på ulike former for analyse for å trekke inferens. Inferens handler om å kunne trekke pålitelige slutninger om en populasjon basert på et utvalg av populasjonen. En vanlig fremgangsmåte for å trekke inferens, er innsamling av data for relevante variabler for deretter å gjøre en statistisk analyse av disse. Regresjoner er et av de vanligste verktøyene og brukes for å undersøke effekten av en variabel på et utfall. Man benytter seg da av begrepet “ceteris paribus” som betyr “alt annet likt” (Woolridge, 2012). Ved å holde alle andre variabler uendret kan man måle effekten av en endring i den uavhengige variabel,  $x$ . Vi kaller variabelen vi ønsker å måle effekten på den avhengige variabelen,  $y$ , ettersom den avhenger av ulike  $x$ -variabler.  $X$ -variablene kalles uavhengige variabler ettersom de er uavhengige av hverandres tilstand.

Gjennom statistisk analyse kan vi fastslå at dersom den uavhengige variabelen  $x$  øker med 1 enhet, eller 1 prosent, øker den avhengige variabelen,  $y$ , med en viss mengde. For å kunne produsere robuste resultater er det flere forutsetninger som må legges til grunn og forholdsregler som må tas i innsamling av data og utarbeidelsen av regresjonsmodeller. Statistikk kan lett bli misvisende dersom forskerne gjør metodiske feil



eller har underliggende motiver som preger funnene. Dette kan for eksempel oppstå ved forskning som samler inn kun den data de ønsker å se og deretter trekker inferens til en hel populasjon fra ikke-representative utvalg.

Videre er det mange ulike tilnærminger og målemetoder som er mulig å anvende med ulike fordeler og anvendelsesområder. Denne oppgaven kommer til å ta for seg målemetodene RCT og Difference-in-Difference, da disse er anerkjente og ofte brukes i forskning, samt mulige for Norad å anvende i sitt arbeid. Disse mener vi er særlig relevante ettersom de tar høyde for underliggende trender og avvik fra disse.

Analysen tar videre høyde for omitted variable bias ved å utvide regresjonsmodellen vår med kontrollvariabler. Omitted variable bias oppstår når uavhengige variabler som har effekt på den avhengige variabelen ikke inkluderes i regresjonen (Woolridge, 2012). Disse havner da i feilledet og deres effekt blir feilaktig tildelt andre uavhengige variabler. La oss illustrere med et eksempel i tråd med normal bistandsproblematikk. En humanitær organisasjon ønsker å forbedre leseferdigheter blant barn i alderen 5-7 år i Tanzania. Barna får derfor utdelt nye skolebøker og gjennomfører et nytt undervisningsopplegg over 3 år. Etter 1 år har samtlige av barna en bedre forståelse av alfabetet og mange leser på et høyere nivå. Innføringen av studieopplegg og nye skolebøker må vel da ha hatt sin tiltenkte effekt? Svaret er mest sannsynlig ja, men uten robuste målemetoder er det umulig å si. I dette tilfellet vil for eksempel naturlig kognitiv utvikling ha en positiv naturlig effekt på leseferdigheter, men ved å utelate naturlig utvikling plukkes denne effekten opp av andre variabler, som skolebøker og studieopplegg. Disse vil da få en oppblåst og ikke reell tildelt effekt ved at kognitiv utvikling ender som en “omitted” variabel (Woolridge, 2012).

### 3.1.2 RCT

RCT eller “Randomized Controlled Trials” er en forskningsmetode som innebærer å randomisere hvilke individer eller grupper som mottar behandling og hvilke som fungerer som kontrollgrupper. På denne måten klarer man å sammenlikne to grupper som i utgangspunktet er helt like når det kommer til observerbare og ikke-observerbare bakgrunnsvariabler, samt deres relative utvikling som resultat av behandling. Et eksempel som kan illustrere dette er ved utprøving av en ny medisin til å behandle en tenkt sykdom. Da kan man samle et utvalg av en populasjon, det vil si et utvalg mennesker med den

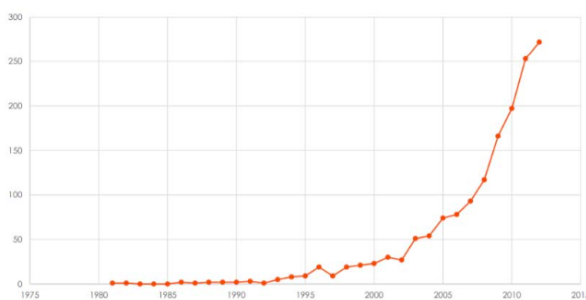
gitte sykdommen, for så å tilfeldig tildele noen medisin og andre placebo/eller ingen behandling. På denne måten skaper man to identiske grupper, hvor eneste forskjellen er tilført behandling eller ikke. Grunnen til at disse gruppene kan anses som like bunner ut i det som kalles Law of Large Numbers (Angrist og Pischke, 2014) (Routledge, 2020). For å forklare denne loven kan vi ta for oss en terning. Kaster vi denne nok ganger vil antall resultat med 1 til 6 være tilnærmet like og forventet verdi vil være 3,5. På samme måte vil behandlingsgruppe og kontrollgruppe tendere mot et gjennomsnitt, det vil si være like, dersom vi trekker nok individer fra en populasjon og fordeler de i de to gruppene. For eksempel kan vi forvente at ved randomisert trekning vil vi finne like mange menn som kvinner i hver gruppe. Viktigere er det dog at vi ved randomisert trekning kan forvente at gruppene vil være like også på områder vi ikke kan måle. Dette legger fundamentet for å kunne bruke en del av populasjonen som kontrollgruppe slik at eneste forskjellen blir innføring av behandling. Alt annet holdes dermed likt - *ceteris paribus* (Woolridge, 2012).

## 3.2 RCT-bevegelsen

RCT har i lang tid har vært den ledende forskningsmetoden innenfor medisinfaget. Grovt forenklet kan man ved å tildele deltakere i et forskningsprosjekt tilfeldig til enten behandlings- eller kontrollgruppe, undersøke den rene effekten av et tiltak målt mot å ikke gjøre noe.

Innenfor økonomifaget – og særlig når det kommer til utviklingsarbeid – fikk RCT en vesentlig økt oppslutning i perioden rundt 2019. Årsaken var utdelingen av Nobelprisen i økonomi som ble tildelt forskerne Abhijit Banerjee, Michael Kremer og Esther Duflo for deres eksperimentelle tilnærming til fattigdomsreduksjon

(Foundation, 2019). De tre forskerne har alle bidratt til viktige funn ved bruk av RCT som metode, og stadig flere har fått øynene opp for denne kvantitative tilnærmingen til forskning innenfor utvikling og fattigdomsbekjempelse. Fra å være en lite diskutert og brukt metode innenfor fagfeltet, omtales nå RCT av mange som gullstandard innenfor



**Figur 3.1:** Publisert bruk av RCT (Cameron et al., 2015)

målemetoder. Dette har skapt mye debatt i fagmiljøer verden over.

Forkjemperne for RCT fremholder særlig at denne metoden er blant de aller beste verktøyene man har for å kunne fastslå kausale sammenhenger og at den eliminerer såkalt selection bias – utvalgsskjevhet – altså at man tildeler behandling på måter som kan påvirke funn. I sin gjennomgang av tilstanden for anvendt økonometri, skriver økonomene Athey og Imbens at når det kommer til å fastslå kausalitet og finne intern validitet, er RCT allment anerkjent i fagmiljøet som en særlig troverdig metode (Athey og Guido W, 2017).

Metoden gjør det nemlig enkelt å administrere én eller flere behandlingstyper og samtidig undersøke effekten av dem isolert, og er dermed et effektivt verktøy å benytte seg av i pilotfaser før prosjektet oppskaleres i full effekt. Der utviklingsøkonomifaget tidligere har konsentrert seg mye om store, makroøkonomiske prosjekter og tendenser, har nobelprisvinnerne i stor grad tatt for seg mikroøkonomi og nærmere undersøkt effekten av små intervensjoner som kan ha store utslag til tross for beskjedne inngrep og kostnader. Ved å gjøre dette har de levert viktige bidrag til fattigdomsreducerende og utviklingsfremmende tiltak som faktisk fungerer. Et godt eksempel er Edward Miguel og Michael Kremer's forskning på behandling av innvollsormer (Miguel og Kremer, 2004). Deres anbefaling om massebehandling mot ormer og funnene deres på effekt for skolegang, står seg fortsatt som et av de fremste tiltakene organisasjonen GiveWell anbefaler å finansiere, 16 år etter at studien ble publisert (Foundation, 2021).

Samtidig er det langt fra full enighet om at RCT er den nye gullstandard for resultatmåling, og det legges stadig frem relevant kritikk mot metoden. Særlig fremholdes det at metoden er kostnadskrevene uten å kunne gi ekstern validitet (Bryan, 2019). Prosjekter som gjennomføres ved bruk av RCT kan altså ikke alltid generaliseres til andre populasjoner, som svekker argumentet for å bruke RCT som pilotgrunnlag for prosjekter som senere skal skaleres. Videre påpekes det at RCT-prosjekter har en tendens til å oppnå lav forskningskvalitet på grunn av rene brukerfeil, som at man ofte benytter for små populasjoner grunnet hvor kostnadskrevene RCT-arbeid er. I tillegg nevnes andre fenomener, som for eksempel pilot-bias, hvor resultatet kan blåses opp ved at organisasjonen som gjennomfører piloten kan være både mer motivert og mer kvalifisert til å gjennomføre tiltaket enn eksempelvis myndighetene som skal overta prosjektet senere.

Veldig mange av disse utfordringene har imidlertid noe til felles: De er ikke unike for RCT, men er snarere problemer for et bredt spekter av målemetoder. Man skal kanskje være forsiktig med å utrope RCT som en gullstandard på alle måleområder, men det lite rom for tvil om at dersom det brukes riktig på de riktige prosjektene, står verktøyet i en særstilling når det kommer til avklaring av intern validitet (Athey og Guido W, 2017).

### 3.3 Difference-in-Difference

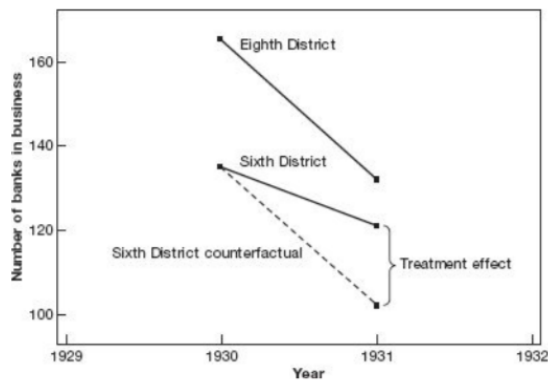
Difference-in-difference, heretter kalt DiD, er en annen tilnærming som er mer anvendelig i praksis der randomisert intervensjon er vanskelig. Denne metoden anerkjenner at behandlingsgruppe og kontrollgruppe kan være forskjellig, men måler likevel den relative forskjellen som følge av en behandling. Ved å predikere trender for behandlingsgruppe og kontrollgruppe kan man måle hvorvidt avvik fra trend er større eller mindre relativt til kontrollgruppen (Angrist og Pischke, 2014).

Et eksempel som kan illustrere bruk av metoden, er finanskrisen i 1930. Etter denne ble det hyppig diskutert blant eksperter hvorvidt pengepolitikk kunne eller ikke kunne ført til færre konkurser og mindre panikk. Da panikken begynte å spre seg på 30-tallet begynte folk å ta ut pengene sine i så stor grad at bankene ikke lenger klarte å etterkomme sine plikter. Gjennom føderale reserver i hvert bankdistrikt kunne også de føderale bankene velge egen pengepolitikk og dermed om de ville heve eller senke lånerenten til distriktets banker. Dette åpner for å gjennomføre en DiD og se på den direkte forskjellen som følge av restriktiv og liberal pengepolitikk ved å sammenlikne to nærliggende distrikter og antall konkurser som følge av finanskrisen. Distrikt 6, Atlanta Federal Bank, hadde en stimuleringspolitikk som gikk ut på å senke renten, mens distrikt 8, St. Louis Federal Bank, sin politikk tilsa at de begrenset kreditt i nedgangstider. Under finanskrisen reagerte bankene som forventet svært ulikt der lån fra Atlanta økte med 40 prosent, mens lån fra St. Louis ble redusert med 10 prosent. 1 juli 1931, om lag 8 måneder etter krisen var det 132 banker åpne i distriktet til St. Louis og 121 banker åpne i distriktet til Atlanta. Selv om dette i utgangspunktet gir inntrykk av at pengepolitikk ikke har hatt effekt, må vi ta høyde for utgangspunktet. 1. juli 1930 var det 165 banker åpne i distriktet til St. Louis og 135 banker åpne i distriktet til Atlanta. Ved å bruke DiD kan vi undersøke den direkte effekten som følger av pengepolitikk (Angrist og Pischke, 2014).

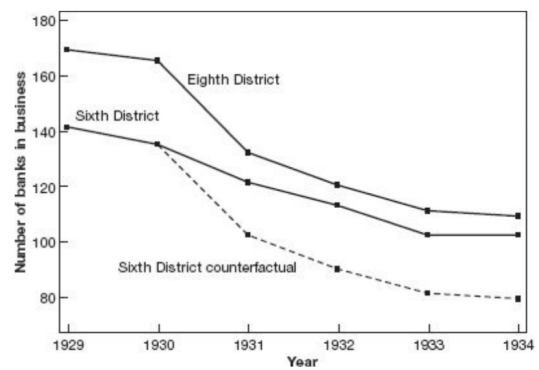
$$\begin{aligned}
 \delta_{DD} &= (Y_{6,1931} - Y_{8,1931}) - (Y_{6,1930} - Y_{8,1930}) \\
 &= (121 - 132) - (135 - 165) \\
 &= -11 - (-30) = 19
 \end{aligned}
 \tag{3.1}$$

Gjennom DiD finner vi  $\delta$  som her er 19 og forskjellen mellom 8 og 6 distrikt. Det vil si at en liberal pengepolitikk reddet 19 banker. Dette kan videre illustreres grafisk hvor vi ser utviklingen i det 8. og 6. distrikt. Her ser vi også den kontrafaktiske utviklingen dersom distrikt 6 ikke hadde innført en liberal rentepolitikk og fulgt tilsvarende utvikling som distrikt 8.

(a) Kortsiktig trend for 6 og 8 distrikt



(b) Langsiktig trend for 6 og 8 distrikt



**Figur 3.2:** Utvikling i 6 og 8 distrikt (Angrist og Pischke, 2014)

## 4 Tidligere evalueringer og nåværende strategi

Norads arbeid med evaluering er ofte selv gjenstand for evaluering. I tillegg til årsrapporter blir det også gjennomført eksterne evalueringer av Norads arbeid med jevne mellomrom. Disse utføres på bestilling fra direktoratets evalueringsavdeling og offentliggjøres på Norads hjemmesider. I dette kapittelet vil essensen av tidligere evalueringer gjengis for å dokumentere kvaliteten på Norads metodiske tilnærming til bistandsevaluering. I tillegg vil Norads langtidsstrategi (Norad, 2021b) gjennomgås da denne fremstår som direktoratets imøtekommelse av kritikken. Kapittelet drøfter avslutningsvis i hvilken grad strategien svarer på den tidligere kritikken, og om direktoratet har tatt nødvendige grep for å unngå slik kritikk i fremtiden.

### 4.1 Kritikk av Norads metode

I januar 2017 ble rapporten “The Quality of Reviews and Decentralised Evaluations in Norwegian Development Cooperation” (Itad og Christian Michelsens Institutt, 2017) publisert. Denne rapporten ble levert som en ekstern evaluering på bestilling fra Norad, og ble forfattet av fagpersoner fra Christian Michelsens Institutt og den globale organisasjonen Itad som spesialiserer seg på slike evalueringer. Basert på 60 desentraliserte evalueringer utført i 2014, det vil si prosjektevalueringer utført på initiativ av mottakerorganisasjon, fant forfatterne store svakheter i rapporteringen som ble gjort. Av de 60 rapportene som ble gjennomgått, ender over halvparten (38) opp med “poor rating» - altså en score på 1 eller 2 på en skala fra 1 til 4 – innenfor kategorien “findings”. Selv om rapporten opererer med langt flere kategorier, er denne særlig relevant i denne sammenhengen ettersom det er denne som indikerer evalueringenes evne til å fastslå evidens og resultatenes beviselighet og robusthet. Sagt med andre ord fastslår rapporten at Norad ikke kan vite om evalueringene av anvendte midler faktisk kan slå fast at norske utviklingsmidler har den effekten de påstås å ha. Eller ytterligere forenklet: Norad kan ikke i tilstrekkelig grad fastslå at bistandsmidlene de forvalter har effekt.

Med rapporten fra 2017 ble det altså tydelig at de desentraliserte evalueringene som gjennomføres ved hvert prosjekt, har klare metodiske svakheter. Denne utfordringen har også vedvart, og i 2020 ble det offentliggjort nok en rapport om tilsvarende tema. Rapporten “Quality Assessment of Decentralised Evaluations in Norwegian Development Cooperation” (Ternström AB, 2020) ble utført av det svenske Ternstrom Consulting og baserer seg på evalueringer fra 2018 og 2019. Rapporten, som skulle følge opp den tidligere strenge kritikken rettet mot evalueringsarbeidet, konkluderer i stor grad med det samme som den tre år yngre rapporten. Forfatterne skriver i denne rapporten at “gjennomgangen av evalueringsrapportene indikerer at et stort antall av de desentraliserte evalueringene ikke baserer seg på data, metode og analyser som er egnet til å gi troverdig informasjon om initiativene og utfallene fra dem”. Som et ytterligere varsko fremheves det fra rapportforfatterne at denne rapporten er basert på frivillig innsendte evalueringer, og de bekymrer seg for at de dermed kan ha mottatt kun de “beste” evalueringene. Dette vil i så fall være ytterligere problematisk for kvalitetsnivået på de eksterne evalueringene.

Forfatterne av 2020-rapporten beskriver godt selv hvorfor det er problematisk med svake metodiske evalueringer. På side 30 i “Quality Assessment of Decentralised Evaluations in Norwegian Development Cooperation” skriver forfatterne følgende om evalueringsrapportenes svake data- og metodegrunnlag:

*“Low quality on the criteria relating to approach and design, methods, sources of evidence, and methodological appropriateness has three main effects. First, if the approach and methods are not appropriate, the findings of the evaluation may not be correct. The correlation coefficient for ‘methodological approach’ and ‘description of design’ is 0.75, the correlation coefficient for ‘methodological approach’ and ‘description of methods’ is 0.73. Second, it becomes difficult or impossible for the reader to understand how the evaluator has selected, collected and interpreted the evidence upon which findings and conclusions are based. Third, transparency of the evaluation process is limited, which not only makes it difficult to follow the line of evidence but also reduces the credibility of conclusions and recommendations. This is a serious weakness of the decentralised evaluations that were rated.”* (Ternström AB, 2020).

For Norad har problemstillingen vært kjent over tid. I snart et tiår har eksterne rapporter jevnlig pekt på utfordringer knyttet til kvalitetssikring og resultatmåling. Likevel var det

ikke før i 2016 at det ble stilt uniforme krav fra Utenriksdepartementet til når prosjekter skulle evalueres (Itad og Christian Michelsens Institutt, 2017). Til gjengjeld gjaldt disse kun for prosjekter over en viss økonomisk rammestørrelse og med en varighet på mer enn to år. Tidligere hadde det vært opp til den ansvarlige tilskuddsansvarlige å avgjøre hvorvidt et prosjekt skulle evalueres. Også i 2014 ble det publisert en ekstern evalueringsrapport fra Itad og CMI (Itad og Christian Michelsens Institutt, 2014). Funnene herfra viser at det har vært en positiv utvikling over årene, men utgangspunktet var også her relativt svakt hva gjelder evidensfokus. I rapporten kommer det frem at av de 20 rapportene som er innsamlet for nærmere undersøkelse, adresserer ingen av de 20 kvaliteten på bevisene og resultatene som er funnet.

Tilsvarende funn bekreftes tre år tidligere i Norads årsrapport for 2011 (Norad, 2011c) hvor det allerede i forordet står beskrevet at “Ingen av evalueringene og studiene bestilt av Evalueringsavdelingen og ferdigstilt i 2011 kunne i tilstrekkelig grad gi informasjon om utviklingsresultater (outcomes) eller virkninger”. I denne rapporten pekes det på at man over tid har hatt mer fokus på hvordan bistandsprosjekter *gjennomføres* og mindre på hvordan de *virker*. I tillegg henvises det også til årsrapporten for 2009 hvor det ble fastslått at man i fremtiden måtte legge større vekt på resultater. Det kan derfor med stor sikkerhet fastslås at Norad lenge har hatt en utfordring knyttet til resultatmåling. Deres egne rapporter dokumenterer en positiv utvikling, men det fremkommer tydelig fra de nyere evalueringene at det fortsatt er markante behov for ytterligere opptrapping av resultatdokumentasjon og datainnsamling knyttet til prosjektene de finansierer.

## 4.2 Norads strategi

Det er tydelig ved gjennomgang av Norads mål og strategi at disse rapportene ikke har gått upåaktet hen. Ifølge Norads egne nettsider er deres hovedoppgave å “sette penger i arbeid for å oppnå resultater” (Norad, 2015b) og i deres resultatportal beskriver de sitt motto som *resultater i kampen mot fattigdom*. Norads hovedoppgave utbroderes videre på deres nettsider gjennom seks overordnede mål for Norads virksomhet. Disse inkluderer blant annet “å forvalte tilskudd i henhold til UD og KOM”, “gi bistandsfaglige råd til UD, KOM og utenriksstasjoner” og “utvikle og drifte forvaltningssystemene for bistand” (Norad, 2015b). Disse tre målene er trukket frem ettersom de er særlig relevante for vår



oppgave, samtidig som de illustrerer Norads arbeid innenfor bistandssektoren, så vel som deres ønske om å bistå, kvalitetssikre og effektivisere norsk utviklingsarbeid på en god måte.

Disse overordnede målene er videre nedfelt i Norads strategi frem mot 2030. I 2021 kom nemlig Norad med sin strategi for 2030. Strategien er bygget på FNs bærekraftsmål, det å sette penger i arbeid sammen med samarbeidspartnere og det å benytte seg av en kunnskapsbasert tilnærming (Norad, 2021b). Med hensyn til det sistnevnte punktet trekker strategien spesifikt frem setningen *fakta har makta* for å illustrere Norads vei videre. Etter mange år med kritikk er det følgelig tydelig at Norad ønsker å imøtekomme denne og utvikle seg ved å sette kunnskap og robuste resultater i fokus.

Kapittel 3 og Norads resultatmåling vitner likevel om svake metoder og upålitelige resultater, selv om det er mulig disse tas høyde for i den nylanserte strategien. I dagens tilgjengelige retningslinjer og manualer, som RAM-systemet og manualer som *Results Management in Norwegian Development Cooperation - a practical guide* (Norad, 2021a) og *Guidance note on risks and results management for applicants* (Norad, 2020b) er det dog fortsatt sterkt fokus på resultater, men med utilstrekkelige krav og lite robuste rammeverk. Norad har selv publisert en rekke evalueringer som peker på at de opererer med for lite robuste metoder og rammeverk. Dermed er det lite trolig at rapporteringsproblematikken har vært ukjent for Norad, og heller mer sannsynlig at det har vært en bevisst avveining å prioritere midler til å hjelpe kontra å måle.

### 4.3 Oppsummering

Eksterne evalueringer fra de siste 10 årene kommer med krass kritikk av Norad og vitner om stillstand hva gjelder Norads resultatmåling. Det har skjedd små, men viktige, endringer de siste årene fra å fokusere på gjennomføring av prosjekter til hvordan de virker i 2014 og krav om baselineverdier i 2015. Til tross for disse, maler rapportene fra 2017 og 2020 et negativt bilde når det kommer til evalueringskvaliteten. Norad evner fortsatt ikke å dokumentere effekt av eget arbeid og implisitt heller ikke å dokumentere om de forvalter norske bistandsmidler på en god måte. I Norads strategi for 2030 forsøker de å ta stilling til dette ved å fremheve setningen *fakta har makta* og hvordan dette skal gjennomsyre Norads arbeid og resultater. Ved gjennomgang av Norads historiske målemetoder i kapittel 3, samt

dagens tilgjengelige manualer og krav, synes robuste målemetoder dog fortsatt å være underrepresentert og -kommunisert både i det daglige virket og i direktoratets strategi. Selv Norads egen “gullstandard”, i form av deres presenterte og foreslåtte rammeverk Norad (2008), er lite robust og ikke tilstrekkelig for å dokumentere effekt. Norads svake metoder og alternative robuste metoder vil illustreres nærmere gjennom kvantitative analyser i neste kapittel.

## 5 Analyse av ungdomsklubber i Uganda

Dette kapittelet har til hensikt å sammenligne Norads målemetoder med mer robuste og kvantitative tilnærminger. Særlig forsøker kapittelet å fremheve viktigheten av å bruke kontrollgrupper og hvorfor fravær av slike vil medføre upålitelige resultater. I kapittelet benyttes tallgrunnlag fra et tidligere prosjekt der RCT ble brukt som målemetode. I tillegg til at prosjektet er basert på bruk av RCT, gjennomfører vi her egne eksempler ved bruk av DiD. Dette for å illustrere hvordan Norads eksisterende målemetoder kan gi misvisende og lite robuste resultater.

### 5.1 Innledning om prosjektet

Hypotesen vår sier at Norad risikerer å gå glipp av avgjørende data for å bevise effekt av prosjektene de finansierer. Begrunnelsen er at de ikke krever at det brukes tilstrekkelig rigorøse metoder for å sikre evidens i analyser og evalueringer utført av tredjepart. For å illustrere dette har vi tatt for oss et annet forskningsprosjekt og anvendt datagrunnlaget deres til demonstrasjon av mer robust kvantitativ metode. Målet er å vise at Norads krav til metode, som i stor grad baserer seg på baseline- og endlinedata for behandlingsgrupper, kan gi andre implikasjoner enn man finner ved bruk av mer omfattende kvantitative målemetoder.

For å forsikre oss om at vi opererer med gode data og at tallene er kvalitetssikret, har vi valgt å hente et åpent datasett tilknyttet en publisert studie fra det anerkjente og fagfelleverderte tidsskriftet *American Economic Journal*. Vi har valgt ut en artikkel fra *Applied Economics*-utgaven, hvor man finner artikler knyttet til utvikling og utdanning. Disse er ofte også særlig relevante for mange av de Norad-finansierte prosjektene. Artikkelen dette kapittelet baserer seg på heter *Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa* (Bandiera et al., 2020). Helt sentralt for oppgaven var det å finne en artikkel av høy kvalitet med åpent datasett og med relevans for oppgavens tematikk. Disse kriteriene oppfylles godt av artikkelen.

*Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa* tar for seg unge kvinner i Uganda og deres grad av "empowerment" som følge av en mangefasettert intervensjon gjennom et eget program. Programmet som følges er iverksatt

av organisasjonen BRAC, en større bistands- og utviklingsorganisasjon med opphav i Bangladesh (BRAC, 2020). De benytter et sammensatt program hvor unge mennesker – i dette tilfellet kvinner – får delta i det de kaller ELA-grupper/-klubber (Empowerment and Livelihood for Adolescents), og mens programmet først ble opprettet i Bangladesh er dette forsøket utført i Uganda. Gjennom deltakelse ønsker BRAC at kvinnene skal læres opp i både yrkesrelevant kunnskap og i livsmestringsegenskaper som skal hjelpe dem med å ta informerte valg om eksempelvis ekteskap og reproduksjon. Vi velger derfor å forstå ordet “empowerment” som myndiggjøring og livsmestring.

Gjennom bruk av RCT ønsker forskerne å undersøke effekten av ELA-møtene på unge kvinner i Uganda, og de finner eksempelvis signifikante økninger i andelen som er selvstendig næringsdrivende og signifikante reduksjoner i ungdomsgraviditeter og ufrivillig samleie hos deltakerne. Funnene er oppløftende for programmet, men for denne oppgaven er det sentrale punktet å merke seg hva slags metodisk grunnlag de benytter seg av for å fastslå dette. Bruken av randomisert tildeling og kontrollgrupper gjør at forskerne kan fastslå effekt med langt høyere sikkerhet enn hva Norad kan med sin bruk av kun baseline/endline. I tillegg gjøres det omfattende undersøkelser av spillover-effekter og andre effekter som kan påvirke funnene de viser til. Videre har de et bevisst forhold til kostnaden per innsats og mål i prosjektet. Dette er også en sentral fordel ved å bruke regresjoner med flere variabler - at man kan se nærmere på effekten fra spesifikke deler av prosjektet. Dermed kan de ha kostnad og effekt i bakhodet når de publiserer evalueringen. Dette lar BRAC iverksette tiltak som fungerer best mulig og ytterligere intensivere innsatsfaktorer som fungerer godt dersom prosjektet replikeres i fremtiden.

Fordi forskerne benytter seg av RCT og fordi artikkelen er fagfellevurdert, antar vi at dataene som benyttes er valide og at funnene er presise og korrekte. Tallene som anvendes skal være randomiserte, og ettersom de baserer seg på rundt 5500 personers deltakelse antar vi i utgangspunktet at vi kan forholde oss til normalfordelte resultater ved baseline i undersøkelsene. Likevel har vi ved enkelte punkter sett at de behandlede gruppene skiller seg signifikant fra kontrollgruppen. For å imøtegå dette har vi brukt DiD for å undersøke den isolerte utviklingen hos gruppene fra baseline til endline. Dette er et vesentlig poeng vi vil komme tilbake til når vi ser på forskjellene mellom Norads metode og metoden Bandiera et al. (2020) anvender.

## 5.2 Metodikk

Alle regresjoner, utregninger og illustrasjoner i analysen er gjennomført i STATA. Selv om funnene illustrerer deler av vår problemstilling, viser de et ufullstendig bilde hva gjelder prosjektets helhetlige omfang og funn. Inferens vedrørende det spesifikke prosjektet, dets deltakere, arbeid og effekt på livsmestring i Uganda burde derfor knyttes til oppgaven som helhet, ikke prosjektet for øvrig.

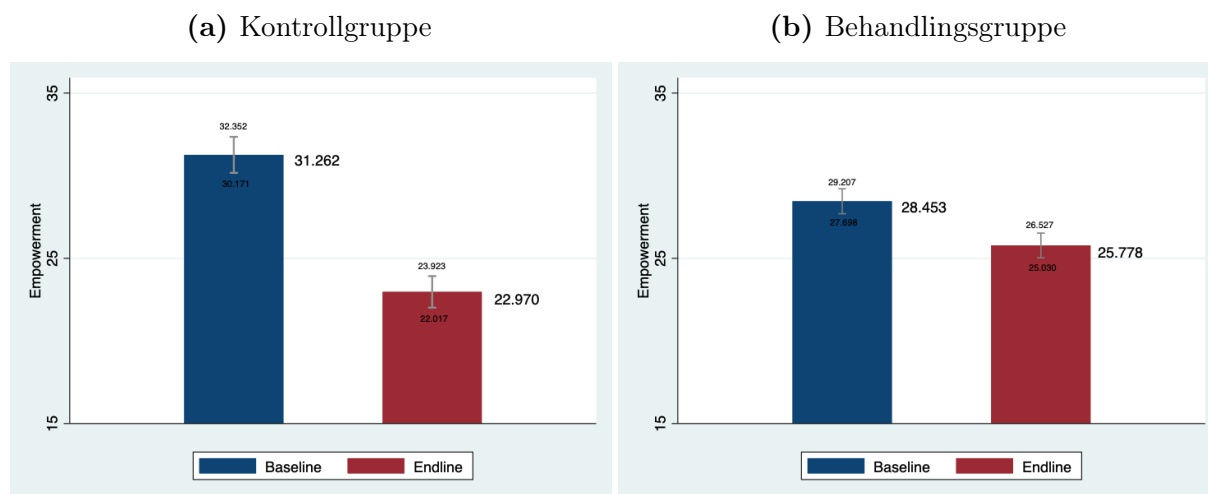
Ved gjennomgang av datasettet har vi valgt ut to variabler som kan illustrere ulike problematikker som kan oppstå ved fravær av kontrollgrupper. Variablene vi har valgt å fokusere på er “empowerment” og “bruk av kondom”, deres relative utvikling i kontroll- og behandlingsgruppe, samt effekt av behandling. Empowerment er en variabel som baserer seg på en spørreundersøkelse der spørsmålene sammen gir en score fra 1 til 100. Her er 100 høyest mulig empowerment, eller livsmestring og myndiggjøring. Bruk av kondom er en dummyvariabel der respondenter som svarer at de alltid bruker kondom ved samleie scorer 1 og andre 0. Ettersom variabelen er en dummy som enten påtar verdien 1 eller 0 vil gjennomsnittet ganget med 100 tilsvare prosent av deltakere som alltid bruker kondom. Både empowerment og bruk av kondom er ved gjennomgang av datasettet valgt fordi de illustrerer vår problemstilling. Like fullt vil vi her vise frem statistiske problemer som er helt normale å finne om man sammenligner datainnsamling med og uten kontrollgrupper. At vi i så måte preges av et “confirmation bias” - altså at vi aktivt søker funn som bekrefter vår egen teori og hypotese - er i denne sammenhengen derfor irrelevant.

### 5.2.1 Empowerment - tabell og grafisk fremstilling

Det finnes ulike metoder for å gjennomføre en DiD, med ulik grad av robusthet. Den første metoden vi akter å gjennomføre er DiD ved bruk av gjennomsnitt. I Figur 5.1 under vises derfor gjennomsnitt for empowerment i kontroll- og behandlingsgruppen på baseline og endline, illustrert gjennom et barplot.

Som man kan se av gruppe (a) har gjennomsnittet sunket med  $(31.26-22.97=8.29)$  8.29 poeng fra baseline til endline. Dette er basert på 2002 observerte personer i kontrollgruppen. Nedre grense for konfidensintervallet på 95 prosent er 30.171 på baseline, og øvre grense tilsvarende på endline er 23.923. Altså kan man med svært høy sannsynlighet fastslå at

man ser en signifikant nedgang i perioden.



**Figur 5.1:** Barplot av empowerment i kontrollgruppe og behandlingsgruppe

I gruppen (b) i Figur 5.1 kan man se utviklingen i behandlingsgruppen. Her har utviklingen også vært negativ, men med kun  $(28.45 - 25.78 = 2.67)$  2.67 poeng. Dersom Norad hadde gjennomført et liknende prosjekt og utelukkende benyttet seg av behandlingsgruppe, kunne de antatt en negativ effekt som resultat av behandling. Sannheten er dog en ganske annen. Den relative nedgangen er som vist,  $8.29 > 2.$ , mindre ved behandling enn for kontroll.

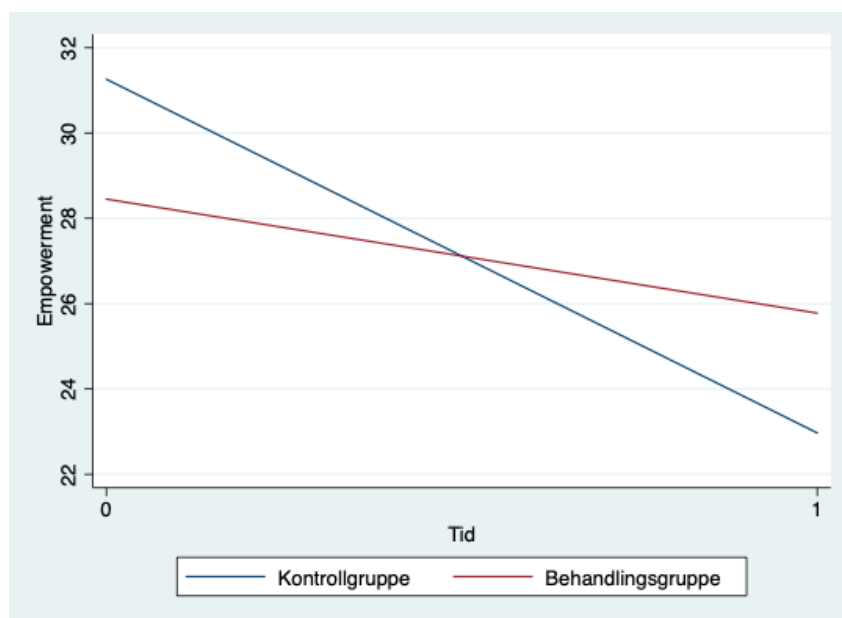
Med tallene over kan vi nå gjennomføre en DiD, som utledet i kapittel 3.3. Dette gjennomføres ved å regne ut forskjellen i gjennomsnittet ved endline og forskjellen i gjennomsnittet ved baseline for så å trekke disse fra hverandre. Vi finner da avviket mellom kontrollgruppen og behandlingsgruppen og dermed den direkte effekten av behandling på empowerment.

$$\begin{aligned} \delta_{DD} &= (Y_{Behandling, Endline} - Y_{Kontroll, Endline}) - (Y_{Behandling, Baseline} - Y_{Kontroll, Baseline}) \\ &= (25.78 - 22.97) - (28.45 - 31.26) = 5.62 \end{aligned} \tag{5.1}$$

Deltakere i behandlingsgruppen scorer dermed 5.62 poeng høyere som følge av behandling. Uten kontrollgrupper ville Norad kunne stemplet dette som et dårlig prosjekt der trenden har vært negativ som følge av behandling. I virkeligheten viser svaret tvert imot at behandling har ført til en redusert nedgang og isolert positiv effekt. Som nevnt

innledningsvis i kapittelet er det valgte prosjektet gjennomført som en RCT. Likevel ser vi signifikante avvik i baseline mellom kontroll og treatment. Som nevnt i kapittel 3.1.2 skulle i teorien *“The law of large numbers”* (Routledge, 2020) medføre like utgangspunkt for kontroll- og behandlingsgruppe. Likevel kan vi se fra gjennomsnittet ved baseline at disse avviker. Ved fravær av randomisert tildeling av behandling og kontroll, vil disse videre potensielt innebære større og mer systematiske avvik. Etersom Norad hverken gjennomfører randomiserte forsøk eller tar høyde for ulike trender, som gjennom DiD, vil prosjektene deres kunne resultere i større og mer omfattende avvik.

Eksempelet som i Figur 5.1 er illustrert ved bruk av DiD, gjennomsnittsberegninger fra baseline til endline og deretter presentert i barplot, kan også vises grafisk for å synliggjøre trendutviklingen. Dette er fremstilt i Figur 5.2 nedenfor. Til tross for at kontrollgruppen startet på et signifikant høyere nivå av selvstendighet/myndiggjøring, opplever denne gruppen en langt skarpere nedgang enn gruppen som mottar behandling. Den avgjørende forskjellen mellom Norads metode og en slik bruk av kontrollgruppe blir altså nettopp kontrollgruppen og hvordan man bruker den til å nyansere bildet som tegnes av den behandlede gruppen.



**Figur 5.2:** Empowerment indeks - grafisk fremstilling

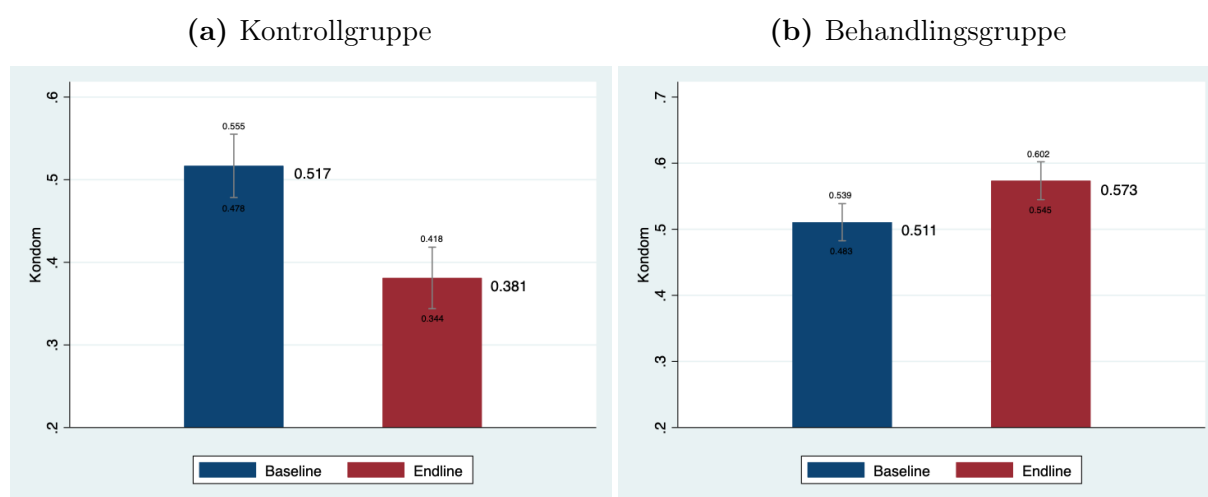
Dersom funnet fra Figur 5.1 stemmer, vil ELA-tiltaket kunne være en betydelig ressurs for å bidra til selvstendighet og myndiggjøring av unge kvinner i Uganda og lignende samfunn. Det er imidlertid ikke sikkert at Norad ville truffet samme konklusjon ved

bruk av sine metoder. Dersom Norad og tredjepartsorganisasjonene skulle funnet samme konklusjon, ville dette krevd omfattende innsikt om utviklingstrekk og demografiske ulikheter i befolkningen over tid. Attpåtil vil denne hypotetiske kunnskapen, uavhengig av hvor nøyaktig og god den skulle vært, være vanskeligere å bevise effekten av med dagens anvendte metode. Årsaken er naturlig nok at god kunnskap er vanskelig å omgjøre til evidens dersom man ikke har metoder for å sannsynliggjøre og bevise den. Dette er også grunnlaget for mye av kritikken Norad har mottatt i tredjepartsevalueringene gjennomgått i Kapittel 4 - Tidligere forskning.

### 5.2.2 Bruk av kondom - tabell og grafisk fremstilling

I neste eksempel ser vi på bruk av kondom fra baseline til endline. Som gjennomgått innledningsvis vil dummyen her påta seg en verdi mellom 0 og 1 og aggregert dermed vise prosentvis andel som alltid bruker eller ikke alltid bruker kondom ved samleie.

Av barplot kan vi se at den prosentvise andelen som alltid bruker kondom ligger like over 50 prosent i kontrollgruppen. Videre kan vi se hvordan andelen har sunket signifikant ved endline til 38%. Dette er en nedgang på 13.6%.



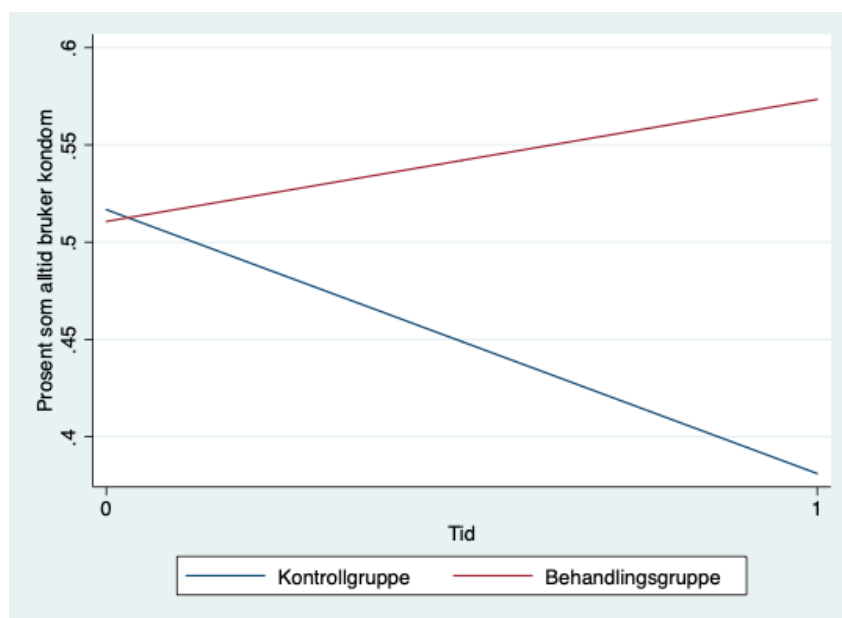
**Figur 5.3:** Barplot av prosent som alltid bruker kondom i kontrollgruppe og behandlingsgruppe

I behandlingsgruppen ser vi derimot en signifikant økning ( $0.573 - 0.511 = 0.062$ ). Vi kan på samme måte som i forrige eksempel her gjennomføre en DiD basert på gjennomsnitt før og etter behandling. Vi finner da avviket fra kontrollgruppen til behandlingsgruppen.



$$\begin{aligned}\delta_{DD} &= (Y_{Behandling,Endline} - Y_{Kontroll,Endline}) - (Y_{Behandling,Baseline} - Y_{Kontroll,Baseline}) \\ &= (0.573 - 0.381) - (0.511 - 0.517) = 0.198\end{aligned}\tag{5.2}$$

Ved samme utregning som tidligere ser vi her at behandlingsgruppen har 19.8% høyere sannsynlighet for å alltid bruke kondom som følge av behandling. Dette tilfellet skiller seg fra det innledende eksempelet ved at Norad her ville regnet prosjektet som en suksess, dog i mindre grad enn det faktisk har vært (19.8% > 6.2%). Faktum er at kontrollgruppen har hatt en sterk negativ utvikling og dermed har effekten av behandling vært desto større enn behandlingsgruppen gir inntrykk for alene.



**Figur 5.4:** Kondom - grafisk fremstilling

En tilsvarende grafisk fremstilling som i Figur 5.3 ble vist for empowerment er også gjort for kondombruk i Figur 5.4. I motsetning til empowerment-eksempelet, ser vi i dette eksempelet så og si identiske utgangspunkt ved baseline. Dette vil i utgangspunktet være å forvente ved randomisering hvor tilfeldig trekning til behandlings- og kontrollgruppen i teorien skal gi forholdsvis like grupper. I den grafiske fremstillingen blir effekten på bruk av kondom som prevensjon veldig tydelig. Der gruppene er så og si identiske ved baseline, er det motstridende trender hos de behandlede sett mot kontrollgruppen. Ved endline er de to gruppene signifikant ulike med god margin. Eksempelet er relevant for Norad fordi vi ser to trendlinjer med ulike fortegn. Dersom man kun hadde sett på de behandlede

utvikling, ville resultatet være langt mindre oppløftende enn det blir ved å se gruppen målt mot kontrollgruppen.

## 5.3 Funn og resultater

### 5.3.1 Empowerment - regresjon

I første eksempel har vi gjennomført en regresjon på “Empowerment” - myndiggjøring og selvstendigjøring - som avhengig variabel. For å benytte DiD har vi i kolonne (1) satt opp regresjonen med behandlingsstatus, tid og en kryssvariabel bestående av dummyvariablene behandling og tid. På denne måten får man den isolerte effekten av behandlingen og ikke den underliggende trenden. I kolonne (2) er regresjonen videre utbrodert med flere relevante variabler for å ta høyde for potensielt utelatte variabler og omitted variable bias. Her er faktorene *har barn*, *alder* og *inntekstgenererende arbeid (IGA)* inkludert. Regresjonslinjen med og uten kontrollvariabler kan skrives slik:

$$\text{Empowerment} = \beta_0 + \beta_1 \text{Behandlingsstatus} + \beta_2 \text{Tid} + \beta_3 \text{Kryssvariabel}$$

$$\text{Empowerment} = \beta_0 + \beta_1 \text{Behandlingsstatus} + \beta_2 \text{Tid} + \beta_3 \text{Kryssvariabel} + \beta_4 \text{IGA} + \beta_5 \text{Alder} + \beta_6 \text{Barn}$$

I kolonne (1) kan vi se at kryssvariabelen viser en koeffisient på 5.62 med signifikans på 1-prosentsnivå. Dette leses som at intervensjonen har gitt en styrking relativt til kontrollgruppen på 5.62 poeng på Empowerments 100-poengsskala. Ettersom skalaen går fra 0 til 100 kan vi tolke dette som at behandlingen gir 5.62 prosent økning, selv om det i utgangspunktet er snakk om poeng på en skala. Videre er det verdt å merke seg at behandlingsstatus er signifikant og negativ. Det tilsier at personer i behandlingsgruppen (behandlingsstatus=1) i snitt scorer 2.81 poeng lavere enn personer i kontrollgruppen ved baseline. Tid har også en sterk negativ effekt på 8.29 poeng (tid=1), noe som tilsier en sterk signifikant negativ utvikling for begge grupper fra baseline til endline. Sistnevnte poenger rundt tid og behandlingsstatus er også synlige i den grafiske fremstillingen.

I kolonne (2) viser funnene at kryssvariabelen styrker seg og forholder seg signifikant på samme nivå. Alle andre inkluderte variabler er også signifikante, men enkelte med så små koeffisienter at de er å regne som irrelevante i det store bildet. Særlig verdt å merke seg

er det at inntektsgenererende arbeid bidrar til nedgang i selvstendigjøring. Tilsvarende finner vi at dummyen for om man har barn virker negativt på de unge kvinnenes grad av selvstendigjøring. Dette er riktignok ikke på 1-prosentsnivå men på 5-prosents, som tross alt er å regne som sterk signifikans like fullt. Årsaken til at vi ikke finner samme signifikans på denne variabelen kan være at den ikke er helt uavhengig av arbeids- og aldersvariablene, men vi hadde i utgangspunktet forventet å se signifikans på høyere nivå knyttet til denne variabelen. Vi har også forsøkt å sette inn dummyvariabler for hvorvidt kvinnene har partner eller ikke samt om de går på skole. Her finner vi effekter, dog uten signifikans for noen av dem. Årsaken kan være at variablene kannibalisierer hverandre og spiser virkning, men uten grundig undersøkelse er dette vanskelig å si noe om på annet grunnlag enn ren intuisjon.

**Tabell 5.1:** Regresjon av empowerment index

	(1)	(2)
	Empowerment	Empowerment m. kontrollvariabler
Behandlingsstatus	-2.81*** (0.65)	-2.90*** (0.67)
Tid	-8.29*** (0.75)	-8.81*** (0.77)
Kryssvariabel (dummy)	5.62*** (0.93)	5.83*** (0.94)
Inntektsgenererende arbeid (dummy)		-1.79** (0.73)
Alder		0.58*** (0.09)
Barn (dummy)		-1.87** (0.81)
Konstant	31.26*** (0.53)	22.74*** (1.49)
Observations	11932	11454
$R^2$	0.01	0.02

Standardfeil i parenteser

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Merk: Tallene brukt i regresjonen er hentet fra datasettet tilhørende studien Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa, publisert i American Economic Journal: Applied Economics VOL. 12, NO. 1, januar 2020 (side 210-59). I kolonne (1) er variablene behandlingsstatus, tid og kryssvariabel benyttet slik standard DiD-metode tilsier. Tilsvarende variabler er også brukt i kolonne (2). Her er i tillegg variablene Inntektsgenererende arbeid, Alder og Barn lagt til. Variablene Inntektsgenererende arbeid og Barn er dummy-variabler med respektive verdier 0 dersom den observerte ikke tjener penger eller har barn, og 1 ved det motsatte tilfellet. Alder gjengir kun hvilken alder den observerte har ved baseline - starttidspunkt for studien.

### 5.3.2 Bruk av kondom - regresjon

I neste eksempel har vi gjennomført en DiD på bruk av kondom gjennom dummyer for behandling og tid, samt en kryssvariabel som kombinerer de to. På samme måte som i forrige eksempel har vi videre gjennomført en regresjon der vi inkluderer plausible kontrollvariabler for å kontrollere for utelatte variabler og omitted variable bias. Vi har da inkludert dummyer for om personen har partner, om personen har barn, om personen bruker annen form for prevensjon og om personen går på skole. Regresjonslinjen med og uten kontrollvariabler kan følgende skrives slik:

$$Kondom = \beta_0 + \beta_1 \text{Behandlingsstatus} + \beta_2 \text{Tid} + \beta_3 \text{Kryssvariabel}$$

$$Kondom = \beta_0 + \beta_1 \text{Behandlingsstatus} + \beta_2 \text{Tid} + \beta_3 \text{Kryssvariabel} + \beta_4 \text{Partner} + \beta_5 \text{Barn} + \beta_6 \text{AlternativPrevensjon} + \beta_7 \text{Skole}$$

I kolonne (1) kan man se at blant personer i behandlingsgruppen (behandlingsstatus=1), vil det være 1% færre som alltid bruker kondom. Dette er dog ikke signifikant og vi kan dermed ikke forkaste nullhypotesen om at behandlingsstatus ikke har noen effekt, det vil si at de to gruppene i utgangspunktet er like. Dette virker også plausibelt dersom vi husker tilbake til den grafiske fremstillingen der baseline for de to gruppene var tilnærmet identisk.

Videre kan vi se nærmere på variabelen Tid. Her kan vi se at det i endline,  $time=1$ , vil være 14% færre som alltid bruker kondom. Dette er signifikant på 1% nivå og vi kan dermed slå fast at det foreligger en forskjell mellom baseline og endline. Dette kan dog være en naturlig utvikling. For å tilskrive behandlingen effekt, må variabelen Treattime undersøkes. Treattime er kryssvariabelen mellom behandling og tid og viser den direkte effekten av behandling på bruk av kondom. Gjennom denne kan vi se at det er 20% flere som alltid bruker kondom som følge av behandling. Dette er også signifikant på 1% nivå.

Videre har vi inkludert kontrollvariabler for å korrigere for omitted variable bias. Behandlingsstatus er fortsatt ikke signifikant, men tid er fortsatt signifikant på 1% nivå og ved endline er det 13% færre som alltid bruker kondom. Kryssvariabelen blir videre redusert med 7% som følge av innføring av nye variabler, men er fortsatt signifikant på 1% nivå. Dette viser at kryssvariabelen i første kolonne var overestimert og plukket opp

mye effekt fra andre variabler. Den er likevel fortsatt på 13% og behandlingen har hatt en signifikant effekt. Gjennom kontrollvariabler ser vi også at blant de som har partner er det 13% færre som alltid bruker kondom, blant de som har barn er det 27% færre, blant de som bruker annen form for prevensjon er det 12% færre og blant de som går på skole er det 11% flere som alltid bruker kondom. Dette maler et bilde der behandling har hatt sterk effekt, men også der andre variabler har en betydelig påvirkning både positivt og negativt. Alle kontrollvariablene er signifikante på 1% nivå.

**Tabell 5.2:** Regresjon av prosent som alltid bruker kondom

	(1)	(2)
	Kondom	Kondom m. kontrollvariabler
Behandlingsstatus (dummy)	-0.01 (0.02)	-0.01 (0.02)
Tid (dummy)	-0.14*** (0.03)	-0.13*** (0.03)
Kryssvariabel (dummy)	0.20*** (0.03)	0.13*** (0.04)
Partner (dummy)		-0.13*** (0.02)
Barn (dummy)		-0.27*** (0.02)
Annen form for prevensjon (dummy)		-0.12*** (0.03)
Går på skole (dummy)		0.11*** (0.02)
Konstant	0.52*** (0.02)	0.61*** (0.02)
Observations	3674	2578
$R^2$	0.02	0.19

Standardfeil i parentes

\*  $p < 0.10$ , \*\*  $p < 0.05$ , \*\*\*  $p < 0.01$

Merk: Tallene brukt i regresjonen er hentet fra datasettet tilhørende studien Women's Empowerment in Action: Evidence from a Randomized Control Trial in Africa, publisert i American Economic Journal: Applied Economics VOL. 12, NO. 1, januar 2020 (side 210-59). I kolonne 1 ble kun variablene behandlingsstatus, tid og kryssvariabel benyttet i tråd med standard DiD-fremgangsmåte. I kolonne 2 ble tilsvarende variabler inkludert, i tillegg til kontrollvariabler for å ta høyde for omitted variable bias. Samtlige variabler er dummyer og tar dermed verdien 0 eller 1.

## 5.4 Begrensninger for inferens fra prosjektet

Gjennomgangen av dette prosjektet er ment for å illustrere hvorfor det er viktig å bruke omfattende måleverktøy når man skal gjennomføre prosjekter hvor mange elementer kan påvirke utviklingen av resultatene man ønsker å måle. At vi i tillegg ser forskjeller i

gruppene, selv etter at de er randomiserte, understreker dette poenget. Samtidig må det også sies at vi opererer med våre egne bias når vi viser frem disse funnene. Årsaken er at vi har funnet frem tall fra forskningsprosjektet som illustrerer poenget vårt. Mange av tallene fra forskningsprosjektet ville kunnet fortelle andre historier, og det ville ikke vært akseptabelt i et normalt forskningsprosjekt å selektere tall slik det har blitt gjort i dette kapittelet. Her har imidlertid formålet vært å vise frem to av mange mulige utfall fra et helt tilfeldig prosjekt, og så er også gjort med både høy signifikans og forholdsvis stor effekt fra variablene vi måler på. Like fullt er det enkelte fenomener som bør adresseres når det gjelder robustheten i funnene:

Ettersom tallene her er hentet fra et prosjekt som er gjennomført og forsket på av andre, er data og variabler tilgjengelig for bruk begrenset til de forskerne selv har ansett som hensiktsmessig å hente inn data på. Bandiera et al. (2020) har etter alle solemerker gjort gode vurderinger på hvilke faktorer det er nødvendig å hente inn data på, men dette begrenser samtidig hva som er mulig å si om mer perifere forhold og datapunkter i studien, som eksempelvis kondombruk, da det er flere faktorer som kan påvirke dette enn vi har tilgjengelige data på. Til tross for dette, er det i utformingen av regresjonene forsøkt å gi et så dekkende bilde som mulig av de faktiske funnene, og det er forsøkt å velge ut de best passende av mange tilgjengelige variabler både for å illustrere det underliggende poenget og for å yte det originale prosjektet den rettferdigheten det fortjener.

Som tidligere gjennomgått ble det funnet betydelige forskjeller mellom behandlings- og kontrollgruppen ved baseline i eksempelet om empowerment. Dette er trolig kun en anomalitet, og den skal ha i tillegg være rettet opp ved å bruke DiD, men det er vanskelig å si noe helt definitivt om hva som er årsaken til og effekten av ulikhetene når disse i utgangspunktet skulle vært rettet opp av randomiseringen.

Gjennomgående i dette kapittelet opererer vi med en forutsetning om at organisasjonene som gjennomfører Norad-finansierte bistandsprosjekter antar null økning i øvrig befolkning dersom de ikke mottar den aktuelle behandlingen. Det er mulig at det finnes eksempler på at dette ikke er tilfelle, men dette er ikke normalen blant de tilgjengelige rapportene på Norads hjemmesider, og det ligger ingenting som tyder på dette i Norads reglement for rapportering da det kun kreves bruk av baseline-rapportering og ingen krav om kontrollgrupper.

## 6 Diskusjon

Dette kapittelet tar for seg Norads målemetoder i lys av funnene fra Kapittel 5 og momentene gjennomgått i tidligere kapitler. Videre forsøker kapittelet å si noe om hvilke metoder Norad kan og bør forlange fra søkerorganisasjoner i fremtiden. En lengre diskusjon om implementering av målemetoder i Norads strategi gjennomføres videre i Kapittel 7, slik at diskusjonen i dette kapittelet i større grad fokuserer på de rent metodiske momentene.

### 6.1 Implikasjoner for Norads resultater

Gjennomgangen av Norads egne krav til resultatmåling i Kapittel 2 og de eksterne evalueringene omtalt i Kapittel 4 viser tydelig at det finnes svakheter i Norads eksisterende krav til målemetoder. Konsekvensene av svakhetene synliggjøres av funnene i Kapittel 5. Dersom Norad skal oppfylle sin overordnede målsetting om å “sette penger i arbeid for å oppnå resultater”, vil krav til målemetoder være et sentralt punkt å ta tak i.

Som vist i Kapittel 2, bruker Norad et eget RAM-system for vurdering av prosjektsøknader. Av kravene som knytter seg til (resultat)måling i dette systemet, finner vi at det i dag forlanges at søkerorganisasjonen dokumenterer outcome, output, targets og målindikatorer (Norad, 2021a). I tillegg stilles det krav til problembeskrivelse og kvantitative grunnlagsverdier. Til tross for fokuset på kvantitative data, inneholder ikke systemet krav til bruk av kontrollgrupper. Inntrykket av at dette er de kravene som stilles i dag forsterkes også av at de harmonerer med kravene som stilles i eksempelvis *Results Management in Norwegian Development Cooperation - a practical guide* og *Guidance note on risks and results management for applicants* (Norad, 2008)(Norad, 2020b). Som demonstrert, kan slik bruk av kvantitative metoder gjøre at man trekker helt andre konklusjoner enn de faktiske forholdene gir grunnlag for.

Som grundig gjennomgått i Kapittel 4, er vi ikke alene om å peke på disse svakhetene. I over 10 år har Norad blitt kritisert for svakheter i kravene til målemetoder, gjennom evalueringer de selv har bestilt. At over halvparten av rapportene som ble evaluert av Itad og CMI i 2017 ble vurdert som metodisk svake i kategorien “findings”, var et varsko for Norad (Itad og Christian Michelsens Institutt, 2017). Tre år senere viste en ny evaluering at det fortsatt stod dårlig til med de metodiske tilnærmingene som ble brukt

til prosjektevalueringer. At rapportene som ble tilgjengeliggjort for evaluering i denne omgangen kun bestod av frivillig innsendte rapporter, ble også understreket av forfatterne bak evalueringen. Dersom dette medførte at de kun fikk se de “beste” rapportene, kunne dette være alarmerende for den totale kvaliteten (Ternström AB, 2020). Konsekvensene av den lave metodiske kvaliteten er at Norad i realiteten ikke kan dokumentere effekt av bistandsarbeidet de finansierer. Implisitt er det ikke mulig å fastslå at pengene de setter i arbeid, faktisk oppnår resultater.

Løsningen for Norad vil være å benytte mer presise evalueringsmetoder. Både RCT og DiD er tidligere gjennomgått i oppgaven og passer godt som valg av kvantitative metoder å benytte seg av for Norad. Den avgjørende forskjellen på disse metodene og de som i dag kreves av Norad, er bruken av kontrollgrupper for å sikre presise effektestimater. I kapittel 5 ble det vist to ulike eksempler på hvordan bruk av kontrollgrupper nyanserer funnene man ellers ville gjort. I empowerment-eksempelet fremkommer det hvordan manglende bruk av kontrollgruppe kan føre til forkastning av prosjektet fordi det har lavere endline enn baseline. I eksempelet om kondom ser man hvordan effekten kan underestimeres dersom man utelater bruk av kontrollgrupper, ettersom kontrollgruppen viser markant nedgang mens det er oppgang i behandlingsgruppen.

Disse to poengene illustrerer kun to av de mulige scenarioene som kan forekomme om man ikke bruker kontrollgrupper; forkastning av et egentlig suksessfullt prosjekt og underestimering av effekt. I virkeligheten er det flere ulike målefeil som kan forekomme. Vi kan illustrere dette ved at Norad opererer med en nullhypotese,  $H_0$ , der behandling/prosjekt ikke har noen effekt og alternativ hypotese,  $H_1$ , der behandling har skapt en positiv effekt. Norad risikerer med dagens metoder å gjøre både såkalte *type 1-feil* og *type 2-feil*. Type 1-feil forekommer dersom man forkaster en sann nullhypotese. Type 2-feil forekommer dersom man ikke forkaster en usann nullhypotese (Woolridge, 2012). Ulike scenarioer der Norad kan feilestimere kan følgelig oppsummeres slik:

1. Feilestimere effekt dersom kontrollgruppe og behandlingsgruppe har hatt lik utvikling
2. Underestimere effekt, dersom kontrollgruppen har hatt en negativ utvikling
3. Overestimere effekt dersom kontrollgruppen har hatt en positiv utvikling

Kontrollgruppen kan følgelig ha en positiv eller negativ utvikling og kun dersom det ikke



forekommer en utvikling i det hele tatt vil fravær av kontrollgrupper fortsatt vise til et korrekt resultat. Fravær av utvikling i kontrollgruppen er dog umulig å si noe om uten å faktisk ta høyde for den, og vi ender dermed på samme konklusjon. Uten bruk av robuste målemetoder kan ikke Norad med sikkerhet slå fast hvorvidt midlene deres har skapt en effekt, hverken på godt eller vondt. Dette kan føre til ineffektiv ressursallokering som over tid resulterer i store dødvektstap. Ved feilestimering av effekt risikerer Norad å støtte prosjekter som ikke evner å skape en positiv endring eller i verste fall støtte prosjekter med negativ påvirkning. Tilsvarende kan de også risikere å skrinlegge prosjekter med god effekt fordi de ikke fanger opp den positive effekten relativt til kontrollgruppen. Ved tildeling av midler til prosjekter uten å kvalitetssikre effekten finnes det altså reell mulighet for at man finansierer mindre gode bistandsprosjekter på bekostning av de bedre.

Gjennom kvalitetssikring og robuste målemetoder vil ressursallokeringen i Norad over tid kunne bedres. Kanskje viser det seg at et prosjekt har større effekt enn antatt ved å se nærmere på kontrollgruppen og dermed fortjener ytterligere støtte. Gode prosjekter kan da videreføres, og tilsvarende vil dårlige prosjekter identifiseres og skrinlegges. En annen og intuitivt god effekt vil være at man får mer presise mål på effekt og virkning av bistandsmidler. Et illustrerende eksempel vil være SOS Barnebyer-eksempelet som ble gjennomgått i Kapittel 2. Her forteller Norad om hvordan 7200 barn utsatt for omsorgssvikt har fått skolegang som følge av prosjektstøtte (Norad, 2020a). Det er liten grunn til å tro at prosjektet ikke har hatt effekt – særlig når det er snakk om direkte finansieringsprosjekter. Samtidig ville det ha verdi for Norad og SOS Barnebyer å vite om det i tillegg var andre, bakenforliggende årsaker enn rene finansierings spørsmål som holdt barna unna skolebenken. Spesielt viktig vil denne kunnskapen være om prosjektet senere skal replikeres og gjennomføres i lignende miljøer. I så fall vil Norad og SOS Barnebyer enklere kunne identifisere faktorer som er viktige å hensynta neste gang et slikt prosjekt skal utformes, finansieres og igangsettes.

## **6.2 Kvalitativt eller kvantitativt - hvilken metode bør benyttes?**

Før et forskningsprosjekt skal innledes, må det bestemmes hva slags metodisk tilnærming man skal ha til prosjektet. Man skiller her mellom kvantitativ og kvalitativ metode.

### **Kvantitative metoder**

Med kvantitative metoder menes innsamling og bruk av data for å teste om det finnes kausale sammenhenger mellom variablene man undersøker. Ved bruk av harde tall søker man å finne generaliserbare sammenhenger for områder, grupper eller andre populasjoner. Slik metode kan også benyttes for å undersøke eksempelvis endringer over tid, virkning av en spesifikk hendelse eller sammenhenger mellom to eller flere utviklingstrekk. Som tidligere gjennomgått, er det her vi plasserer RCT og bruken av DiD som to sterke verktøy. I Kapittel 3 ble det gjort rede for de vanligste kritikkene som rettes mot RCT, og kritikken er som regel også gjeldende for både DiD og andre rigorøse målemetoder. Bruk av kontrollgrupper koster gjerne mye å gjennomføre, uten at det nødvendigvis gir ekstern validitet (Bryan, 2019). I tillegg kan slike metoder ofte være tidkrevende og til tider vanskelig å gjennomføre på en god måte. Like fullt er RCT ofte regnet som det fremste verktøyet for å sikre objektivitet og generaliserbare resultater innen forskning (Webber og Prouse, 2018). For vår del er både RCT og DiD særlig relevante metoder ettersom de sikrer bruk av kontrollgrupper. DiD er dog mindre ressurskrevende og derfor potensielt mer anvendelig for Norad i første omgang. Dette skyldes omfanget av nødvendig datainnsamling ved RCT og medfølgende kostnader. Samlet sett sikrer begge metodene likevel høy intern validitet og objektivitet i funn, noe som er viktige momenter for Norad. Samtidig finnes det også tilfeller hvor rigorøs kvantitativ måling ikke er like godt egnet. I prosjekter hvor mellommenneskelige faktorer, subjektive forhold og erfaringer er viktige, vil kvantitative metoder være mindre funksjonelle å benytte for å måle resultater (Hammarberg et al., 2016).

### **Kvalitative metoder**

Kvalitative metoder anses gjerne som motstykket til de kvantitative. Denne metoden benyttes særlig når det er vanskelig å tallfeste effekter eller påvirkning. Hammarberg, Kirkman og de Lacy oppsummerer dette som metoden man bruker når man vil undersøke “spørsmål om erfaringer, meninger og perspektiver – gjerne fra et deltakerstandpunkt” (Hammarberg et al., 2016). Metoden er altså godt egnet til å utforske mer subjektive erfaringer og å skaffe seg innsikt i tematikk der tall ikke nødvendigvis ville vært formålstjenlig. De vanligste verktøyene inkluderer gjerne dybdeintervjuer, observasjon og notater utført både før, underveis og etter prosjektets gjennomføring. Styrken ved metoden er gjerne det unike innblikket det kan gi i prosjektet, dets særegenhet og de

påvirkede. Intervjuer med forsøkspersoner eller andre som påvirkes kan gi svar man sjeldent ville fått ved å kun forholde seg til rent tallmateriale. Snur man på dette, kan dette unike innblikket også være en utfordring ved metodebruken. Ved både subjektive preg på utformingen av spørsmål og subjektive svar, kan generaliserbarheten av funn være vanskelig. Dette kan gjøre replikering av funn vanskelig, særlig dersom man undersøker mange variabler i spesifikke kulturer. Metoden får også kritikk for å være kostnadskreven og lite ressurseffektiv (Eyisi, 2016). For en organisasjon som Norad kan disse utfordringene være vanskelig å se bort fra og det vil dermed være en avveining for dem mellom ressursbruk og robuste og replikerbare funn.

### 6.3 Oppsummering og anbefalinger

Som vi har sett i gjennomgangen av Norad, finnes det både elementer av kvalitative og kvantitative metoder i Norads og tredjepartsorganisasjonenes evalueringer av bistandsprosjekter. Der det tidligere har blitt påpekt at Norad har vært mer opptatt av gjennomføringen av bistandsprosjektene heller enn selve output, har bruken av intervjuer, observasjon og informanter vært nyttige verktøy (Hammarberg et al., 2016). Over tid har bruken av mer kvantitativ metode blitt standard, og i de eksterne evalueringene gjennomgått i Kapittel 4 ser man fra 2016 og utover at bruk av tall fra baseline og kvantifisering av resultater er løftet frem som krav (Norad, 2015a). For Norad og tredjepartsorganisasjonene er det mange gode argumenter for å benytte seg av begge formene for metode. De burde først og fremst i større grad benytte kvantitative metoder for å finne kausale sammenhenger og replikerbare resultater. For å få dette til er de følgelig nødt til å bruke mer rigorøse metoder enn dagens. Samtidig er det også sentralt at de utførende organisasjonene forstår menneskene de skal hjelpe og at bistand utformes sammen med mottakerne. Dette kan bidra til legitimitet og oppslutning i mottakerland, og dermed spiller kvalitative metoder som intervjuer, informanter og observasjon viktige roller (Hammarberg et al., 2016). I denne oppgaven forsøker vi dermed ikke å argumentere mot bruken av kvalitative metoder, men heller for at det arbeidet som gjøres på det kvantitative må styrkes og gjøres mer robust.

## 7 Implikasjoner og implementering i Norads strategi

### 7.1 Strategikart

Strategikartet benyttes for å belyse sammenhengen mellom *resultater*, *mottakere*, *aktiviteter* og *ressurser*. Ved å analysere disse punktene kan Norad implementere nødvendige ressurser, utføre gitte aktiviteter, målrette arbeidet mot identifiserte mottakere og oppnå ønskede resultater. I denne delen av analysen knyttes først Norads virksomhet opp mot det teoretiske rammeverket, før implikasjoner av en slik resultatkjede drøftes ved hjelp av en tilpasset versjon av strategikartet.

#### **Resultater/mål**

Norad har som overordnet mål å sette penger i arbeid for å oppnå resultater (Norad, 2015b). Resultatene Norad ønsker å oppnå, i tråd med deres tiårsstrategi, er å sette penger i arbeid på en kunnskapsmessig måte, bidra til Bærekraftsmålene og skape robuste og troverdige resultater (Norad, 2021b). Robuste og kvalitetssikrede resultater, i tråd med tidligere viste metoder, settes dermed som et mål og er følgelig utgangspunktet for både strategikartet og Norads virksomhet. Gjennom ønskede resultater kan Norad videre bestemme hvilke kunder de ønsker å nå ut til, aktiviteter dette krever og dermed hvilke ressurser de er avhengig av å opparbeide (Lien et al., 2016).

#### **Kunder/mottakere**

Kunder er neste ledd i strategikartet. I denne oppgaven vil det være hensiktsmessig å se på kunder som mottakere av bistand ettersom disse er avgjørende for å oppnå ønskelige resultater. For å kunne bidra til Bærekraftsmålene må bistanden rettes mot spesifikke grupper og samfunn. Ved eksperimentell bistand, slik som i kapittel 5, burde også grupper og/eller områder som kan måles prioriteres. Dette legger føringer for hvilke “kunder” – bistandsmottakere – Norad burde prioritere i strategien sin og når disse burde prioriteres.

#### **Aktiviteter**

Aktivitetene som kreves for å nå disse kundene er tildeling av midler til organisasjoner som opererer innenfor ønskede fokusområder. Dette er noe Norad styrer selv i dag gjennom

utlysninger, tildelingsbrev og fokus på Bærekraftsmålene (Norad, 2021). For å oppnå robuste resultater kreves dog en omlegging av Norads tilnærming og hvilke krav som stilles til søkerorganisasjoner. Norad må legge om aktivitetene både gjennom bedre kunnskapsformidling, bedre oppfølging og strengere krav til søkerorganisasjoner om robust metodikk og kvalitetsikring. Også dette må gjenspeiles i Norads strategi for valg av prioriterte mottakere av midler, både for å styre Norads virksomhet i riktig retning og for å utvise transparens i prioriteringene av finansiering.

### **Ressurser**

Dersom disse aktivitetene skal kunne gjennomføres i samsvar med ønsket om å sette penger i arbeid på en mer strategisk måte, må robuste målemetoder identifiseres som en sentral ressurs i Norads strategi og strategikart. Dette fordrer imidlertid omdisponeringer av midler, slik at evaluering og kvalitetssikring prioriteres i høyere grad. Dette kan innebære å stille strengere krav eller øremerke deler av midler til spesifikke evalueringsformal. For at dette skal gjøres må Norad besitte kompetanse til å følge opp organisasjoner, hvordan de evaluerer prosjekter og stille høyere krav til kvalitetssikring av resultater. Dette befester seg videre i et behov for humankapital med kompetanse for resultatmåling i Norad. Kompetanse kan dog kompenseres ved hjelp av endrede rutiner, opplæring, retningslinjer og bruk av eksterne evalueringsbyråer.

#### **7.1.1 Betydning for Norad og sirkularitet**

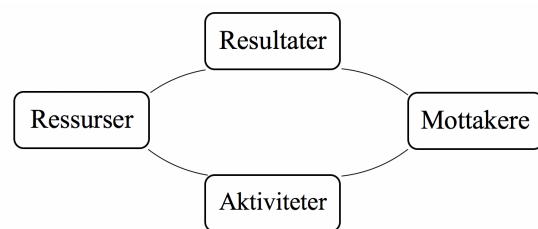
Ved gjennomgang av Norads historikk, retningslinjer og krav til søkerorganisasjoner kommer det tydelig frem hvordan hverken Norad eller søkerorganisasjoner i dag har overskudd til å kvalitetssikre bistand gjennom kontrollgrupper. Som gjennomgått i kapittel 6 vil dette kunne gi utslag på mange ulike måter, men på bunntinjen blir dette stående som manglende dokumentasjon på hvorvidt norske bistandsmidler har blitt satt til god bruk. Strategikartet over synliggjør dermed de presserende tiltakene Norad må implementere i sitt virke og dermed integrere i sin langsiktige strategi. Strategien vil da fungere som førende for implementering av tiltak i tiden fremover. Kun ved å utarbeide en robust strategi i tråd med overnevnte strategikart, vil Norad evne å gjennomføre de nødvendige aktivitetene og treffe de riktige mottakerne for å oppnå ønskelige resultater.

En endring i Norads målemetoder vil kunne ha flere positive utfall utover kun pålitelige

resultater. Ved implementering av robuste metoder vil Norad kunne optimalisere bistandsarbeidet sitt ved å replikere gode prosjekter, eliminere kontraproduktive praksiser og optimalisere drift. Dette vil kunne føre til en mer effektiv ressursallokering og mer effekt per krone. Videre risikerer Norad ved stillstand å svekke norsk bistands omdømme. Bistandssektoren som helhet har allerede blitt tungt kritisert for lite robust bruk av midler. Når Norad nå nedfeller mål i sin strategi om at *fakta har makta*, fordrer dette at mottoet følges opp (Norad, 2021b). Dersom Norad implementerer de foreslåtte tiltakene vil de kunne distansere seg fra denne kritikken og oppnå et styrket omdømme. Med stadig voksende BNI vil det for fremtiden være naturlig å forvente en diskusjon om hvorvidt bistandsbudsjettene også bør vokse i samme takt (SSB, 2014). Økt legitimitet og effekt vil da kunne styrke saken for utviklingen i budsjettpostene. På denne måten vil de foreslåtte metodegrepene kunne spille en viktig rolle i å bevare den relative størrelsen på det norske bidraget. Som et siste moment vil en slik skissert praksisendring kunne bidra til å stille strengere krav til andre aktører og lands bistandsarbeid gjennom demonstrasjon av best practice.

### Sirkularitet

Ved å implementere robuste målemetoder som en ressurs er det mulig å tilpasse strategikartet fra en lineær til en sirkulær resultatkjede. Vanligvis vil strategikartet nemlig benyttes fra ressurser til resultater eller resultater til ressurser (Lien et al.,



**Figur 7.1:** Sirkulært strategikart

2016). Ved å gjøre dette kan man som forklart se sammenhengene mellom leddene, der ressurser åpner for aktiviteter, som åpner for kunder og som deretter medfører resultater. Et vanlig strategikart kan da benyttes i begge retninger, men stanser likevel i hver ende. Vi mener derimot at robuste resultater i Norads tilfelle vil kunne føre til legitimitet, tillit og et bedre omdømme, som igjen vil være en ressurs. Dette vil som tidligere nevnt kunne lede til ytterligere ressurser, som økte midler, ved at styrket omdømme legitimerer fortsatt økte midler i takt med BNI. På denne måten sluttet sirkelen, og ressurser fører igjen til aktiviteter, som fører til kunder, som fører til resultater, som igjen fører til ressurser og en sirkulær årsakssammenheng.

Med dette tankesettet vil en omlegging av Norads virksomhet ikke kun medføre kostnader,

men også føre til reelle og approprierbare ressurser. Nye metoder vil da gi ringeffekter som bidrar til nye ressurser, og disse vil igjen plukkes opp i kretsløpet. Dette blir videre et argument i avveining av balansen mellom å hjelpe og å måle. En omlegging av Norads resultatmåling og krav vil på sikt kunne gi Norad forutsetninger og midler til å sette mer penger i arbeid og dermed hjelpe flere. Ved en utbygging av deres strategi med nevnte punkter, kan Norad oppnå en effektiv implementering av robuste metoder i deres virke. Ved å gjøre dette vil Norad videre være bedre rustet til å oppfylle et sentralt mål i langtidsstrategien, nemlig å bidra til Bærekraftsmålene, samtidig som fakta har makta.

## 7.2 Implementering av robuste metoder i Norads strategi og virke

Skal Norad endre dagens evalueringspraksis og inkorporere mer robuste metoder, bør dette som forklart forankres i Norads strategi og langsiktige arbeidsplaner. I organisasjonens strategidokument skisseres det også en villet utvikling mot mer kunnskaps- og faktabasert arbeid. Under slagordet “fakta har makta” har de selv satt opp fem overskrifter som skal prege Norads utvikling i inneværende tiår (Norad, 2021b):

1. Sette penger i arbeid på en mer strategisk måte
2. En sentral partner innen bærekraftig utvikling
3. Bidra til en grønnere verden
4. Være nyskapende i utvikling av bistanden
5. Styrke og systematisere utvikling, deling og bruk av kunnskap i Norad

I dette langsiktige strategiarbeidet harmonerer punktene 1 og 5 svært godt med anbefalingene fra denne oppgaven. Punktene fremstår som en erkjennelse av at det nødvendig å forbedre den kunnskapsbaserte tilnærmingen til hvordan bistandsarbeid velges ut, utføres og vurderes. Til tross for denne erkjennelsen, fremstår like fullt Norads strategi som noe uklar. Her burde det fremgå tydeligere av strategien at man ønsker å legge seg på en linje der bruk av rigorøse metoder med kontrollgrupper faktisk er en viktig del av verktøykassen. Dette betyr igjen at kravene også må inn i dokumenter som *Results Management in Norwegian Development Cooperation - a practical guide* og *Guidance note*

*on risks and results management for applicants* (Norad, 2008)(Norad, 2020b).

Skal rigorøse målemetoder bli en del av den nye standarden, må Norads ledelse sørge for at disse implementeres på en god måte. Videre i dette underkapittelet vil noen avgjørende faktorer for en vellykket implementering i strategi og virke gjennomgås.

### **Internkommunikasjon**

For å vellykket implementere slike endringer må den langsiktige verdien av rigorøse metoder tydeliggjøres i strategien, både til direktoratet som helhet og til relevante partnere. Norad har allerede identifisert deling og bruk av kunnskap som et utviklingsområde i sin strategi, men direktoratets nåværende gullstandard for måling og kunnskapsbruk er langt fra det vi vil anse som robust (Norad, 2021a)(Norad, 2020b). Derfor er det særlig viktig å løfte frem hvorfor ytterligere robuste metoder er avgjørende slik at Norad og alle deres ansatte kan forstå viktigheten av eget arbeid, hvorfor en omlegging av metoder er nødvendig og hvilke verdier dette vil skape på lang sikt. Ved å underbygge dette med tidligere argumentasjon rundt en sirkulær resultatkjede, samt formidle dette til alle berørte parter, vil Norad også ha større sjanse for en vellykket endringsprosess (Neill, 2018).

### **Delt ambisjon med politiske og organisatoriske stakeholdere**

Skal Norad få aksept og forståelse for at en større andel av bistandsmidlene skal gå til datainnsamling og måling, burde dette også forankres politisk og organisatorisk. Her må Norad i første omgang gå i dialog med Utenriksdepartementet for å sørge for at det også er politisk vilje til å følge opp Norads egne strategidokument. Dersom dette løses, må man som et neste steg sørge for å holde tredjepartsorganisasjoner involvert i utviklingen av nye krav og verktøy knyttet til rapportering og måling. Basert på erfaringene fra informasjonsmøter da Norad stilte mer systematiske krav til baseline-måling, bør Norad ta lærdom av dette og være tidlig ute med å informere sentrale aktører om kommende krav og hva som forventes av søkere som mottar prosjektmidler (Norad, 2015a). Disse grepene er viktige fordi en slik omdisponering av midler vil kreve at de sentrale aktørene i bistandsarbeidet spiller på lag ved en eventuell omlegging.

### **De store først, deretter de små**

Å innføre slike målemetoder er et forholdsvis kostbart grep og kommer til å kreve ekstra ressurser hos tredjepartsorganisasjonene, både i form av midler og ansatte. Det er naturlig å se for seg at kompetanse til å følge opp de skisserte kravene til resultatmåling vil



være mangelvare i de aller fleste søkerorganisasjoner. Da Norad begynte å stille krav om baseline-måling som en standard, gjorde man dette først i de store prosjektene med lengre varighet, som typisk gjennomføres av større organisasjoner (Norad, 2015a). Dette burde de også gjenta ved innføringen av nye metoder, fordi det vil være enklere å stille slike krav til tyngre organisasjoner. Norad kan på sin side bistå med å identifisere ressursnettverk som kan bistå med det metodiske arbeidet i praksis. Dette vil trolig være nødvendig fordi det kreves vesentlig høyere forståelse for statistikk og forskning for å arbeide med kontrollgrupper og rigorøs metode enn det gjør med de langt enklere baseline-kravene som stilles i dag. Målet på sikt bør være å normalisere bruken av mer rigorøse kvantitative metoder innen enkelte bistandsprosjekter, men i innføringsfasen bør man heller fokusere på at dette gjøres riktig enn at det gjøres raskt.

### **Norads egen kompetanse**

For at direktoratet skal være i stand til å stille gode krav til søkerorganisasjonene, må de selv også besitte kompetanse til å evaluere rapporter som gjennomføres av søkerorganisasjonene. Skal kravene til kvantitativ metode skjerpes, er det viktig at Norad også etter denne endringen besitter slik kompetanse. Norads ansatte later i dag til å besitte stor kompetanse – også når det kommer til kvantitative målemetoder (Norad, 2014). I utgangspunktet virker det dermed som at direktoratet burde være i stand til å ha en overordnet rolle ved slike skjerpede krav og at ansatte i eksempelvis Resultatavdelingen og Evalueringsavdelingen har kunnskapen som er nødvendig for å bedømme graden av kvalitet på prosjektevalueringer. Like fullt vil det være nødvendig for Norad å gjennomgå kompetansenivået internt i direktoratet og sørge for at relevante ansatte har tilstrekkelig kunnskap om nye målemetoder. Ved mangler kan dette løses med både faglig påfyll for eksisterende ansatte og gjennom nyansettelser hvor man kan stille krav til forkunnskaper. Tatt i betraktning hvor viktig en slik praksisendring fremstår for Norad, bør dette være en både nødvendig og overkommelig kostnad for direktoratet.

## **7.3 Når bør rigorøs kvantitativ metode anvendes?**

Hittil har vi vist at Norad gjennomgående har et måleproblem når det kommer til å dokumentere effekten av midlene de distribuerer. Ved å unnlate bruk av kontrollgrupper og kun se på effekter fra baseline til endline, klarer ikke evalueringene deres å troverdig

fastslå evidens (Ternström AB, 2020)(Riksrevisjonen, 2019). Mer robuste kvantitative metoder vil videre, som tidligere argumentert for, både kunne styrke effekten av bistand over tid og bidra til økt troverdighet og oppslutning knyttet til norske bistandsprosjekter. Implikasjonen blir dermed at Norad burde anse økt bruk av rigorøs kvantitativ metode som en investering i kvalitet og legitimitet. Med dette som bakteppe vil det være et logisk neste steg å foreslå at Norad burde implementere strengere krav til metodebruk i alle prosjekter. Vi anerkjenner dog at slike krav til data og metode er både tidkrevende og kostbare (Bryan, 2019). Dermed må nødvendige avveininger mellom kostnad og nytte gjøres før Norad selv tar stilling til policyimplikasjonene fremmet i denne oppgaven.

Som vi tidligere har diskutert i kapittel 6, finnes det absolutt prosjekter hvor kvalitativ metode har klare fordeler. I prosjekter hvor brukermedvirkning vil være avgjørende enten for hele eller innledende faser, vil kvalitative metoder har klare styrker (Hammarberg et al., 2016). Der brukere eller andre lokale oppfordres til å bidra med ideer, meninger og tanker, vil intervjuer og andre kvalitative verktøy kunne være mer hensiktsmessig å benytte. Når de benyttes riktig vil kvalitativ metode være viktig for å dokumentere at midler blir disponert riktig og at prosjektet utføres som det skal. Intervjuer og brukermedvirkning kan også vise seg å være gode verktøyer for å bidra til varsling, korrupsjonsbekjempelse og compliance. Årsaken er tettere dialog med de involverte i bistandsprosjektet og nærmere oppsyn med gjennomføringen.

Videre er det ingen automatikk i at rigorøse målemetoder er nødvendig i alle tilfeller. Norad bidrar over tid til å realisere mange prosjekter som ligner eller som er identiske med hverandre (Norad, 2021). Har man allerede funnet gode arbeidsmetoder og tilnærminger til spesifikke metodikker, vil kostbar bruk av målemetoder kunne føre til ineffektivitet og til å trekke viktige ressurser bort fra selve bistanden. Skal det gjennomføres prosjekter med enkle målindikatorer, vil tiden brukt på omfattende forskning kunne oppfattes å være bortkastede ressurser. Det er sentralt å bevare respekt og fokus på å bidra mest mulig per krone tildelt. Samtidig har Norad lite belegg per i dag for å fastslå at metodene deres har god effekt. Momentet om gode, eksisterende arbeidsmetoder, vil derfor være mer relevant i fremtiden.

Sammen utgjør innsiktsarbeid (for eksempel demokratiseringsprosjekter eller brukermedvirkningsprosjekter) og enkle realiseringsprosjekter (eksempelvis bygg

av skoler og saniteringsanlegg) en vesentlig andel av Norads arbeid (Norad, 2021). Disse vil det i mange tilfeller ikke være nødvendig å binde opp store ressurser for å drive omfattende kvantitativt målearbeid på. Like fullt vil saken for rigorøs kvantitativ metode stå sterkt innenfor en rekke prosjekter som finansieres av Norad. På dette grunnlaget er det særlig to områder som skiller seg ut som særskilt egnet for bruk av rigorøs kvantitativ metode: Eksperimentell og presedensdannende bistand.

### **Eksperimentell bistand**

I mange tilfeller av bistand finnes det talløse måter å nå et mål på. Ulike intervensjoner, ulike demografiske grupper og andre valg som gjøres ved utformingen av et bistandsprosjekt, er alle variabler som kan få store utslag på utfallet. Ofte ser man bistandsprosjekter som utføres ved bruk av nye metoder og som utforsker nye tematikker med nye virkemidler. Et rimelig krav til slike bistandsprosjekter vil være å fastslå at dersom man skal utforske nytt terreng, bør man tegne opp et kart for de som vil vurdere å legge ut på den samme reisen i ettertid. Her er det i våre øyne behov for å benytte seg av mer robuste kvantitative metoder. Dersom man skal innvilge midler til nye metoder, burde man også kunne dokumentere om de virker.

Vi mener derfor at krav til rigorøse kvantitative metoder vil ha gode effekter på sikt om man skal utføre prosjekter med eksperimentell utforming. Hva som skal regnes som eksperimentelt vil kreve en grundigere gjennomgang enn det denne oppgaven tillater. Likevel bør det være en tommelfingerregel å regne det som eksperimentelt om det finnes mange ulike veier til målet og det ikke har vært utført lignende prosjekter tidligere som man kan replikere eller basere seg på. I disse tilfellene bør Norad i det minste vurdere om et slikt prosjekt bør finansieres med forbehold om kvantitativ målemetode med høyere robusthetsgrad enn de nåværende metodene sikrer.

### **Presedensdannende arbeid**

Til tross for at bistand fordeler seg på store antall mål og grupper, er det mange prosjekter som søker å løse de samme problemene. Er det tilstrekkelig mange som arbeider for å nå de samme målene, vil kunnskapsdeling knyttet til metoder og tiltak være et udiskutabelt gode. Det samme vil være tilfelle dersom det utføres prosjekter som man kan forvente at vil bli kopiert eller virke som inspirasjon i senere tid. I disse tilfellene bør Norad også erkjenne at de eksisterende metodene som brukes for måling, ikke er tilstrekkelig gode

slik det skisseres i de eksterne evalueringene av desentraliserte evalueringer som Norad har mottatt og som beskrevet i Kapittel 4. Dersom Norad har grunn til å tro at flere lignende prosjekter vil gjennomføres i løpet av rimelig kort tid, vil det være klokt som et minstevalg å benytte seg av rigorøs kvantitativ metode hos deler av disse prosjektene. Årsaken er at det kan bidra til å utvikle beste praksis både på tvers av prosjekter og på tvers av landegrenser. Alle er tjent med at bistand fungerer, og Norge står i en særskilt stilling i form av at man her kan sette av midler til evaluering samtidig som man bidrar med mer midler enn det FN-landene er enige om som felles mål.

## 7.4 Kan norske bedrifter lære noe av våre funn?

Norsk nærings- og organisasjonsliv opererer i et landskap hvor ESG og bærekraft tar stadig større del i den daglige driften (Ditlev-Simonsen, 2021). Viljen og evnen til å bidra i positiv forstand til å bekjempe klimaendringer, ulikhet og andre omfattende samfunnsutfordringer virker historisk stor. Dette medfører at mange bedrifter og organisasjoner tar grep og implementerer grønnere og mer sosiale tiltak i egen drift. Stadig flere selskaper stiller krav til leverandører og til partnere, og det later til at forbrukerne i økende grad ønsker seg ESG-orienterte tiltak fra selskapene de møter (Andreassen og Lervik-Olsen, 2021). Implikasjonene fra oppgaven kan dermed også gjøre seg gjeldende for flere enn Norad.

I takt med forventninger fra stakeholdere og fra politisk hold, har utvikling i arbeid med bærekraft blitt normalt innhold i norske bedrifters årsrapporter (PWC, 2020). Også her er det vanlig med forholdsvis enkle baseline/endline-analyser og av åpenbare grunner mindre fokus på evidens og forklaringskraft. Å forlange et like rigorøst måleregime fra norske bedrifter som rapporterer på ESG som vi forlanger fra Norad, ville vært urimelig kostbart for norske bedrifter som allerede i stor grad gjennomfører ESG-arbeidet som frivillig innsats. Like fullt burde disse bedriftene merke seg hvilke feilkilder som kan oppstå ved en slik metodisk tilnærming. I de statistiske eksemplene vi presenterer i denne oppgaven, og ikke minst i historien om Millennium Villages, fremkommer det hvor feil man kan ta i analyser dersom man ikke tar hensyn til utviklingen i samfunnet for øvrig – eller i en kontrollgruppe hvor man holder alle andre faktorer like.

Slike måleproblemer burde man være observant på. Blant de mer opplagte konsekvensene kan dårlige måleredskaper føre til over- eller underrapportering som kan gi ubehagelige

overraskelser over tid. Mer alvorlig vil det være dersom dårlig måling gjør at man over tid satser på feil hest ved å velge ineffektive eller skadelige metoder i sin søken etter forbedringer på ESG-feltet. Dette kan medføre store kostnader, negativ oppmerksomhet og krevende snuoperasjoner for bedrifter som i utgangspunktet har operert i god tro om at de bidrar positivt. Presis bruk av KPIer og gode metoder for resultatmåling bør derfor være noe flere i nærings- og organisasjonsliv er opptatte av og særlig noe som følger med i bestillingen når man skal få satt opp mål for ESG.

## 7.5 Oppsummering og delkonklusjon

I dette kapitlet har vi gjennomgått hvordan Norad kan implementere sterkere målemetoder gjennom sin langtidsstrategi. Å trappe opp de metodiske kravene til bistandsprosjekter kan vise seg å være svært kostnadskrevende, men det vil også trolig gagne Norad og norsk bistand over tid. Troverdige resultater og oppslutning rundt bistandsarbeidet er ressurser for Norad som er viktig for å opprettholde avsetningsnivået knyttet til bistand. Skulle det bli mer kjent i befolkningen at man i dag ikke kan fastslå på troverdig vis at bistand har effekt, vil dette tilsvarende kunne svekke både legitimiteten og oppslutningen rundt det norske bistandsarbeidet og følgelig svekke bevilgningene.

Tidligere i oppgaven har vi vært inne på det norske 1-prosentsmålet som sier at 1 prosent av BNI skal settes av til bistand. Med lav grad av evidens knyttet til resultater av bistand vil det kunne bli vanskeligere å rettferdiggjøre stadig økende bistandsbudsjetter. Betyr dette i så tilfelle at norsk bistand burde reduseres? Vi mener at svaret er nei. Samtidig burde Norge som en sterk pådrivernasjon for internasjonalt utviklingsarbeid ha et mer bevisst forhold til resultater og stille strenge krav til kvalitet. Mangel på reell kvalitet og resultater kan gi (berettiget) grobunn for bistandsskepsis, og bistandens natur gjør den kritisk avhengig av oppslutning hos både giver- og mottakerland. Så lenge Norad arbeider for å forbedre kvaliteten på målemetodene sine, er det likevel fremgang som i større grad rettferdiggjør bruken av midler og som bringer arbeidet nærmere målet for kvalitet.

## 8 Konklusjon

I denne oppgaven har vi vist at Norad har et betydelig forbedringspotensial når det kommer til dokumentasjon av bistandsmidlenes effekt. Gjennom flere eksterne rapporter er det påpekt at brorparten av evalueringene knyttet til Norad-finansierte bistandsprosjekter, ikke er tilstrekkelig gode metodisk. Som vist i oppgaven, kan dette ha mange ulike utslag for det norske bistandsarbeidet. Fellesnevneren for alle disse utslagene er imidlertid muligheten for at man ikke finansierer de riktige eller beste prosjektene. Som vi har gjennomgått, er årsaken i stor grad at Norad benytter seg av forholdsvis enkle kvantitative og kvalitative metoder for å dokumentere effekt. For de kvantitative metodetilfellene innebærer dette ofte at det kun samles inn data for populasjonen som mottar behandling og at man måler utvikling fra baseline til endline uten bruk av kontrollgrupper. Implisitt bygger disse metodene i praksis på antakelser om nullutvikling i øvrig populasjon når man måler utvikling. Med andre ord baserer Norad seg på at behandlingsgruppen ikke ville hatt noen utvikling uten at tiltaket hadde vært gjennomført. Som vi både vet og har vist, er dette et sjeldent tilfelle. For å sirkle tilbake til oppgavens aller første del, virker det som at Norad følger i fotsporene til Jeffrey Sachs med alle metodiske fallgruver det kan medføre.

I tillegg til å være godt beskrevet og gjennomgått av eksterne evalueringer, har vi også med metodiske eksempler i kapittel 3 og 5 vist hva utfallet kan bli med Norads metode. Ikke bare fører dagens metodebruk til mindre robuste slutninger; det kan også føre til at man trekker direkte feil slutninger dersom man ikke har særdeles gode kunnskaper om forholdene man opererer i. Dette kan igjen føre til at man fortsetter å finansiere prosjekter med lav effekt, avslutter prosjekter med stor effekt og at man innordner norske bistandsmidler på en mindre effektiv måte over tid enn man ville gjort med mer presise målemetoder.

På bakgrunn av disse funnene har vi også illustrert Norads strategikart for å se på hvor eventuelle endringer må forekomme. Ved å opparbeide rigorøse, kvantitative målemetoder som en strategisk ressurs i organisasjonen, kan dette gi netto positive resultater for Norad og deres tjenestetilbud til tross for den forhøyede kostnaden det innebærer. Vi har illustrert dette ved bruk av det vi kaller et sirkulært strategikart, hvor forbedrede målemetoder gir økt presisjon i identifikasjonen av effektive formål, som igjen kan gi forbedret bistand.

---

Dette kan styrke legitimiteten til avsetninger til bistand og dermed også styrke saken for å opprettholde 1-prosentsmålet.

Til tross for at 1-prosentsmålet til tider er omdiskutert, er det bred politisk enighet om at det norske bidraget til internasjonal fattigdomsbekjempelse og utviklingsarbeid skal ligge på rundt dette nivået. Med den norske overoppfyllelsen av FNs mål for bistandsmidler, mener vi forholdene ligger godt til rette for at Norad skal kunne øke ressursbruken knyttet til resultatmåling. Mens bruk av RCT eller andre rigorøse, kvantitative målemetoder trolig vil medføre betydelige kostnader, bør det særlig kunne være rom for dette i en nasjon som allerede gir mer enn man har forpliktet seg til. Når vi i tillegg kun anbefaler slik bruk innenfor eksperimentell eller presedenssettende bistand, vil dette også begrense den ekstra kostnaden man må påregne.

I oppgaven har vi blant mange mulige målemetoder identifisert to metoder - RCT og DiD - Norad bør vurdere å implementere i sine krav. Ved å benytte seg av disse i særlig eksperimentelle og presedensdannende bistandsprosjekter, vil direktoratet oppnå mer robuste funn. Kort oppsummert vil fordelene av mer robuste funn være bedret ressursallokering, økt effekt av bistand og høyere legitimitet både for Norad og for bistandsarbeid generelt. Det er i oppgaven også skissert hvilke hensyn Norad bør ta ved implementering av slike målemetoder, både i deres langtidsstrategi og virke. Dermed mener vi at oppgaven både viser at hypotesen som ble stilt innledningsvis (at Norad med dagens målemetoder risikerer å gå glipp av avgjørende data) stemmer, og at oppgavens problemstilling er besvart utfyllende.

Med dette mener vi at vi har dokumentert at det finnes bedre tilgjengelige målemetoder enn de som benyttes i dag og at det vil være tilrådelig å trappe opp kvaliteten på resultatmålingen. I tillegg har vi pekt på andre fordeler ved en slik implementering, og vi har påpekt flere helt konkrete negative utfall som kan forekomme dersom man ikke gjør disse grepene. Med denne oppgaven håper vi å ha levert et bidrag til debatten om Norads videre utvikling og understreket hvor viktig det er å prioritere midler til metode selv om det allerede er kamp om begrensede ressurser. Vi håper også Norad tar til seg disse anbefalingene, da vi mener det vil kunne føre til viktige fremskritt både for Norad og for bistandsarbeidet internasjonalt. Forhåpentligvis kan Norad fortsette å bidra mer effektivt – én evaluering om gangen.

## 8.1 Robusthetsanalyse

Avslutningsvis er det hensiktsmessig å gjennomføre en robusthetsanalyse der oppgavens svakheter gjennomgås. I den allmenne fortolkningen av funn og resultater fra denne oppgaven, er det flere momenter som burde nyanseres avslutningsvis.

Først bør det nevnes at funnene og deler av argumentasjonen i denne oppgaven med all sannsynlighet ikke er ukjent for Norad, men snarere noe de bevisst har tatt stilling til tidligere så vel som i sin forelagte tiårsstrategi. Vår konklusjon vil like fullt stå seg, ettersom vi mener det vitner om passivitet og mangel på nødvendige tiltak for at Norad kan møte sine ansvarsområder og målsettinger. Det er også verdt å reflektere rundt at diskusjonen rundt Norads målemetoder er en problematikk som har blitt drøftet siden 2009 - og trolig enda lengre tilbake i tid. Det har altså allerede gått over et tiår uten å få på plass adekvate målemetoder. Dette styrker utvilsomt grunnlaget for å hevde at det eksisterende omstillingsarbeidet går for sakte. Videre burde det trekkes frem at oppgaven i liten grad drøfter det praktiske rundt implementering av slike endringer. Vi har forståelse for at dette vil ta tid og innebære drastiske byråkratiske endringer, men igjen er viktigheten av implementeringen noe vi mener nødvendiggjør umiddelbar handling. Det ligger ikke innenfor denne oppgavens formål og omfang å skissere en spesifikk implementeringsplan for Norad, men snarere påpeke områder med forbedringspotensial og mane til endring i direktoratets metoder.

Det kan også innvendes at vi i litteraturgjennomgang og analyse har benyttet DiD, som er en mindre omfattende tilnærming enn gullstandarden RCT. Til tross for at vi anbefaler bruk av RCT der det lar seg gjennomføre, finnes det en rekke ulike metoder for statistiske analyser og robust resultatmåling. Det overordnede poenget med krav om kontrollgrupper står likevel uavhengig av målemetode og kan dermed regnes som robust. Vi tar også høyde for eventuelle feil i illustrerte eksempler i form av potensielt utelatte relevante variabler, samt at prosjektet kan vise et begrenset bilde. Dette er et godt eksempel på når gode kvalitative indikatorer i større grad kan å øke Norads sannsynligheten for å knytte effekt til prosjekt. Det sentrale poenget her er ikke å argumentere for én spesifikk målemetode, men å vise at kontrollgrupper er avgjørende og deretter anbefale anerkjente metoder. Uten bruk av kontrollgrupper har ikke Norad per i dag mulighet til å treffe metodisk funderte konklusjoner, og det er dette som gjennomgående argumenteres for i oppgaven. Analysen



og medfølgende argumentasjonsrekke kan dermed regnes som robust.

I analysen har vi videre utledet et strategikart der vi argumenterer for en sirkulær årsakssammenheng hvor robust måling kan resultere i styrket legitimitet og økte midler i takt med BNI. Dette er en ren antakelse, og vi har ingen empiriske holdepunkter for å si at økt legitimitet vil ha en ren påvirkning på bevilgningsstørrelser. Vi anser dette som delvis robust ettersom bistandssektoren som helhet har mottatt kritikk over tid og at å motbevise i hvert fall deler av denne trolig vil kunne resultere i positive effekter for Norad.

Til slutt bør det nevnes at Norad kan ha foretatt mulige tiltak siden den eksterne evalueringen utgitt i 2020 og at disse ikke er tatt høyde for. Strategien for 2030 (Fakta har makta) ble lansert i mars 2021, og kan inneholde konkrete grep vi ikke har fått tilgang til enda. All annen tilgjengelig informasjon fra Norads nettsider og nyheter vedrørende deres virksomhet vitner imidlertid ikke om noen nært forestående endringer. Selv om tiårsstrategien fremhever bedre kunnskapsformidling og å sette penger i arbeid på en mer strategisk måte er det i tilgjengelige manualer og retningslinjer lite som legger til rette for strengere krav og resultatmåling. Vi anser det dermed som lite sannsynlig at Norad har tatt tilstrekkelige grep. Oppgaven som helhet kan følgelig regnes som robust, men overnevnte punkter burde likevel tas høyde for ved eventuell anvendelse.

## Referanser

- Andreassen, T. W. og Lervik-Olsen, L. (15. januar 2021). Bærekraft tiltrekker seg flere kunder. Hentet 26. mai 2021, fra: <https://www.dagensperspektiv.no/2021/baerekraft-tiltrekker-seg-flere-kunder>.
- Angrist, J. D. og Pischke, J.-S. (2014). *Masterings metrics*, volume 1. Princeton University Press.
- Athey, S. og Guido W, I. (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives*, 31(2):3–32.
- Bandiera, O., Buehren, N., Burgess, R., Goldstein, M., Gulesci, S., Rasul, I., og Sulaiman, M. (2020). Women's empowerment in action: Evidence from a randomized control trial in africa. *American Economic Journal: Applied Economics*, 12(1):210–259.
- Banerjee, A. (2005). New development economics' and the challenge to theory. *Economic and Political Weekly*, 40(40):4340–4344. Hentet 9. april 2021, fra: <http://www.jstor.org/stable/4417230>.
- BRAC (2020). Press room: About brac. Hentet 26. mai 2021, fra: <http://www.brac.net/index.php/partnership>.
- Bredvei, D. (2021). Norad: Pandemien gjør verden "mergavmild. Hentet 20. april 2021, fra: [https://www.nrk.no/urix/norad\\_-\\_pandemien-gjor-oss-mer-gavmilde\\_-\\_ikke-mindre-1.15463059](https://www.nrk.no/urix/norad_-_pandemien-gjor-oss-mer-gavmilde_-_ikke-mindre-1.15463059).
- Bryan, K. (2019). What randomisation can and cannot do: The 2019 nobel prize. Hentet 1. mars 2021, fra: <https://voxeu.org/article/what-randomisation-can-and-cannot-do-2019-nobel-prize>.
- Bump, J., Clemens, M., Demombynes, G., og Haddad, L. (mai 2012). Concerns about the millennium villages project report. *The Lancet*, 379. 10. april 2021, fra: [https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736\(12\)60787-9.pdf](https://www.thelancet.com/pdfs/journals/lancet/PIIS0140-6736(12)60787-9.pdf).
- Cameron, D. B., Mishra, A., og Brown, A. N. (2015). The growth of impact evaluation for international development: how much have we learned? *Journal of Development Effectiveness*, 8(1):1–21.
- Clemens, M. og Demombynes, G. (september 2013). The new transparency in development economics: Lessons from the millennium villages controversy. 15. april 2021, fra: [https://www.cgdev.org/sites/default/files/Clemens-Demombynes-new-transparency\\_2.pdf](https://www.cgdev.org/sites/default/files/Clemens-Demombynes-new-transparency_2.pdf).
- Cooper, M., Kharas, H., og Müller, B. (22. Oktober 2020). Are we on track to end global hunger. Hentet 6. april 2021, fra: <https://www.brookings.edu/blog/future-development/2020/10/22/are-we-on-track-to-end-global-hunger/>.
- Coughlan, S. (09. Juli 2019). Global literacy targets already off track - unesco. Hentet 6. april 2021, fra: <https://www.bbc.com/news/education-48909386>.
- Devex (2020). Millennium villages project. Hentet 29. mai 2021, fra: <https://www.devex.com/organizations/millennium-villages-project-47592>.
- Ditlev-Simonsen, C. D. (15. januar 2021). Esg-rapportering: Det holder ikke lenger med de hyggelige historiene. Hentet 26. mai 2021,

- fra: <https://finansavisen.no/nyheter/debattinnlegg/2021/05/07/7665066/esg-rapportering-det-holder-ikke-lenger-med-de-hyggelege-historiene>.
- Easterly, W. og Pfutze, T. (2008). Where does the money go? best and worst practices in foreign aid.
- Easterly, W. og Sachs, J. (2006). The big push déjà vu: A review of jeffrey sachs's the end of poverty: Economic possibilities for our time". *Journal of Economic Literature*, 44(1):96–105. Hentet 26. mai 2021, fra: <http://www.jstor.org/stable/30032297>.
- Eyisi, D. (2016). The usefulness of qualitative and quantitative approaches and methods in researching problem-solving ability in science education curriculum. *Journal of Education and Practice*.
- FN (08. Januar 2021). Fns bærekraftsmål. Hentet 6. april 2021, fra: <https://www.fn.no/om-fn/fns-baerekraftsmaal>.
- Foundation, G. (2021). Givewell - list of causes to donate to. Hentet 21. mai 2021, fra: <https://www.givewell.org/aggregator/sources/7>.
- Foundation, N. (2019). Nobel prize economics 2019. Hentet 28. mai 2021, fra: <https://www.nobelprize.org/prizes/economic-sciences/2019/press-release/>.
- Frafjord, J. (2020). Skal vise frem bistandsbudsjettet. Hentet 10. april 2021, fra: <https://finansavisen.no/nyheter/aktuell/2020/12/01/7593733/krf-ulstein-skal-vise-frem-bistandsresultater>.
- Hagen, R. J. (2018). Mangler det 190 millioner eller vilje til reform. Hentet 10. april 2021, fra: <https://www.aftenposten.no/meninger/kronikk/i/8wdRr1/mangler-det-190-millioner-eller-vilje-til-reform-rune-jansen-hagen>.
- Hammarberg, K., Kirkman, M., og de Lacey, S. (2016). Qualitative research methods: when to use them and how to judge them. *Human Reproduction*.
- Hunger Notes (2018). 2018 world hunger and poverty facts and statistics. Hentet 6. april 2021, fra: <https://www.worldhunger.org/world-hunger-and-poverty-facts-and-statistics/#:~:text=The%20United%20Nations%20Food%20and,lower%2Dmiddle%2Dincome%20countries>.
- Itad og Christian Michelsens Institutt (april 2014). Can we demonstrate the difference that norwegian aid makes? Hentet 1. april 2021, fra: <https://www.oecd.org/derec/norway/Can-We-Demonstrate-the-Difference-that-Norwegian-Aid-Makes.pdf>.
- Itad og Christian Michelsens Institutt (januar 2017). The quality of reviews and decentralised evaluations in norwegian development cooperation. Hentet 1. april 2021, fra: <https://www.norad.no/contentassets/41a0207f132c4e69a338791bac88dda7/the-quality-of-reviews-and-decentralised-evaluations-in-norwegian-development-cooperation.pdf>.
- Lien, L. B., Knudsen, E. S., og Baardsen, T. (2016). *Strategiboken*, volume 1. Vigmostad Bjørke AS.
- Miguel, E. og Kremer, M. (2004). Identifying impacts on education and health in the presence of treatment externalities. *Econometrica*, 72(1):159–217.

- Neill, M. S. (August 2018). Change management communication: Barriers, strategies messaging. Hentet 5. mai 2021, fra: [https://www.researchgate.net/profile/Marlene-Neill/publication/327057178\\_Change\\_Management\\_Communication\\_Barriers\\_Strategies\\_Messaging/links/5b757e2845851546c909dc07/Change-Management-Communication-Barriers-Strategies-Messaging.pdf](https://www.researchgate.net/profile/Marlene-Neill/publication/327057178_Change_Management_Communication_Barriers_Strategies_Messaging/links/5b757e2845851546c909dc07/Change-Management-Communication-Barriers-Strategies-Messaging.pdf).
- Norad (16. mars 2021b). Strategi mot 2030. Hentet 1. april 2021, fra: <https://www.norad.no/globalassets/publikasjoner/publikasjoner-2021/norads-strategi-mot-2030.pdf>.
- Norad (17. Juli 2014). Evalueringsmetoder. Hentet 10. april 2021, fra: <https://www.norad.no/evaluering/om-evaluering/evalueringmetoder/>.
- Norad (18. februar 2015b). Norads hovedoppgaver. Hentet 1. april 2021, fra: <https://www.norad.no/om-norad/norads-hovedoppgaver/>.
- Norad (2008). *Results management in Norwegian development cooperation*. Norad, Oslo, Norge.
- Norad (2011c). Årsrapport 2011 - evaluering av norsk utviklingssamarbeid. Hentet 1. april 2021, fra: [https://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/evalueringssavdelings-filer/arsrapport-2011\\_evaluering-av-norsk-utviklingssamarbeid.pdf](https://www.norad.no/globalassets/import-2162015-80434-am/www.norad.no-ny/filarkiv/evalueringssavdelings-filer/arsrapport-2011_evaluering-av-norsk-utviklingssamarbeid.pdf).
- Norad (2012). Hvilke prosjekter tilfredsstill kriteriene for å være godkjent som utviklingshjelp. Hentet 20. mars 2021, fra: <https://www.norad.no/tilskudd/sok-stotte/hvilke-prosjekter-tilfredsstill-kriteriene-for-a-vare-godkjent-som-utviklingshjelp/>.
- Norad (2015a). Krav til baseline for sivilsamfunnsorganisasjoner. Hentet 12. april 2021, fra: <https://www.norad.no/tilskudd/sok-stotte/sivilt-samfunnfrivillige-organisasjoner/krav-til-baseline-for-sivilsamfunnsorganisasjoner/>.
- Norad (2020a). Barn utsatt for omsorgssvikt har fått bedre skoletilbud. Hentet 18. april 2021, fra: <https://resultater.norad.no/resultathistorie/barn-utsatt-for-omsorgssvikt-har-faatt-bedre-skoletilbud>.
- Norad (2020b). *Guidance note on results and risk management for applicants*. Norad, Oslo, Norge.
- Norad (2021). Bistandsresultater. Hentet 10. mars 2021, fra: <https://resultater.norad.no/ sektor>.
- Norad (2021a). *RAM-modellen*. Norad, Oslo, Norge.
- Norad (27. oktober 2011a). Norsk utviklingspolitikk. Hentet 1. april 2021, fra: <https://www.norad.no/om-bistand/norsk-utviklingspolitikk/>.
- Norad (28. september 2011b). Slik styres norad. Hentet 1. april 2021, fra: <https://www.norad.no/om-norad/slik-styres-norad-/>.
- Norad, Utenriksdepartementet, og Klima- og miljødepartementet (2021). Om portalen. Hentet 5. april 2021, fra: <https://resultater.norad.no/om-portalen>.
- NTB (26. september 2007). Lover 1 milliard dollar til kvinnebistand. Hentet 27. mai 2021, fra: <https://www.bt.no/nyheter/utenriks/i/1v6LB/lover-1-milliard-dollar-til-kvinnebistand>.

- OECD (16. april 2020). Aid by dac members increases in 2019 with more aid to the poorest countries. Hentet 6. april 2021, fra: <https://www.oecd.org/dac/financing-sustainable-development/development-finance-data/ODA-2019-detailed-summary.pdf>.
- OECD (2019). Official development assistance 2019. Hentet 6. april 2021, fra: [https://public.tableau.com/views/ODA-GNI\\_15868746590080/ODA2019?:display\\_count=y&publish=yes&:origin=viz\\_share\\_link?&:showVizHome=no#1](https://public.tableau.com/views/ODA-GNI_15868746590080/ODA2019?:display_count=y&publish=yes&:origin=viz_share_link?&:showVizHome=no#1).
- OECD (2021). Development finance standards. Hentet 26. mars 2021, fra: <https://www.oecd.org/development/financing-sustainable-development/development-finance-standards/>.
- Our World in Data (23. januar 2019). How many children are not in school? Hentet 6. april 2021, fra: <https://ourworldindata.org/how-many-children-are-not-in-school>.
- Palladium (23. september 2020). New report says sdgs won't be achieved until 2092 - now is the time for action. Hentet 6. april 2021, fra: <https://thepalladiumgroup.com/news/New-Report-says-SDGs-Won't-Be-Achieved-Until-2092-Now-is-the-Time-for-Action>.
- PWC (2020). Bærekraft 100. Hentet 29. mai 2021, fra: <https://www.pwc.no/no/tjenester/risk-advisory-services/barekraft-samfunnsansvar/pwc-barekraft-100-2020.pdf>.
- Regjeringen (2021). Rekordhøy norsk bistand. Hentet 10. april 2021, fra: [https://www.regjeringen.no/no/aktuelt/oecd\\_tall/id2844317/](https://www.regjeringen.no/no/aktuelt/oecd_tall/id2844317/).
- Riksrevisjonen (2019). Riksrevisjonens undersøkelse av informasjon om resultater av bistand til utdanning. Hentet 25. mai 2021 fra: [https://www.riksrevisjonen.no/globalassets/rapporter/no-2018-2019/bistand\\_utdanning.pdf](https://www.riksrevisjonen.no/globalassets/rapporter/no-2018-2019/bistand_utdanning.pdf).
- Routledge, R. (2020). Law of large numbers. Hentet 26. mai 2021, fra: <https://www.britannica.com/science/law-of-large-numbers>.
- Sachs, J. (2005). *The end of poverty: Economic possibilities for our time*, volume 1. Penguin Press.
- Sachs, J. (januar 1999). The big push, natural resource booms and growth. Hentet 23. april 2021, fra: <https://www.jeffsachs.org/journal-articles/rngjzfszgjsw32588n9g9tm6ydd6lj>.
- Solberg, E. (22. januar 2014). Jenters rett til utdanning. Hentet 27. mai 2021, fra: <https://erna.no/2014/01/22/jenters-rett-til-utdanning/>.
- SSB (2014). Offisiell norsk bistand, nivå og andel av bni. Hentet 15. mai 2021, fra: <https://www.ssb.no/natur-og-miljo/barekraft/offisiell-norsk-bistand-niva-og-andel-av-bni>.
- Ternström AB (oktober 2020). Quality assessment of decentralised evaluations in norwegian development cooperation (2018–2019). Hentet 1. april 2021, fra: <https://www.norad.no/contentassets/e7901444a02a4be97462a50773e535c/6.20-quality-assessment-of-decentralised-evaluations-in-norwegian-development-cooperation-2018-2019.pdf>.
- The World Bank (07. Oktober 2020). Poverty. Hentet 6. april 2021, fra: <https://www.worldbank.org/en/topic/poverty/overview#:~:text=The%20global%20extreme%20poverty%20rate,%245.50%20a%20day%20in%202017>.

- The World Bank (2017). Poverty headcount ratio. Hentet 6. april 2021, fra: <https://data.worldbank.org/indicator/SI.POV.DDAY?locations=1W&start=1981&end=2015&view=chart>.
- UN (2020). Ending poverty. Hentet 10. april 2021, fra: <https://www.un.org/en/global-issues/ending-poverty>.
- United Nations (2015a). Millennium goals. Hentet 25. mai, fra: <https://www.un.org/millenniumgoals/>.
- United Nations (27. september 2015b). Transforming our world: the 2030 agenda for sustainable development. Hentet 27. mai 2021, fra: <https://sdgs.un.org/2030agenda>.
- Utenriksdepartementet (26. mars 2021). Utenriksdepartementets tildelingsbrev til direktoratet for utviklingssamarbeid. Hentet 1. april 2021, fra: [https://www.regjeringen.no/contentassets/82928efb0ef545cd904e03ab51b51470/norad\\_tildelingsbrev\\_2021.pdf](https://www.regjeringen.no/contentassets/82928efb0ef545cd904e03ab51b51470/norad_tildelingsbrev_2021.pdf).
- Webber, S. og Prouse, C. (2018). The new gold standard: The rise of randomized control trials and experimental development. Hentet 10. mai 2021, fra: [https://www.researchgate.net/profile/Marlene-Neill/publication/327057178\\_Change\\_Management\\_Communication\\_Barriers\\_Strategies\\_Messaging/links/5b757e2845851546c909dc07/Change-Management-Communication-Barriers-Strategies-Messaging.pdf](https://www.researchgate.net/profile/Marlene-Neill/publication/327057178_Change_Management_Communication_Barriers_Strategies_Messaging/links/5b757e2845851546c909dc07/Change-Management-Communication-Barriers-Strategies-Messaging.pdf).
- Woolridge, J. M. (2012). *Introductory Econometrics*, volume 5. South-Western College Publishing.

# Appendiks

Figur A0.1: Results Monitoring Framework Template

RESULTS FRAMEWORK*					
RESULTS	Planned implementation period	Indicators	Baseline	Targets	Means of Verification
<b>IMPACT (and target group):</b>		<i>Indicator:</i>			
<b>Outcome 1:</b>		<i>Indicator:</i>			
Output 1.1		<i>Indicator/Delivered?</i>			
Output 1.2		<i>Indicator/Delivered?</i>			
Output 1.3		<i>Indicator/Delivered?</i>			
<b>Outcome 2:</b>		<i>Indicator:</i>			
Output 2.1		<i>Indicator/Delivered?</i>			
Output 2.2		<i>Indicator/Delivered?</i>			
Output 2.3		<i>Indicator/Delivered?</i>			
<b>Outcome 3:</b>		<i>Indicator:</i>			
Output 3.1		<i>Indicator/Delivered?</i>			
Output 3.2		<i>Indicator/Delivered?</i>			
Output 3.3		<i>Indicator/Delivered?</i>			
<b>Outcome 4:</b>		<i>Indicator:</i>			
Output 4.1		<i>Indicator/Delivered?</i>			
Output 4.2		<i>Indicator/Delivered?</i>			
Output 4.3		<i>Indicator/Delivered?</i>			

\*This is an example only. Please customize the number of outcomes and outputs to your project. For more information about results framework, see "Guide to Completing the Application for Project and Program Funding" and "Results Management in Norwegian Development Cooperation - A practical guide" on Norad.no.

Figur A0.2: SOS Barnebyer - utvidet resultatrammeverk

RESULTS	Indicators	2015 Baseline	2019 Achieved	2019 Targets
<b>Outcome 3:</b> Increased access to and benefit from quality education or vocational training for target children and youth in programme areas	# local schools providing quality education to target children	20 schools	54 schools	55 schools
<b>Output 3.1.1:</b> Improved school statistics for target children in the programme	% enrolled, attending school	82% boys 83% girls	94% boys 94% girls	95% boys 95% girls
	% Dropout	3% boys 3% girls	5% boys 5% girls	2% boys 2% girls
	% of educational performance	55%	85% boys 86% girls	80% boys 80% girls
	% Transition to secondary school	56% boys 47% girls	74% boys 72% girls	75% boys 75% girls
<b>Output 3.1.2:</b> Children with disabilities (CwD) from target families access quality education	# of school age CwD enrolled into school	83 boys 62 girls	802 boys 744 girls	115 boys 190 girls
	# of CwD supported with special aids	75 boys 56 girls	154 boys 146 girls	110 boys 85 girls
<b>Output 3.1.3:</b> Youth complete vocational skills development	# youth completing vocational training	99 boys 81 girls	351 boys 167 girls	275 boys 275 girls
<b>Output 3.1.4:</b> Schools supported and facilities rehabilitated	% rehabilitated (boreholes, sanitary facilities)	27%	67%	83%
	# School Management Committees & Parent Teacher Associations trained	21	61	66
	Teachers trained in quality education	4% M 53% F	84% M 87% F	90% M 90% F

## Do file for STATA

```
*Do file for analyse
*Datasett kan lastes ned fra følgende link:
clear
capture log close
set more off

*endre working directory til posisjon paa egen pc
cd "/Users/skagelem/Downloads/ELA master/"

use "/Users/skagelem/Downloads/ELA master/Uganda ELA Panel wide.dta".dta

*endre navn paa variabler saa datasettet kan bli "reshapet" fra wide til long
rename empowerment empowerment1

rename Rempowerment empowerment2

rename always_condom cond1

rename Ralways_condom cond2

*reshape datasettet saa endline ligger under baseline for empowerment og kondom
*Reshaper med id og lager ny variabel year med verdiene 1 and 2
reshape long empowerment cond, i(id treatment) j(year)

*lager dummy for tid slik at 2=1 og 1=0
gen time=(year==2)

*installerer package for aa inkludere 95% konfidensintervall til barplot
ssc install cibar

mean empowerment if treatment==0, over(time)
mean empowerment if treatment==1, over(time)

mean cond if treatment==0, over(time)
mean cond if treatment==1, over(time)

cibar empowerment if treatment==0, over(time)
cibar empowerment if treatment==1, over(time)
```



```
cibar cond if treatment==0, over(time)
cibar cond if treatment==1, over(time)

*bruker graph editor for aa manuelt endre navn og legge til verdier

*graf for empowerment i behandlings- og kontrollgruppe fra baseline til endline
graph twoway (lfit empowerment time if treatment==0)/*
*/(lfit empowerment time if treatment==1)
*lagre som png

*graf for kondom i behandlings- og kontrollgruppe fra baseline til endline
graph twoway (lfit cond time if treatment==0) (lfit cond time if treatment==1)
*lagre som png

*lager kryssvariabel for DiD
gen treattime=treatment*time

*DiD empowerment med kun treatment, time og kryssvariabel
reg empowerment treatment time treattime
est sto regemp1

*DiD med kontrollvariabler
reg empowerment treatment time treattime any_iga age M_children
est sto regemp2

*eksporterer regresjon til fil som kan settes inn i latex
#delimit;
esttab regemp1 regemp2 using regemp.rtf,
    b(%4.2f) se(%4.2f) r2(%4.2f) replace label star(* 0.10 ** 0.05 *** 0.01)
    unstack mtitle("Empowerment" "Empowerment m. kontrollvariabler")
    title("Regresjon av empowerment index");
#delimit cr

*DiD kondom med kun treatment, time og kryssvariabel
reg cond treatment time treattime
est sto regcond1

reg cond treatment time treattime partner M_children/*
```

```
*/ other_contraceptive E_Denrolled
est sto regcond2

*eksporterer regresjon til fil som kan settes inn i latex
#delimit;
esttab regcond1 regcond2 using regcond.rtf,
    b(%4.2f) se(%4.2f) r2(%4.2f) replace label star(* 0.10 ** 0.05 *** 0.01)
    unstack mtitle("Kondom" "Kondom m. kontrollvariabler")
    title("Regresjon av prosent som alltid bruker kondom");
#delimit cr
```