

Oddgeir Osland
Merethe Dotterud Leiren
Annelene Holden Hoff
Aud Tennøy
Sverre Strand
TØI rapport 929/2007

tøi : Transportøkonomisk institutt
Stiftelsen Norsk senter for samferdselsforskning



Evalueringsstradisjoner og kvalitetskriterier i samferdselssektoren



Evalueringstradisjoner og kvalitetskriterier i samferdselssektoren

Oddgeir Osland
Merethe Dotterud Leiren
Annelene Holden Hoff
Aud Tennøy
Sverre Strand

ISSN 0808-1190

ISBN 978-82-480-0829-3 Papirversjon

ISBN 978-82-480-0830-9 Elektronisk versjon

Oslo, desember 2007

Tittel: Evalueringstradisjoner og kvalitetskriterier i samferdselssektoren

Forfatter(e): Oddgeir Osland, Merethe Dotterud Leiren, Annelene Holden Hoff, Aud Tennøy, Sverre Strand,

TØI rapport 929/2007

Oslo, 2007-12

87 sider

ISBN 978-82-480-0829-3 Papirversjon

ISBN 978-82-480-0830-9 Elektronisk versjon

ISSN 0808-1190

Finansieringskilde:

Norges Forskningsråd

Prosjekt: 3092 Kvalitetskriterier for evaluering av samferdselstiltak

Prosjektleder: Oddgeir Osland

Kvalitetsansvarlig: Arvid Strand

Emneord:

Evaluering; kvalitetskriterier; samferdselstiltak

Sammendrag:

Rapporten gir en beskrivelse av og sammenfatter eksempler på ulike tradisjoner i evaluering av samferdselstiltak, fra de mer formaliserte som nytte-kostnadsanalyse til mindre formaliserte som kvalitative organisasjonsanalyser. Rapporten analyserer disse med bakgrunn i etablerte kriterier for kvalitet ved evaluering, og søker å lokalisere kritiske punkter ved ulike tradisjoner.

Title: Evaluation traditions and criteria in the transport sector

Author(s): Oddgeir Osland, Merethe Dotterud Leiren, Annelene Holden Hoff, Aud Tennøy, Sverre Strand,

TØI report 929/2007

Oslo: 2007-12

87 pages

ISBN 978-82-480-0829-3 Paper version

ISBN 978-82-480-0830-9 Electronic version

ISSN 0808-1190

Financed by:

the Research Council of Norway

Project: 3092 Criteria for evaluation of transport measures

Project manager: Oddgeir Osland

Quality manager: Arvid Strand

Key words:

Summary:

The report gives an overview and summary of different traditions for evaluation in transport research. The different approaches are analyzed based on established criteria for evaluation. This analysis leads to a localization of distinct and general critical points that have to be addressed in the different traditions in order to fulfill the

Language of report: Norwegian

*Rapporten kan bestilles fra:
Transportøkonomisk institutt, Biblioteket
Gaustadalleen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no*

*The report can be ordered from:
Institute of Transport Economics, The library
Gaustadalleen 21, NO 0349 Oslo, Norway
Telephone +47 22 57 38 00 - www.toi.no*

Forord

Denne rapporten er finansiert av Norges forskningsråd, gjennom det strategiske instituttprogrammet ”Kvalitetskriterier for evaluering av samferdselstiltak”. Oddgeir Osland har hatt hovedansvaret og skrevet alle kapitlene. Aud Tennøy har skrevet et førsteutkast til kapittel 2, Annelene Holden Hoff har bidratt til første del av kapittel 3, mens Sverre Strand har skrevet flere arbeidsnotater som har ligget til grunn for viktige deler av kapittel 3. På noen spørsmål har det vært ulike oppfatninger om hvordan tradisjoner skulle framstilles og drøftes. Det er derfor grunn til å understreke at det endelige resultatet er hovedforfatters ansvar. Merethe Dotterud Leiren har medvirket i omskriving av alle kapitler og vært en sentral samtalepartner i arbeidet. Takk også til Knut Sandberg Eriksen og Petter Næss for kommentarer på de deler av rapporten som omhandler deres arbeider. Arvid Strand har stått for kvalitetssikring av rapporten, mens den endelige tekstbehandlingen er utført av Camilla Olsson.

Oslo, desember 2007
Transportøkonomisk institutt

Lasse Fridstrøm
instituttssjef

Arvid Strand
avdelingsleder

Innhold

Sammendrag

Summary

Kapittel 1	1
1.1 Tematisk og foreløpig definisjonsmessig avgrensing.....	2
1.2 Fire perspektiver på evaluering	3
1.2.1 Evaluering som evalueringsmetodikk.....	3
1.2.2 Det informerte målrettede oppdragsgiverperspektivet	5
1.2.3 Det statsvitenskapelige informerte meta-perspektivet.....	9
1.2.4 Det profesjons- og kunnskaps sosiologiske perspektivet	12
1.3 Oppsummering av analytisk rammeverk og rapportens videre oppbygning	14
2 Evaluering av tiltakspakker for kollektivtransport	18
2.1 Bakgrunn.....	18
2.2 Evalueringens tema, teori og metodikk	19
2.2.1 Problemformulering/forskningsspørsmål	20
2.2.2 Teoretisk utgangspunkt.....	20
2.2.3 Metoder, datagrunnlag og devalueringer	22
2.4 Aggregert analyse av isolerte tiltakseffekter	23
2.4.1 Noen substansielle hovedfunn	24
2.4.2 Samfunnsøkonomiske analyser.....	25
2.5 Imøtekommer evalueringen validitetskriteriene?	26
2.6 Oppsummering.....	28
3. Store samferdselsinvesteringer. To evalueringstradisjoner – to hovedutfordringer.....	30
3.1. Kostnadsoverskridelser og feilbudsjettering og trafikkprognoser – noen funn og observasjoner	30
3.2 Ex post nyttekostnadsanalyse av store samferdselsinvesteringer.....	33
3.2.1 Kort om nyttekostnadsanalyse	33
3.2.2 Nærmere om ex post nyttekostnadsanalyser av samferdselsprosjekter	34
3.2.3 Nærmere om to konkrete jernbaneprosjekter – Sandeparsellene og Nationalteatret stasjon	36
3.2.4. Oppsummering av nyttekostnadsetteranalyse som metodikk for kontinuerlig læring.....	38
3.2.5 Interdisiplinær kritikk mot nyttekostnadsanalyser	39
3.3. Organisatoriske forklaringer på feilbudsjettering	43
3.3.1. Organisasjonsvariabler og strategisk budsjettering som tema.....	43

3.3.2 Aalborg-miljøtmiljøet og Flyvbjergs undersøkelse av budsjettoverskridelser	45
3.3.3 Kritikk av Aalborg prosjektet og Flyvbjergs argumentasjon	46
3.4.1 Oppsummering av organisatoriske og institusjonelle forklaringer på feilslåtte analyseforutsetninger	47
3.4 Kort om felles implikasjon av to ulike tradisjoner	49
4. Evalueringer som organisasjonsanalyse	51
4.1. Evaluering av arbeidsmiljøtiltak i bussbransjen	51
4.1.1. Teoretisk modell og avgrensning.....	52
4.1.2. Metodisk opplegg og avgrensning.....	53
4.1.3. Trekk ved undersøkelsen	56
4.2. Evaluering av TP 10 som prosess	56
4.2.1. Bakgrunn.....	56
4.2.2 Teoretisk og metodisk opplegg.....	57
4.2.4 Design, datainnsamling og særlige utfordringer.....	58
4.2.5 Hovedkonklusjoner	59
4.2.6 Oppsummering med bakgrunn i analytisk skjema.....	60
4.3. Manglende effektvariabler og ekstern validitet som metodisk utfordring..	61
4.4. Oppsummering.....	64
5. Oppsummering og avsluttende diskusjon	66
5.1. Oppsummering av kjennetegn ved ulike evalueringer	68
5.1.1 Tiltakspakkene – evaluering som multidimensjonal variabelanalyse ..	68
5.1.2 Ex post nyttekostnadsanalyse: Evaluering som systematisk læring – med ett kritisk unntak	70
5.1.3 Evaluering av mega-prosjekter som organisasjonsanalyse.....	72
5.2. Avsluttende drøfting – implikasjoner for videre arbeid	76
Referanser.....	79

Sammendrag:

Evalueringstradisjoner og kvalitetskriterier i samferdselssektoren

Formålet med denne rapporten har vært å kartlegge kvalitetskriterier som tas i bruk ved ulike typer evalueringer, hvilke likheter og forskjeller vi finner på tvers av evalueringstradisjoner og -tema, og hvilke særegne utfordringer og dilemmaer en står overfor når en skal realisere disse kriteriene.

Vi definerer evaluering som en *systematisk etterundersøkelse av om effektene av offentlig virksomhet er i samsvar med intensjonene, med det formål at slike undersøkelser skal ha relevans for pågående eller fremtidige beslutningsprosesser.*

Når en setter effekter i sentrum for analysen, må en forholde seg til kontrafaktisitetetsproblemet som det avgjørende spørsmålet ved evaluering, dvs. spørsmålet om effektene ville vært annerledes dersom bestemte tiltak ikke hadde vært introdusert.

Fire perspektiver

Vi analyserer evaluering med bakgrunn i fire perspektiver. Det første perspektivet som forstår evaluering som evalueringsmetodikk, har sin bakgrunn i Elviks forståelse av at det er metoden som skiller en god og en dårlig evaluering – ikke resultatet. Avgjørende er det om evalueringen tilfredsstillende kravene til statistisk, teoretisk, intern og ekstern validitet.

Det andre perspektivet, det informerte oppdragsgiverperspektivet, forstår evaluering som en målrasjonell prosess. Evaluering tas i bruk når saksforholdet er vesentlig, dvs. har stor økonomisk eller prinsipiell betydning. Evalueringsdesignet bør fortrinnsvis være formalisert for å kunne sammenligne ulike evalueringer, og det bør være basert på indikatorer som er gode styringsparametre.

Det tredje perspektivet, det statsvitenskapelig informerte meta-perspektivet, slik vi finner det hos Vedung, er pluralistisk. Det bygger på en antagelse om at en kan benytte ulike perspektiver til ulike evalueringsformål, samtidig som han avgrenser evaluering til etteranalyser.

Det fjerde perspektivet, det profesjons- og kunnskapssosiologiske perspektivet, innebærer den forståelse at teorier "begriper" likeartede fenomener på ulike måter. Forskere "finner ikke" empirien, de skaper den også gjennom de måtene teoriene begriper fenomenene. Dette gjør at validitetsspørsmål i noen grad er teori-interne, ulike funn og vurderinger kan være resultater av ulike teorier og modeller, noe som gjør sammenlignbarhet vanskelig og interdisiplinære kontroverser utbredt.

Videre er ulike begreper ”verdilatet”, slik at bestemte teorier gir støtte til noen hensyn og verdier og svekker andre.

To av perspektivene - det informerte oppdragsgiverperspektivet og Vedungs metaperspektiv – er her benyttet avgrensende: De definerer hvilke typer evalueringer vi inkluderer i denne rapporten, og hvilke vi ekskluderer. Denne funksjonen har, i all hovedsak, også Elviks perspektiv om evaluering som evalueringsmetodikk. Vi tar utgangspunkt i at metoden, ikke funnene, bestemmer om evalueringen er god eller dårlig. Det profesjons- og kunnskapssosiologiske perspektivet introduseres for å få fram at et meta-perspektiv må åpne for spørsmålet om sammenlignbarhet mellom perspektiver på ulike fenomener.

Klassifikasjonsskjema for sammenligning av evalueringer

Disse perspektivene gir grunnlag for følgende klassifikasjonsskjema for sammenligning og vurdering av evalueringer:

- Hva er *temaet*, *problemstillingen* eller *formålet* med evalueringen?
- Hva kjennetegner *metode* og *datagrunnlag* (kvalitative eller kvantitative data, eventuelt type underkategori)? Hvor omfattende er undersøkelsen, hvor mange variabler og typer variabler har den?
- Hva kjennetegner *teorien* som blir benyttet, hvilke *forutsetninger* bygger den på, hvilke *antagelser* søker den å undersøke, hva er de sentrale *begrepene*, har den en klar *disiplinforankring*, hva er dens *gyldighetsområde*?
- Er mulige/sannsynlige utilsiktede effekter godt dekket?
- Kumulativitet: Hvordan antas evalueringen å kunne føre til ny kunnskap (substansielt/metodisk/teoretisk)?

I klassifikasjonsskjemaet inkluderes også *kvalitetskriterier*: Hvilke kvalitetskriterier tar de ulike evalueringene i bruk? Har de et design som i utgangspunktet kan tilfredsstillende krav til:

- *Statistisk validitet* (dvs. reliabilitet, representativitet, analyseteknikker og statistiske sammenhenger),
- *Teoretisk validitet* (eksplisitt teorigrunnlag, operasjonalisering, testbare hypoteser, gir funnene støtte til teorien),
- *Intern validitet* (om det i tillegg til den statistiske sammenhengen også kan fastslå årsaksretning, og om denne holder for kontroll med andre forklaringer)
- *Ekstern validitet* (om en kan generalisere resultatene av undersøkelsen utover den spesifikke konteksten, i tid, rom og eventuelt til andre saksområder)

Anvendelse av skjemaet

I rapporten anvender og prøver vi ut dette klassifikasjonsskjemaet på noen utvalgte evalueringer innen samferdselssektoren, primært med sikte på å få fram hypoteser om det spesifikke ved bestemte typer evalueringer eller det vi mer løselig kan omtale som evalueringstradisjoner. Vi skiller her mellom tre typer: Multidimensjonal variabelanalyse, nyttekostnadsanalyse og organisasjonsanalyse. De fleste av de evalueringene som blir gjennomgått har til felles at de har pågått over lengre tid med relativt omfattende ressurser. Det er derfor ikke grunn til å anta at de på det området skiller seg negativt ut fra arbeidssituasjonen ved evaluering generelt, slik at kvaliteten ikke burde være lavere enn i andre tilfeller.

Multidimensjonal variabelanalyse

Den første evalueringen vi gjennomgår og vurderer, er evalueringen av tiltakspakkene for kollektivtransport (kapittel 2). Formålet med denne typen evaluering antar vi at er representativ for det vi har gitt benevnelsen multidimensjonal variabelanalyse: Den omhandler og analyserer en rekke ulike variabler på ulike analysenivå som både omhandler tiltak, rammebetingelser, holdninger og handlinger blant målgruppen (i dette tilfellet passasjerer). Gitt at en, som i dette tilfellet, har diakrone data, er dette et design som i utgangspunktet kan tilfredsstillende kravene til statistisk validitet. I denne konkrete evalueringen finner vi klare svakheter ved selve analysen av data. Dels gjelder dette operasjonell validitet, problemer med å benytte opplevd endret tilfredshet med tilbudet og opplevde endrede reisevaner som indikatorer for tilbudsending og endrede reisevaner. Svakheter ved operasjonelle variabler, kombinert med manglende data for passasjerstatistikk, kan også antas å svekke de aggregerte analysene. Det er et problem med presentasjonen av data-analysen at slike forhold i liten grad drøftes og presenteres på en måte som gjør grunnlaget for regresjonsanalysen transparent. Dermed er det også et paradoks at resultater presenteres med høy grad av presisjon.

De teoretiske begrepene, som ”holdninger”, ”preferanser”, ”faktisk atferd”, er operasjonaliserbare, og i liten grad forankret i en bestemt teoretisk disiplin. Den teoretiske modellen er relativt omfattende, med lite klart formulerte hypoteser. Funn gir derfor ikke grunnlag for å diskutere selve modellens relevans, og det blir heller ikke gjort. Manglende disiplinforankring gjør at ekstern validitet ikke kan drøftes i forhold til bestemte, mer overgripende teorier. I prinsippet kunne en ha diskutert ekstern validitet i betydningen om funnene kan reproduseres på samme type fenomen (tiltak for økt kollektivandel i byområder). Heller ikke dette blir gjort her.

Designet innebærer at evalueringen potensielt kan være kumulativ på to måter - metodisk og substansielt. Metodisk ved oppbygging av databaser som gir grunnlag for komparative statistiske analyser over tid, men dette forutsetter at databasen forbedres og utvikles. Substansielt ved økende grad av kunnskap om faktorer som påvirker kollektivandeler i by. Et hovedfunn er at tiltakene ikke har vesentlig effekt på transportmiddelfordelingen, men i noen tilfeller gir passasjervekst for kollektivtransporten. Også på dette området drøfter forfatterne i liten grad den eksterne validiteten til de konklusjonene som trekkes, som at bedret

frekvens er den viktigste forklaringen til endret reisehyppighet, og at dårligere tilbud gir større passasjerbortfall enn forbedringer gir passasjervekst.

Ex post-nyttekostnadsanalyser

I kapittel tre drøftes to hovedtilnærminger til analyser av feilbudsjettering og kalkyle i større samferdselsinvesteringer. De to tilnærmingene gir grunnlag for en diskusjon om tverrfaglige kontroverser. Den ene typen evaluering som gjennomgås, nyttekostnads-etteranalyser, innebærer en antagelse om at en viktig årsak til feilkalkyler er metodiske svakheter. Etteranalyser ble forsøkt gjennomført for flere jernbaneprosjekter, parallelt med vurdering av tidligere og nåværende metodehåndbøker. Dokumentasjonen for tidligere prosjekter var imidlertid så dårlig at selve etteranalysen bare hadde illustrativ verdi. Derimot ga gjennomgangen grunnlag for konkrete forslag til forbedringer i metodehåndboka.

Formålet med nytte-kostnadsanalyse som etteranalyse er nettopp å forbedre ex ante-metodikken gjennom å: 1) Kartlegge avvik mellom de beregnede samfunnsøkonomiske virkningene i forundersøkelsen og etterundersøkelsen og 2) å lokalisere hvilke av disse som skyldes ulik metode, ytre forutsetninger, prosjektavhengige avvik.

Denne typen analyse er disiplinforankret i velferdsøkonomisk teori, der formålet er å undersøke om et tiltak gir et positivt eller negativt samfunnsøkonomisk resultat. Dette skjer bla ved at en sammenstiller virkninger som så langt mulig tilstrebes gitt en felles pengeverdi. Utgangspunktet for slike analyser er begrepene om konsument- og produsentoverskudd. Et prosjekt hvor den samlede neddiskonterte nytten (målt i kroneverdi) er større enn de samlede neddiskonterte kostnader (også i kroneverdi) kan sies å være samfunnsøkonomisk lønnsomt og representerer følgelig akseptabel ressursbruk for samfunnet.

En nytte-/kostnadsanalyse skal vise tiltakets effekt i forhold til ulike aktører i samfunnet. Her skilles det gjerne mellom noen hovedkategorier av interessenter: Det offentlige, brukerne og samfunnet for øvrig. I samferdselssektoren framkommer ofte effektene for det offentlige i form av investeringskostnader, endringer i drifts- og vedlikeholdskostnader og offentlig kjøp av transporttjenester. Brukerne er både operatører og trafikanter, som kan oppleve endret tidsbruk (reisetid, ventetid, gangtid) og endring i direkte økonomiske størrelser som driftskostnader og utlegg (som for eksempel billettkostnader). Samfunnet for øvrig uttrykker det som også kalles eksterne effekter, som for eksempel endringer i lokale, regionale og globale utslipp, støy og naturinngrep.

Det teoretiske rammeverket er preget av at antagelser og forutsetninger er eksplisitt og klart formulert. Undersøkelsen tilfredsstillende kravene til et klart teorigrunnlag ved at begreper er operasjonaliserbare. Det mest kritiske punktet i operasjonaliseringsprosessen er usikkerhet ved operasjonalisering av komponenter som i utgangspunktet ikke er beskrevet i økonomiske termer: Tidskomponenter og eksterne kostnader som miljø- og ulykkeskostnader. Teorien gjør det også mulig å teste antatte sammenhenger mellom sentrale variabler, men (selvsagt) ikke den teoretiske forutsetningen at dette er de relevante variablene. I sum tilfredsstillende denne typen evaluering godt kravene til teoretisk validitet.

Når det gjelder statistisk validitet, opererer denne type analyse med to typer statistiske sammenhenger, for det første de sammenhenger som uttrykkes gjennom selve nytte-kostnadsanalysen, som det krever spesialisert kunnskap for å vurdere og er lite transparent, og for det andre framstillingen av avvikstabeller som gir en oversiktlig avviksanalyse.

Selve det metodiske opplegget er designet for å ivareta de fleste sidene ved intern validitet: Systematisk gjennomtenkning av årsaks-virkningsforhold. Ekstern validitet og generaliserbarhet, er primært knyttet til utvikling av metodikken, dels som gjennomgang og forbedring av selve metodehåndbøkene og forbedret og mer ensartet metodikk for veg og jernbane. Det er også i form av utvikling av metodeverktøy at denne typen undersøkelser er kumulativ. Oppsummert er dette en evalueringstradisjon preget av høy grad av formalisering, som imøtekommer de mest grunnleggende krav Elvik har formulert til evalueringsmetodikk. Formålet er å utvikle og forbedre metodeverktøyet slik at en gjør færre feil i framtidige nytte-kostnadsanalyser.

Men det er særlig én side ved prosjektgevinster og -kostnader som ikke blir empirisk etterprøvd gjennom et slikt opplegg, og det er usikkerheten knyttet til konvertering av tids-, miljø- og ulykkeskostnader til økonomiske kategorier. For andre typer variabler kan ex post nyttekostnadsanalyse sees på som et forsøk på å lokalisere metodiske svakheter som kan medføre feilbudsjettering og -prioritering innen samferdselssektoren.

Kritikken som er reist mot denne metoden (dvs. mot nyttekostnadsanalyse per se og ikke primært mot etteranalyser) er knyttet til forutsetninger som gjøres og som ikke testes. Dette gjelder særlig spørsmålet om tidsverdsetting og verdsetting av kostnader knyttet til ulykker, støy osv. Slike problemstillinger er kontinuerlig diskutert og analysert innenfor den typen fagmiljø som jobber med nyttekostnadsanalyse. Typisk i denne sammenheng vil være å basere verdsettingen på individenes egne verdsettinger, eventuelt utvidet til kartlegging av ulike trafikantsegmenters verdsettinger. Kritikken som reises utenfra, fra forskere i andre fagtradisjoner, er at selve de forutsetningene som gjøres for slike analyser er problematiske. I rapporten gjengis Sverre Strands kritikk, der han peker på at tidsverdien gjerne vil være den største verdien i mange nyttekostnadsanalyser av samferdselsprosjekter, og at det er visse grunnlagsproblemstillinger knyttet til verdsetting av tid. Han skiller mellom tre trinn i tidsanalyser: 1) Måling av tid og tidsbesparelser, 2) aggregering av tidsbesparelser og 3) pengeverdering av tidsbesparelser. Spørsmål 1 omhandler tema som hvor mye større enn null en enkeltbesparelse må være for at den skal kunne omtales som signifikant nok til at tiden kan anvendes alternativt. Problem 2 omhandler under hvilke forutsetninger tidsbesparelser kan aggregeres, mens det tredje punktet handler om konvertering av tidsverdier til pekuniære verdier. Strand hevder at alle tre er problematiske, men at standardtilnærmingen til problemet vil være at en gjennom tidsverdsetting løser 1 og 2.

Tilsvarende rettes det systematisk kritikk mot nytte-kostnadsanalyse for en rekke andre antagelser som den bygger på, som aktørforutsetninger innen konsumentteori som konverterer det kritikere oppfatter som moralske, kollektive problemstillinger til spørsmålet om individuelle preferanser, og problemer knyttet til bruk av neddiskonteringsraten på miljøspørsmål. Denne fageksterne kritikken

kan sees på som ledd i en interdisiplinær kritikk, som hevder at det å benytte denne type analyse medfører å fremme bestemte verdier. Kritikerne peker også på problemet med manglende transparens ved slike modeller. Rapporten drøfter i noen grad kritikken av kritikken, men understreker at dette mest av alt viser behovet for offentlig, saklig tverrfaglig debatt.

Organisasjonsevalueringer (I)

Den andre tradisjonen for evaluering av overskridelser og feilbudsjettering i samferdselsinvesteringer som drøftes i denne rapporten, er Aalborg-miljøets forskning på feltet. I tillegg til å peke på metodiske forklaringer som årsak til slik feilbudsjettering, analyserer en i denne tradisjonen også organisatoriske og institusjonelle variabler. Aalborg-miljøets forskning blir i første rekke presentert gjennom Bengt Flyvbjergs arbeider. Flyvbjerg forstår overskridelser som resultat av desinformasjon og lokaliserer tre hovedmodeller for å forklare slik desinformasjon: Tekniske modeller som refererer til feil i data og metoder, psykologiske modeller som refererer til kognitive feil og politisk-økonomiske modeller i form av strategisk misrepresentasjon og løgn. Han argumenterer for at de to første modellene ikke kan forklare desinformasjon og feilbudsjettering, blant annet fordi en da skulle forvente læring og redusert feilbudsjettering over tid. Fordi de andre forklaringsmodellene ikke finner støtte gjenstår den siste. Den beste forklaringen blir da at insitamentstrukturen er slik at fordelene med å desinformere er større enn ulempene og resultatet blir ”survival of the unfittest” – det er ikke det beste prosjektet som vinner, men det som har best desinformasjon. Botemidlene for dette problemet er dels utvikling av bedre metoder for å estimere kostnader, nytte og risiko gjennom ”reference class forecasting”, dels endrede insitamentstrukturer som vil begrense desinformasjon og fremme riktig informasjon.

Teoretisk er tilnærmingen til Flyvbjerg eklektisk uten en klar disiplinforankring. Fordelen med en slik tilnærming er at den spenner opp et relativt bredt analytisk rammeverk som inkluderer mange relevante variabler - både metodiske, psykologiske og organisatoriske. På den andre siden er aktørforutsetningene og antagelsene om institusjoners virkemåte noe underspesifiserte, men drøftingen indikerer at disse er forankret i institusjonell økonomi. Det kritiske punktet i analysen er den teoretiske og interne validiteten. Det er for eksempel få klare empiriske indikatorer på løgn, og denne forklaringsvariabelens kraft fastlegges gjennom et kontrafaktisk resonnement. Det metodiske problemet Flyvbjerg står overfor er det klassiske for de som forsker innen tradisjoner som omfattes av mistankens hermeneutikk, det er vanskelig å finne empirisk støtte for at løgnere lyver, enten de vet det selv eller ikke.

I rapporten argumenterer vi for at det analytiske inntaket til Flyvbjerg – ”organisation matters” – representerer en fruktbar dreining og et viktig supplement til utvikling av metodeverktøyet. Vi argumenterer imidlertid også for at viktige elementer i selve tolkingen av disse organisatoriske trekkene og aktørforutsetningene, er problematiske. En mulig retning for forbedring av denne forskningen vil være å introduserer et mer finmasket variabelsett både når det gjelder organisasjonsvariabler og trekk ved aktørene. Forskningsprosjekter med dette temaet kunne gi mer kunnskap om bakgrunnen for overskridelser og slik ha

praktisk-politisk betydning for hvordan vi organiserer beslutningsprosessen, samt være et bidrag i den internasjonale faglige debatten.

Organisasjonsevalueringer (II)

I det siste kapitlet drøftes organisasjonsevalueringer mer systematisk. Kapitlet innledes med en evaluering der betingelsene for en systematisk undersøkelse av effekter av organisasjonsendring lå godt til rette, fordi en både hadde den tid, de ressurser og det datamateriale som tillot kartlegging av endring både på effektvariablene (trivsel og sykefravær) og på tiltaksvariabel (introduisering av team). Kapitlet fortsetter med en gjennomgang av en organisasjonsanalyse der en ikke hadde effektvariabler: Evalueringen av TP 10 som prosess. Denne tar for seg forsøk på samordnet areal- og transportpolitikk i ti større norske byområder og har status som en prosessevaluering. Den teoretiske modellen for evalueringen er lite formalisert, men avgrenser seg til å presentere en analytisk modell som beskriver sentrale aktører, prosjektdesign, mål for og organisering av lokale prosesser, og endelige produkter i form av planer og vedtak. Effektene av slike vedtak måles ikke, blant annet fordi de antas å være av så vidt langsiktig karakter.

Det analytiske skjemaet impliserer ikke spesifisering av hvilke variabler som kan antas å påvirke utfallet. Dataene baserer seg også her på flere typer metoder; intervjuer, seminarer, spørreskjema. Basisen for tolkningen er mer omfattende beskrivelser der noen sentrale begreper (som sentrale aktører, føringer) har heuristisk karakter og strukturerer drøftingen. Analysen presenterer ikke et eksplisitt teoretisk grunnlag, men presenterer noen begreper og antatte sammenhenger forankret i statsvitenskap og organisasjonsteori. Designet bygger ikke på statistisk analyse og kan dermed heller ikke sies å tilfredsstille kravene til statistisk validitet.

Manglende eksplisitt teoretisk grunnlag gjør det vanskelig både å vurdere den interne og den eksterne validiteten. Designet er slik lite formalisert og vurdering av evalueringen må basere seg på vurdering av de substansielle tolkingene til forskere og andre. Gyldighetsområdet og den eksterne validiteten avgrenser seg dermed til samordnet areal og transportpolitikk i norske byområder. På dette feltet har den imidlertid også hatt effekter. En strøm av utredninger om svakheter ved organiseringen av den lokale areal- og transportpolitikken fulgte i kjølvannet av denne evalueringen. Den har altså vært kumulativ i den forstand at den har vært oppfattet å representere sentrale kjennetegn ved sektoren.

Generelle utfordringer

For det første er det et fellestrekk med de evalueringene som har vært gjennomgått i denne rapporten, at de i liten grad drøfter ekstern validitet, og dermed i liten grad andre undersøkelser om samme emne. Det mest klare unntaket er nytte-kostnadsanalysen, der utvikling av modellverktøyet er sentralt. For det andre er det også, med samme unntak, et fellestrekk at undersøkelsene i liten grad eksplisitt knytter seg til overordnede teoritradisjoner. Begge deler innebærer at det er liten grunn til å anta at slike evalueringer vil bli underlagt den kvalitetskontrollen og muligheten for nyansering og kunnskapsutvikling som følger av systematisk komparasjon og teoretisk diskusjon. Det er dermed

avgjørende at en enten legger forholdene til rette for slik videreutvikling, eller at en etablerer funksjonelle ekvivalenter til den type kritikk som ligger i den offentlige debatt.

Videre avdekker rapporten to typer transparensproblemer. Det transparensproblemet som følger med komplekse statistiske analyser, som øker når datagrunnlaget ikke er offentlig, innebærer at utenforstående vanskelig kan få innsikt i fenomenet. Avgjørende er her den kontrollen som ligger i kollegial kritikk. Denne type transparensproblem kan assosieres til det Elster har kalt "hard obskurantisme" – selv om dette begrepet også inkluderer at kunnskapen i tillegg til å være lite transparent også er lite relevant. Det andre transparensproblemet finner vi i kvalitative evalueringer der modeller er lite formaliserte, slik at antagelser og resonnementer vil være vanskelig tilgjengelige. På dette området er den valideringen som ligger i møtet med informantene og tilbakemeldinger et sentralt kontrollelement. Disse to ytterpunktene representerer også ulike måter der evaluerer kan påvirke resultatet, bevisst eller ubevisst. I statistiske analyser kan dette skje gjennom rekoding- og modelleringsprosessen, i kvalitative analyser i sammenfatningen og tolkingen av data. Til forskjell fra kvantitative analyser, vil en organisasjonsanalyse uten effektvariable, ikke kunne håndtere kontrafaktisitetproblemet.

Vår rapport viser hvordan de hovedtypene av kvalitetskriterier som Elvik stiller gjennom sin analyse, i all hovedsak er sentrale og fruktbare redskaper for å vurdere evalueringer, og å få fram kritiske punkter ved ulike evalueringstradisjoner. Å stille krav om teoretisk, intern og ekstern validitet medvirker til å skjerpe blikket for styrker og svakheter ved evalueringer, og å etablere en felles målestokk. Derimot er det gode argumenter for å fravike kravene til statistisk validitet, fordi dette i noen tilfeller vil innebære å forkaste muligheten for å få fram kunnskap som er vesentlig, men ikke kan skaffes gjennom statistiske analyser.

Implikasjonene av det profesjons- og kunnskapsteoretiske perspektivet kommer fram i drøftingen av evalueringer av større samferdselsinvesteringer. For det første viser to ulike tilnærminger, nyttekostnadsetteranalyser og organisasjonsanalyser, hvordan ulike tilnærminger kan fremskaffe komplementær kunnskap. For det andre er kritikken og diskusjonen om nyttekostnadsanalyser et eksempel på interdisiplinære kontroverser som bla omhandler hvordan modellene kan fremme bestemte verdier.

Profesjonsperspektivet benyttes også til å sammenfatte trekk ved koblinger mellom evaluerere, oppdragsgivere og grad av transparens. Følgende hypotese formuleres: Epistemiske samfunn som er organiserte rundt spesialisert kunnskap som rommer lite transparente modeller, kan representere bestemte problem – for omgivelsene. Manglende metodiske overveielser og kritikk kan, i de tilfeller det epistemiske samfunn er organisert rundt komplekse statistiske analyser, medføre entydige og sikre vurderinger på spørsmål som er flertydige og sammensatte – som rangordning av prosjekter i nyttekostnadsanalyser. I andre tilfeller der slike epistemiske samfunn er organisert rundt bestemte politiske problemer og har liten affinitet til alternativanvendelse av ressurser, vil manglende metodiske overveielser gi støtte til en ekspanderende policy-sektor. Logikken i resonnementet vil være at når dette tiltaket ikke virket, så øker vi medisinen,

Summary:

Evaluation traditions and criteria in the transport sector

This report discusses different evaluation types within the transport sector, in order to track important challenges that evaluators face.

Initially, we distinguish between different approaches of evaluation. Elvik's methodological approach of meta-evaluations emphasises that evaluations have to fulfil key methodological criteria such as statistical, theoretical, internal and external validity. The employer/principal approach emphasises that evaluations should improve knowledge of to what degree public policies leads to goal attainment, whereas Vedung's approach emphasises the importance of pluralism – pluralism in the sense that different approaches are adequate for different purposes. However, all three perspectives share an emphasis on the counterfactual as a key challenge in evaluation. An evaluation should answer, or at least address, the question of what would have happened if the measure introduced and evaluated had not been introduced. In addition, we introduce an approach rooted in the sociology of knowledge, emphasising that different evaluations are rooted in different disciplines, and based on different assumption, concepts and models that ultimately leads to different “findings”. Hence, this approach accentuates the question of theoretical comparability and compatibility between disciplines.

Several different evaluations are described and discussed in the report. In the first chapter measures introduced in order to improve public transport are discussed. The project evaluates measures introduced by principals. The local authorities (the principals) provide the data to the evaluators. Also, the choice of cases follows funding. The design gathers data from before and after the introduction of measures, a design that in principle should be able to produce data that would meet the criteria of statistical validity. In this evaluation, however, this requirement is only partially met due to lack of adequate data, in particular on the dependent variable which is passenger-growth. The theoretical framework is based on “middle-range” concepts (such as preferences, attitudes, options, actions) that may be related to more general theories in different disciplines (sociology, psychology), but they are merely presented as correlations. There is also a lack of internal and theoretical validity due to missing statistical validity. Also the external validity is not discussed. This evaluation is viewed as typical for multidimensional variable analyses - it is an evaluation with high potential for analysing change based on correlations between variables. Its external validity, has to be addressed either through theoretical or substantial comparison.

The principal approach's imperative is that evaluations should be made in cases of high significance or risk. Following this, the third chapter addresses evaluations of large-scale infrastructure projects. The chapter refers to research that find

miscalculations and cost overruns as a major problem in transport policy, both nationally and internationally. Two approaches try to explain this problem.

First, ex post cost benefit analyses are based on the assumption that these miscalculations may be reduced by improving the methods. Such analyses represent an exemption in our report in the sense that it is an evaluation anchored in a theoretical discipline (welfare economics), has clear assumptions, and has a design that – given its validity area – has a high potential for meeting statistical, theoretical, internal and external validity.

However, in the cases of ex post cost benefit analyses described in our report, the data provided to the evaluators was scarce. Due to lack of information of the ante situation, the evaluators were not able to make proper ex post cost benefit analyses. For that reason they could not make conclusions of miscalculations, but they contributed with advices on how to improve documentation and methods.

Our report points out that, although some aspects of the valuation of time, environment and accidents are continuously debated, tested and sophisticated within the discipline of economics, some aspects are merely assumptions – and hence not tested. On one hand, ex post cost benefit analyses are paradigmatic as a type of evaluation that meets the criteria – or ideal – of meta-evaluation. On the other hand, such analyses are based on assumptions that may be questioned. The valuation of time illustrates our argument. It is an important variable in the determination of the cost benefit ratio, but the assumption of aggregation of time values is controversial. Our report gives an overview of criticism against cost benefit analyses in general. It concludes that although some criticism is inconsistent, such interdisciplinary controversies show that there is uncertainty connected to the results. Hence, discussions on key assumptions of cost benefit analyses in the interdisciplinary and public debate is requested, not the least because cost benefit analyses hold an important position in the administrative process - yet less legitimacy in the political process.

The second approach to evaluation of large-scale infrastructure projects is the organisational approach. This is an approach based on the assumption that there are not only methodological, but also organisational explanations of miscalculations and cost overruns. The registration of data, the calculations of costs and benefits, as well as the presentation of results are made in a certain organisational and institutional setting, e.g. calculations may be carried out by organisations that may have economic interests in the results, and in an institutional environment with scarce public resources, and in which the actors compete on projects.

Our report presents and discusses the Danish pioneer research by Flyvbjerg in particular. His work aims at explaining why transport projects tend to be more costly than budgeted. The design of his evaluations is characterised by middle-range theoretical concepts and is interdisciplinary anchored. The design seems to meet statistical and internal validity, however, there is a lack of such internal validity in Flyvbjerg's work due to his interpretation of the data. Moreover, the theoretical validity in his work is problematic. Our report argues that some of his conclusions are inchoated, in particular when it comes to the notion of a 'fooling game' which results in 'the survival of the unfittest.' Also the conclusion that

projects that are the misrepresented (fooling the principal) will win, rather than the best projects, is problematic.

Moreover, our report finds that the approach lacks variance on the actor variable as well as the institution variable, e.g. Flyvbjerg assumes that all actors are equal (all actors are opportunistic, rather than e.g. also acting in accordance with scientific norms), and all institutional incentive structures promote lies. However, the very improvements of incentive structures that the researchers propose, such as increased transparency, local cofinancing, and lump-sum rather than earmarked financing, are characteristics that vary across time and space. Hence, such variation should be included in the analyses. Moreover, such analyses would be of significance for further public and interdisciplinary debate on cost overruns and miscalculations of large infrastructure investment projects.

The third chapter thoroughly discusses organisational evaluations. The importance of such evaluations is acknowledged, firstly because organisational change often is introduced as a measure in itself, and secondly, because public measures such as new regulations, new economic incentives, and new procedures are always transposed through organisations. This implies that organisational characteristics affect the outcome of a change. However, as argued in our report, organisational evaluations are difficult in cases where an organisation is an independent or mediating variable, but there is a lack of information of the dependent variable. In such cases the question of the counterfactual remains unsolved.

This is a challenge of the organisational evaluation, the process evaluation of transport plans in ten Norwegian cities (TP 10), also addressed in the third chapter. Its evaluation design is based on an analytical framework based on concepts such as actors, processes, and local projects. The case studies in this evaluation are thoroughly described, but both the internal and theoretical validity is difficult to examine because it is not made explicit in the analyses. However, the results seem to have been validated through feed-back from interviewed actors.

We argue that the general conclusion is based on a premise that is not tested. The main argument is that TP 10 had limited effects because other important elements in the institutional environment remained unchanged. Despite the institutional changes which were introduced in order to reduce e.g. road building incentives, the incentive structure still made it rational for local authorities to prioritise road building as it maximized economic external resources (resources from the state and not local funding). The argument following from this diagnosis was that the organisational change was not strong enough.

However, counteracting hypotheses, i.e. additional changes in incentive structure would not alter priorities, were not addressed. In our report we argue that the evaluator of TP 10's argument may be right, but it is not empirically confirmed.

To sum up, it is a common characteristic of the evaluations discussed in our report, that they barely discuss the external validity of findings. If this is a general pattern, it implies that findings from evaluations are not given the critical scrutiny and control that follows from systematic comparison and critique.

Furthermore, our report enhances two types of transparency problems: the transparency problem related to complex statistical analyses, in which critique must be formulated by peers, and the problem related to qualitative analyses in which the validation sometimes is carried out by the informants. The potential bias of such transparency problems are likely to increase in strong epistemic communities of evaluators, administration and practitioners, sharing common beliefs of a given political problem, and how it is to be solved, and having monopoly of definitions of and discussions on these problems.

Kapittel 1

Temaet for denne rapporten er evalueringer av offentlig politikk. Formålet er å sammenfatte kriterier for å vurdere kvaliteten av evalueringer og evalueringsopplegg, og å anvende disse på ulike typer evalueringer for å lokalisere typiske utfordringer som disse står overfor.

Enkelt sagt bygger politiske myndigheters bruk av evaluering innen offentlig virksomhet på antagelser om at det er mulig:

- i) Å oppnå en mer effektiv bruk av de ressursene som er stilt til rådighet av og for fellesskapet for å realisere målsetningene som det samme fellesskapet har besluttet gjennom sine folkevalgte organer, og
- ii) Å vurdere og dermed diskutere denne ressursbruken på en fornuftig måte. Dette innebærer at vi kan utvikle og etablere felles kriterier for en slik vurdering og diskusjon.

Bakgrunnen for denne rapporten er en antagelse om at kjennetegn ved og omfanget av kriterier for å vurdere kvaliteten av evalueringer, slik de *de facto* er i bruk i dag, varierer. For det første finner vi varierende kriterier i bruk i evalueringer. Denne variasjonen har sitt opphav i forhold som ulike fagtradisjoner, problemstillinger og temaområder. Noe av denne variasjonen kan være velbegrunnet. Det kan for eksempel være hensiktsmessig å ha andre evalueringskriterier når en skal vurdere nytten av en bestemt veistrekning i Nasjonal Transportplan (NTP) enn når en skal vurdere selve NTP som beslutningsprosess. En slik variasjon i kriterier kan være sektorspesifikk og -forankret: Det kan finnes gode grunner til å ha ulike kriterier når en skal vurdere effekten av konkurranseutsetting av lokal busstransport og konkurranseutsetting av sykehjem.

Variasjon i kriteriebruk kan imidlertid også være ubegrunnet: Uenigheter som ytrer seg som kontroverser mellom fagtradisjoner eller kontroverser mellom oppdragsgiver og evalueringer, kan skyldes at den ene parten tar i bruk argumenter som ved nærmere ettersyn ikke bunner i kvalitetsvurderinger, men i helt andre forhold. Det at vi kan ha ulike, og mer eller mindre velbegrunnede oppfatninger om hvilke kriterier som skal benyttes i evalueringer, aksentuerer det som er tema her, hvilke kriterier en bør ta i bruk når en skal vurdere kvalitet *ved* evalueringer gjennom meta-evalueringer, dvs. evalueringer av evalueringer. Igjen kan denne variasjonen være velbegrunnet eller ubegrunnet, ureflektert og/eller feilaktig.

Spørsmålet om hva som skiller en god evaluering fra en dårlig evaluering er påtrengende både i forsknings-/utredningsmiljøet og hos oppdragsgivere. Og det er et vanskelig spørsmål. Det rommer spørsmålet om hvilke kriterier som er gyldige uavhengig av skiller mellom fag, tema og problemstillinger, og hvilke kriterier som er betinget av slike forskjeller.

Formålet med denne rapporten er å sammenfatte kriterier for evaluering av evalueringer, å kartlegge og sammenfatte noen evalueringer innen samferdselssektoren, søke å lokalisere noen særegne utfordringer disse står overfor, og slik søke å si noe av mer allmenn interesse om evalueringstradisjoner.

Vi vil dermed undersøke:

- Hvilke kvalitetskriterier tas i bruk?
- Hvilke likheter og forskjeller kan vi finne på tvers av evalueringstradisjoner og tema?
- Hvilke særlige utfordringer og dilemmaer står en overfor, når en skal anvende disse kriteriene?

Empirisk tar vi evalueringer gjennomført ved vårt eget institutt, TØI.

Innledningsvis i dette kapitlet skal vi gi en nærmere definisjon og avgrensning av hva vi forstår med evaluering og hvordan det benyttes i denne rapporten. Dette gjør vi ved å risse opp fire ulike perspektiver på evaluering, perspektiver som ikke er gjensidig utelukkende, men som gjør oss i stand til å sette opp grensesteiner i forhold til hva en evaluering ikke er og, ikke minst, hvilke sider ved evaluering vi vektlegger i denne rapporten. Basert på denne gjennomgangen, avsluttes framstillingen med en sammenfatning av sentrale variabler og dimensjoner som vi benytter til å analysere ulike evalueringer.

1.1 Tematisk og foreløpig definisjonsmessig avgrensning

Vår tematiske avgrensning er *evaluering av offentlig politikk* og spesifikt samferdselspolitikk. Vi kan skille mellom en bred og en smal definisjon av begrepet evaluering. En bred definisjon er en systematisk undersøkelse av virkninger og i hvilken grad de er i samsvar med intensjonen. Gitt denne definisjonen inkluderer evaluering også *ex ante* økonomiske analyser, for eksempel av vurderte eller planlagte infrastrukturprosjekter i tillegg til evalueringer av gjennomførte tiltak.

En smalere, innflytelsesrik definisjon, finner vi hos Vedung (1998:11). For han er evaluering en ”systematisk etterhåndsvurdering av utfall, sluttprestasjoner eller forvaltning i offentlig virksomhet, med det formål at denne skal spille en rolle i praktiske beslutningssituasjoner”. Definisjonen avgrenser evaluering til offentlig virksomhet. Til forskjell fra den første definisjonen, inkluderer den *ex post*, men ikke *ex ante* analyser. Samtidig åpner den for evaluering av pågående prosesser. Med begrepet systematisk viser den til kvalitetskrav og forskningsmetodikk, og med henvisning til praktiske beslutningssituasjoner viser den at forskningen skal være relevant. Dette er en definisjon som både dekker iverksettingen/forvaltningen av offentlige tiltak eller virksomhet og effektene av disse.

Vi skal i denne rapporten benytte et begrep om evaluering som er snevrere enn Vedungs. Dels ønsker vi klart å framheve at slike undersøkelser av iverksetting eller administrasjon primært har vår interesse fordi dette påvirker resultatet eller effektene av offentlig politikk. Forvaltningen har ikke vår interesse *per se* – den har interesse kun i den grad den påvirker innholdet i politikken, i form av de prioriteringer som gjøres og de effekter dette får. Nå er det slik, ut i fra en organisasjonsteoretisk antagelse, at det vil forvaltningen oftest gjøre: Valg av organisasjon og måten en iverksetter et offentlig tiltak på har effekter på output.

snarere enn å diskutere 0-hypotesen: At andre forhold er avgjørende. Mens evalueringen i mange tilfeller burde ha ledet til politisk nøkternhet og kanskje til og med resignasjon, blir den en satsplanke for fornyet optimisme.

Den antatte relevansen av en slik hypotese styrkes av at det er et fellestrekk for alle de evalueringene vi har gjennomgått, at de i en viss forstand er teoriimmune. De er preget av at en ikke undersøker om teorien som ligger til grunn for undersøkelsen støttes av resultatet av undersøkelsen. Dette innebærer at det er mulig og sannsynlig at det etableres epistemiske samfunn på tvers av evaluering og forskning, som baserer seg på "common beliefs" som ikke lar seg avkrefte. Dette øker betydningen av tverrfaglig og tverrsektoriell kommunikasjon og kritikk, og dermed betydningen av institusjonelle ordninger som medvirker til dette.

Vi vil imidlertid fremheve at det – i denne sammenheng – bare er interessant som en mellomliggende variabel.

Dette poenget og begrunnelsen for de avgrensninger vi gjør, vil bli nærmere diskutert nedenfor. Vi vil imidlertid allerede her presentere vår avgrensning av evalueringsbegrepet i denne rapporten, der vi gir evaluering en definisjon som også eksplisitt forholder seg til politikk og politisk virksomhet som intenderte handlinger. Vi definerer evaluering som en *systematisk etterundersøking av om effektene av offentlig virksomhet er i samsvar med intensjonene, med det formål at slike undersøkelser skal ha relevans for pågående eller fremtidige beslutningsprosesser.*

1.2 Fire perspektiver på evaluering

Innledningsvis skal vi skissere fire perspektiver på evaluering. Vi benevner disse:

- i) Evaluering som evalueringsmetodikk
- ii) Det informerte målrettede oppdragsgiverperspektivet
- iii) Det statsvitenskapelig informerte meta-perspektivet
- iv) Det profesjons-/kunnskapssosiologiske perspektivet

Perspektivene er stiliserte og forenklete for å få fram noen sentrale aspekter ved ulike måter å definere evaluering på, og de er ikke gjensidig utelukkende. Vårt formål er snarere å avgrense temaet vårt, å sette opp noen merkesteiner, med bakgrunn i hvordan perspektivene ”overlapper”, for slik å avgrense anvendelsesområdet for analysen.

1.2.1 Evaluering som evalueringsmetodikk

Utgangspunktet vårt er at det å foreta en evaluering innebærer å stille det kontrafaktiske spørsmålet: Hva ville ha skjedd dersom tiltaket ikke ble iverksatt? Utviklingen av evalueringsmetodikk i form av meta-analyser er et forsøk på å imøtekomme dette problemet. I Elviks avhandling (1999) om dette emnet tar spørsmålet om utarbeidelse av meta-analyser feste i de etablerte kvalitetskrav vi normalt stiller til forskning per se (som logisk konsistens, etterrettelighet, etterprøvbarehet, generaliserbarhet), men utvider dette med de krav en stiller til det vi her kan kalle normal samfunnsvitenskapelig virksomhet, for presentasjonen av forskningen slik at en kan vurdere intern og ekstern validitet (se for eksempel Rose 1982).

Elvik har formulert et ”tilnærmet objektivt” sett med kriterier for hva om er gode og dårlige undersøkelser i evalueringsforskningen. Med ”tilnærmet objektivt” mener han at kriteriene for hva som er god forskning

1. kan formuleres så klart at de ikke gir rom for sterkt divergerende tolkninger, og at
2. de bygger på normer for god forskningsmetode som har svært bred tilslutning blant forskere, og ikke minst at

3. de er uavhengige av innholdet i resultatene av en undersøkelse og dermed uavhengige av om vi "liker" eller "ikke liker" resultatene.

Kriterier for gode og dårlige undersøkelser i evalueringsforskning er i avhandlingen formulert med utgangspunkt i Cook og Campbells (1979) validitetssystem. Validitet defineres som graden av tilnærming til sannheten. Jo nærmere sannheten vi har grunn til å tro at resultatene av en undersøkelse eller et sett av undersøkelser ligger, desto høyere er validiteten. I avhandlingen presenterer Elvik i alt 20 kriterier for validitet i evalueringsforskning. Kriteriene er knyttet til fire hovedformer for validitet: Statistisk, teoretisk, intern og ekstern validitet. Vi skal her avgrense oss til å presentere de fire hovedformene og de underkriteriene som er mest relevant for denne rapporten.

Tabell 1. Hovedformer for validitet

Validitet	Grunnlag for grad av validitet
Statistisk	Tallmessig nøyaktighet, feilfrihet og representativitet
Teoretisk	Klart formulert teorigrunnlag
Intern	Årsakssammenheng mellom uavhengig og avhengig variabel
Ekstern	Sammenfall og stabilitet i et sett av undersøkelser

Kilde: TØI-rapport 929/2007

I sin framstilling av **statistisk validitet**, (se vedlegg 1), opererer Elvik med ni kriterier, hvorav de sentrale her er representativitet, reliabilitet og analyseteknikker.

Teoretisk validitet betegner i hvilken grad en undersøkelse bygger på et klart formulert teorigrunnlag som forklarer resultatene av undersøkelsen. Kriterier for teoretisk validitet omfatter:

1. I hvilken grad det er formulert et eksplisitt teorigrunnlag for en undersøkelse, for eksempel i form av hypoteser som skal testes
2. Om teoretiske begreper som brukes i en undersøkelse kan operasjonaliseres tilfredsstillende
3. Om teorien som er formulert kan forklare hvordan det undersøkte tiltaket kan virke på det problemet det er ment å løse (eksempelvis trafikkulykker eller personskader i trafikkikkerhetsforskning)
4. Om teorien som ligger til grunn for en undersøkelse støttes av resultatene av undersøkelsen

Intern validitet gjelder spørsmålet om i hvilken grad en undersøkelse, eller et sett av undersøkelser, gir grunnlag for å hevde at det er en årsakssammenheng mellom det undersøkte tiltaket og de endringer som kan påvises i den eller de avhengige variablene. Det er formulert sju kriterier for kausalitet i evalueringsforskning (se tabell 2 i vedlegg). Sentralt i kriteriene for intern validitet er at det, for å kunne snakke om en årsakssammenheng, er en nødvendig, men ikke tilstrekkelig forutsetning at det foreligger en statistisk sammenheng mellom årsaksvariabelen og virkningsvariabelen. Denne statistiske sammenhengen må videre holde ved

kontroll for andre mulige forklaringer/variabler og det må være mulig å kunne spesifisere årsaksmekanismen som forklarer sammenhengen.

Elvik peker også på betydningen av **ekstern validitet**: Muligheten for å generalisere resultatene av undersøkelsen utover den spesifikke konteksten undersøkelsen er utført i, i tid, i rom og eventuelt til andre saksområder. Dette handler om sammenhengen mellom årsak og virkning er reproduisert under varierende betingelser i flere undersøkelser.

Elvik benytter disse kriteriene i meta-analyser av trafikksikkerhetstiltak, i hovedsak gjennom statistiske analyser. Selv om vi i denne rapporten også inkluderer evalueringer av ikke-statistisk art, dvs. ikke kvantifiserte og til dels ikke kvantifiserbare data, inntar vi samme grunnposisjon som Elvik: Evaluering er i første rekke en bestemt metodikk, det er *metoden*, og ikke resultatene som avgjør om det er en god eller dårlig evaluering. En evaluering innebærer her å undersøke effekter av tiltak eller handlinger. Det har et uomtvistelig element av årsaksanalyse ved seg, og enhver evaluering må romme en metodikk for å vurdere eller kontrollere for slike årsakssammenhenger. På dette punktet er Elviks utgangspunkt også vårt utgangspunkt.

Hvor mange av disse kvalitetskriteriene en skal kreve oppfylt for at en skal kunne snakke om en god evaluering, er et annet spørsmål. Det kan være argumenter for at listen over felles kriterier på tvers av fag og metoder bør gjøres kortere. Det gjelder særlig kravet om statistisk validitet, som avgrenser seg til og er forankret i kvantitativ metode og dermed statistiske analyser. En slik forutsetning ville innebære at en utdefinerte kvalitativ evaluering a priori. Vi skal imidlertid ikke i utgangspunktet verken korte ned Elviks liste, eller utdefinere evalueringsdesign som i utgangspunktet bryter med den. Vi legger til grunn disse validitetskravene og kommer tilbake til diskusjonen om hvorfor disse eventuelt bør frafalles eller nyanseres i en diskusjon om kvalitetskrav på tvers av evalueringstradisjoner.

1.2.2 Det informerte målrettede oppdragsgiverperspektivet

Dette perspektivet finner vi typisk formulert i manualer eller veiledere for evaluering, slik disse benyttes av nasjonale myndigheter og innen EU (kfr for eksempel Finansdepartementets veileder eller veileder i "Evaluating EU expenditure programmes" (Nagarajan og VanHeukelen, 1997)).

En typisk formulering om evaluering i dette perspektivet vil være den vi finner i Reglement for økonomistyring i staten: Alle virksomheter skal sørge for at det gjennomføres evalueringer for å få informasjon om *effektivitet, måloppnåelse, og resultater* innenfor hele eller deler av virksomhetens ansvarsområde og aktiviteter (vår utheving).

Med andre ord, her forstås evaluering som en målrasjonell prosess, noe som blant annet betyr at målsetningene må være klart formulerte. Av dette følger også behov for formalisering av styrings- og kontrollparameter som fungerer som gode operasjonaliseringer for måloppnåelse både innen virksomhetsområdet og når en skal evaluere virksomhetsområdet. Dette kan være knyttet til økende grad av tilstrebelser til formalisert målstyring.

Et annet trekk ved dette perspektivet vil være at siktemålet ofte er avgrenset i en noe annen betydning. Oppdragsgivere vil være interessert i effekten av det tiltaket de har ansvar for, og i mindre grad ha interesser knyttet til relevansen av slik kunnskap på andre områder. Dermed vil de også være opptatt av substansiell kunnskap, kunnskap om særegne forhold ved en sektor eller et saksområde som har betydning for tolking av data, kunnskap om når modellene utdefinerer vesentlige forhold, osv

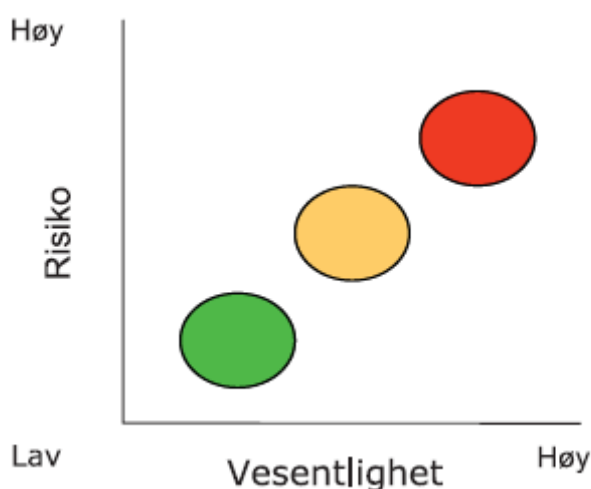
På den andre siden vil dette innebære at de er mindre opptatt av ekstern validitet, dvs. funnens gyldighet i andre kontekster (som land eller sektorer). Dermed vil de nedtone den kontroll av funnene som følger av oppmerksomhet i forhold til ekstern validitet.

Videre vil oppdragsgiverne sitte med beslutningsansvar for spørsmål av typen:

- Når skal en evaluere?
- Hvordan skal en gjennomføre evalueringen (metodikk, ressurser med videre)?
- Hvem skal gjennomføre evalueringen (egen etat, ekstern utfører, eventuell konkurranse i form av anbud)?

Av dette følger også behov for formalisering eller tommelfingerregler for type evalueringer. En måte å sammenfatte slike kriterier for valg av evalueringsdesign, er presentert i vedlegg 3.

Det er også behov for vesentlighetsvurdering: Når er det behov for å foreta en evaluering? Det gjelder slike forhold som saksområdets økonomiske og prinsipielle betydning, herunder om det er et tiltak som, dersom det er vesentlig, kan få konsekvenser for framtidig virkemiddelbruk. En annen dimensjon av relevans er risiko. Det kan være risiko knyttet til politiske forhold, økonomi og finansiering med mer. I Finansdepartementets veileder blir det understreket at vesentlighet og risiko henger sammen, at grensene for hva som er akseptabel risiko ser annerledes ut på områder av høy vesentlighet enn ved lav vesentlighet (se figur nedenfor).



Figur 1. Sammenheng mellom vesentlighet og risiko

Kilde: TØI-rapport 929/2007

Endelig medfører dette, gitt den økte oppmerksomheten mot evaluering innen offentlig sektor, behov for intern kompetanseoppbygging, både som bestiller av evaluering og som evaluerer av egen virksomhet.

Oppsummert er sentrale krav knyttet til evalueringer i et oppdragsgiverperspektiv at disse skal være *målrettede, formaliserte* og fortrinnsvis knyttet til gode styringsparameter. I tillegg vil oppdragsgiver i første rekke være interessert i om en svarer på problemstillingen – det vil si effekten av det evaluerte tiltaket. Det er dermed intern validitet og den sektorspesifikke, substansielle kunnskapen som er av særlig interesse for oppdragsgiver. Ut i fra ønsket om gode resultater og transparens er det videre ønskelig at en skal/bør ha utviklet gode kriterier for når en skal evaluere (vesentlighetsvurderinger), og for valg av evaluerer og evalueringsmetodikk.

I det oppdragsgiverperspektivet vi har skissert her, er det flere typer evalueringer som det offentlige foretar som er utdefinert, som forvaltningsrevisjon/evaluering, det vil si spørsmål om en organisasjon fungerer i samsvar med retningslinjene. Det som gjerne blir omtalt som prosessvurdering eller organisasjonsanalyse er også bare av interesse for oss dersom det gir kunnskap om effektene av tiltaket eller grader av måloppnåelse. Det er helt sentrale retninger innen statsvitenskapelig og sosiologisk forskning, som analyserer institusjonelle og organisatoriske strukturer og prosesser, men der de effekter slike strukturer har for innholdet og prioriteringer i offentlig politikk, er underkommunisert eller i liten grad eksplisert. Antagelsen her, som vi skal komme tilbake til avslutningsvis, er at manglende fokus på effektproblemet innen prosessevaluering (Baklien 2000) er en hovedutfordring innen denne retningen.

Synspunktet over er ikke ukontroversielt. Tilsvarende vil mange – ikke minst med fagbakgrunn i sosiologi og statsvitenskap – hevde at den tolkningen vi her gir av det informerte målrettede oppdragsgiverperspektivet, har to viktige svakheter. For det første fordi en idealiserer oppdragsgivers intensjon og handlingslogikk og de imperativer oppdragsgiver er underlagt. Oppdragsgiver vil i mange tilfeller ha andre siktemål med evalueringen enn å undersøke effekter. Evalueringen kan være et uttrykk for ønske om legitimering, og er mer som symbolsk handling å regne. Dette er et relevant argument dersom vi i denne rapporten var opptatt av å undersøke effekter av evalueringer på offentlige beslutninger, men siden vi er opptatt av kvalitetskriterier som er av relevans for oppdragsgiver og evaluerer, gitt at disse faktisk primært ønsker kunnskap om et tiltak virker eller ikke, er dette irrelevant. Når vi her setter kvalitetskriterier ved evaluering i sentrum, legger vi til grunn at oppdragsgiver baserer seg på et målrasjonelt perspektiv.

For det andre kan en argumentere for at vi idealiserer, eller rettere innskrenker, rasjonalitetsforutsetningen og entydig knytter evaluering til målrasjonell virksomhet, mens en kan/bør stille seg andre målsetninger, for eksempel utvikle bredere forståelse eller ha som fokus læring på lengre sikt. En annen innvending består i at det er vanskelig å skille mål fra verdier, siden enhver saksvurdering har et normativt element. Også dette er adekvate argumenter i en generell diskusjon om evaluering, men vi holder oss på målrasjonalitetens arena. I denne rapporten diskuterer vi kvalitetskriterier til evalueringer, gitt oppdragsgivers verdier og målhierarki. I dette kan ligge at evaluerer medvirker til at oppdragsgiver gjør sine verdier mer eksplisitt for seg selv, og at evaluerer er kjent med sine egne

normative forutsetninger. Vi legger imidlertid til grunn det etablerte, om enn omdiskuterte, vitenskapssynet som tar utgangspunkt i sontringen mellom sak og vurdering (jfr. Weber 1971:161f)

Det målrasjonelle perspektivet støter imidlertid på en bestemt type tilbakevendende problemer: Problemer knyttet til (utilsiktede) sideeffekter og ikke-tematiserte konkurrerende målsetninger. Mange offentlige inngrep som er gjort for å realisere en bestemt målsetting vil ha sideeffekter som kan påvirke realiseringen av andre mål og/eller andre viktige samfunnsmessige verdier. I mange tilfeller vil slike mål være forankret i andre deler av offentlig virksomhet. Mens Vegdirektoratet skal fremme trafikksikkerhet og vil være opptatt av trafikkovervåkning, har Datatilsynet ansvar for å motvirke unødig overvåking. Et (for) snevert oppdragsgiverperspektiv vil kunne medføre at en bare er opptatt av effekter innen eget myndighetsområde. Vi legger derfor til grunn at det er hensiktsmessig å inkludere slike sideeffekter i analysen, og dermed effekter på alle relevante variabler for ulike målsetninger innen offentlig virksomhet.

Et annet problem er begrensninger i antall avhengige variabler: I noen tilfeller vil oppdragsgiver være opptatt av å ikke inkludere en rekke potensielt vesentlige uavhengige variabler i analysen, fordi disse variablene ligger under andre organers myndighetsområde eller av andre grunner ikke ønskes tematisert. Et eksempel ville være dersom Samferdselsdepartementet søkte å undersøke effekter av offentlig kjøp av lokal kollektivtransport uten å undersøke selve finansieringsordningen, fordi denne ble administrert av Kommunaldepartementet og ikke kjøpsorganet. Et annet eksempel ville være å undersøke effekter av en reform med overføring av ansvarsområde (for eksempel som forsøksordningen for kollektivtransport i by) uten å forholde seg til de ansvarsområdene og ressursene som ikke ble overført. Dersom slike uavhengige variabler som ekskluderes fra analysen i realiteten kan forklare mye av variansen på den avhengige variabelen, mangler de faglige forutsetningene for en adekvat analyse.

Vi legger til grunn at et informert målrettet oppdragsgiverperspektiv impliserer et adekvat grunnlag for beslutninger, dvs. at alle vesentlige variable kan inkluderes i analysen. Det er grunn til å anta at det informerte oppdragsgiverperspektivet vil være opptatt av kumulativ kunnskap i betydningen systematisk kunnskap om årsak- og virkningsforhold av effekter innen eget ansvarsområde og eventuelle sideeffekter for annen offentlig virksomhet. De vil primært være opptatt av at kunnskapen er valid for gjeldende saksforhold. Derimot er det grunn til å anta at oppdragsgiverne sjeldnere vil være opptatt av den teoretiske relevansen – og dermed den eksterne validiteten til en undersøkelse. Imidlertid er det nettopp gjennom kontroll for ekstern validitet en får kritisk etterprøvd også funnene i den konkrete, foreliggende evalueringen.

Oppsummert: Vi inkluderer vi i denne rapporten prosess- og organisasjonsanalyse i arbeidet i den grad de er viktige mellomliggende variabler for å forklare effekter på andre variabler. Vi diskuterer ikke organisasjonsanalyse per se, heller ikke oppdragsgivers bruk av evalueringen (om denne har effekter eller ikke), men vi inkluderer begrepet utilsiktede sideeffekter som del av evalueringens domene.

1.2.3 Det statsvitenskapelige informerte meta-perspektivet

Vedungs definisjon og tilnærming

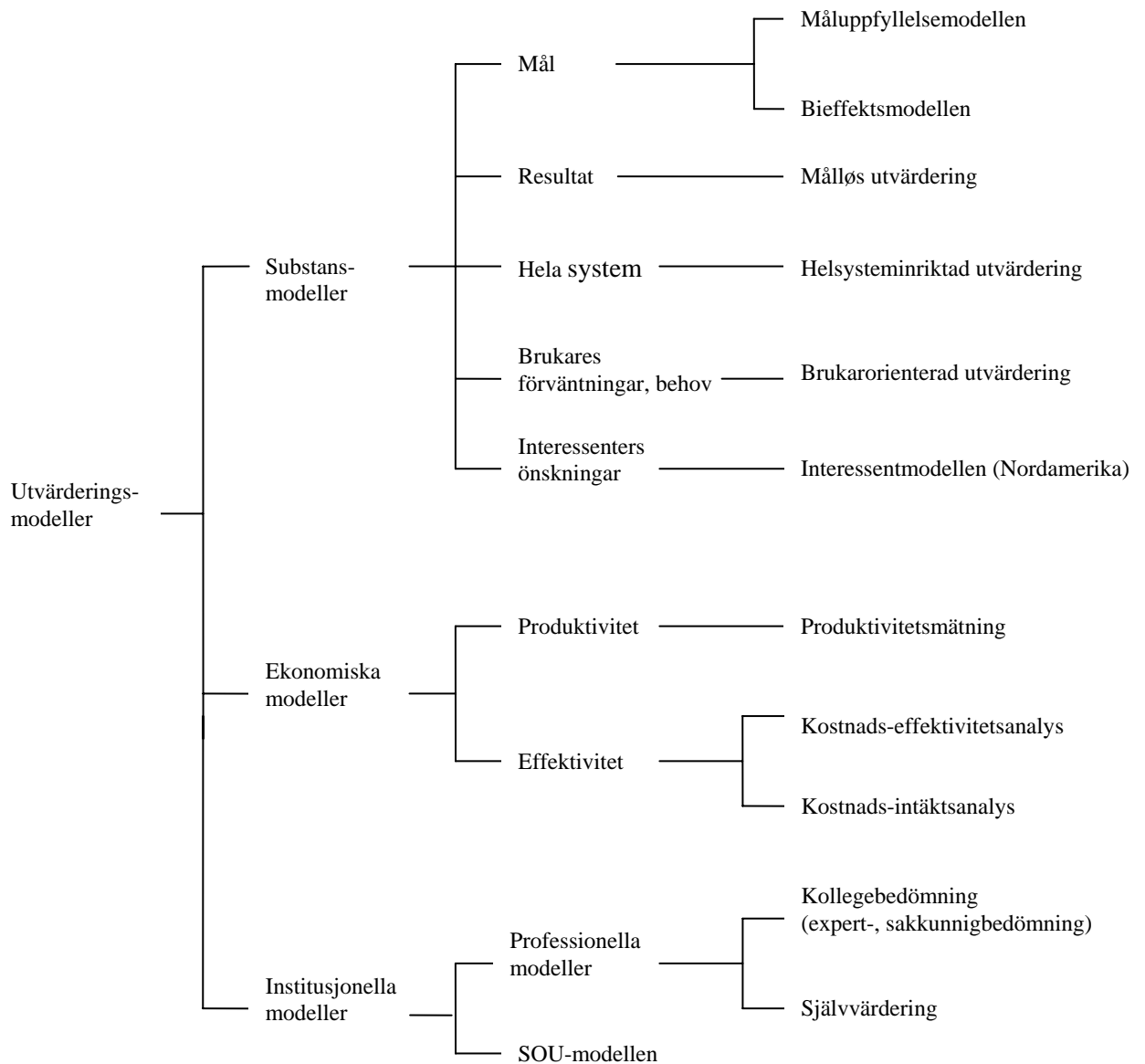
Som vist i definisjonen innledningsvis, avgrenser Vedung evaluering fra ex ante vurderinger. Slik står arbeidet i kontrast til den type evaluering vi finner vektlagt i deler av forvaltningen, herunder i både Finansdepartementets veileder og den som preger Senter for statlig økonomistyrings (SSØ) arbeid. For vårt formål er dette en rimelig avgrensing, i det vi avgrenser oss til evalueringer i den betydningen at de må være basert på foreliggende empiri.

To innvendinger bør nevnes: I en modellering av en bestemt karakter, som kan ligge til grunn ex ante, er erfaringene ”bygd inn i” modellen. Dette er nettopp rasjonale for den kontinuerlige ex ante-ex post prosessen; at modellen skal raffineres. I vår avgrensing ligger nettopp at input i modellen baserer seg på nye data. Spissformulert er vår forestilling om empirisk kunnskap at den ikke omhandler det som ennå ikke har hendt. (Dermed har vi ikke sagt at det ikke er forskjell på en vitenskapelig basert prediksjon og eventuelle fordommer, bare at den substansielle forskjellen (til forskjell fra den formale) bare lar seg undersøke basert på empiri.)

En særegen problemstilling vedrørende ex post-begrepet gjelder analyser av beslutninger i det politisk-administrative systemet. Slike kan være av flere slag, avhengig av hvor langt de er kommet i beslutningsprosessen. Utfallet av en politisk beslutningsprosess kommer først i form av vedtak. Disse foreligger som analyserbare data. Neste trinn er at disse implementeres gjennom det administrative apparatet for eksempel gjennom retningslinjer eller tildelingsbrev. I neste trinn er de observerbare tiltak, for eksempel i form av investeringer, iverksetting av et trafikkikkerhetstiltak, bygging av en støyskjerm. I neste trinn gir slike tiltak effekter på handlingsmønstre utenfor det politisk-administrative systemet (for eksempel reisevaner), og i siste trinn som målbare effekter på forhold som miljø, dødsulykker osv. Gitt vår avgrensing ovenfor, retter vi i denne rapporten oppmerksomheten mot evalueringer som ikke bare undersøker beslutningsprosessens effekter i form av vedtak, men også inkluderer observerbare effekter utenfor det politisk-administrative systemet. Dermed avgrenser vi oss i all hovedsak også fra ex post analyser av beslutningsprosesser som bare har effekter i form av vedtak.

Vedungs perspektiv er preget av en vektlegging av pluralisme, av ulike typer evalueringer for ulike formål og presenterer en bredt anlagt typologi som illustrerer ulike typer evalueringer (se figur under). Skillet mellom substansmodeller og økonomiske modeller er at de førstnevnte ikke vurderer kostnader. De fleste typer evalueringer som Vedung (1998:50-85) diskuterer, bortsett fra de institusjonelle, vil være inkludert i det evalueringsbegrep vi legger til grunn her. Et etablert skille som ikke er framtrødende i Vedungs diskusjon, er skillet mellom formativ og summativ evaluering (se Foss 2000 og Baklien 2000). Summativ evaluering er identisk med det vi er har omtalt som effektevaluering, mens formative evalueringer gjennomføres i tilknytning til gjennomføring av tiltaket, for å støtte implementering av tiltaket. Slik sett har det likhetstrekk med prosessevaluering, men prosessevaluering kan også inngå i en effektevaluering – og gi forklaring på effektfordeling. Vi inkluderer den siste typen

prosessevaluering i vår drøfting her, en drøfting som også vil åpne for en kritisk drøfting av prosessevaluering uten effektmål.



Figur 2. Vedungs oversikt over ulike evalueringsmodeller

Kilde: Vedung, 1995:50

Kravet om pluralisme, som følger av Vedungs tilnærming, kan gis en sterk og en svak betydning. I den svake betydningen innebærer den at en anerkjenner at ulike typer problemstillinger og datatilfang bør tilsi ulike metoder eller tilnærminger. I en sterkere betydning impliserer det at en bør ta i bruk flere ulike metodiske og teoretiske innfallsvinkler til undersøkelsen av ett og samme fenomen.

Et slikt imperativ om metodetriangulering (Grønmo, 1996) – så vel som teoritriangulering – er begrunnet i at flere perspektiver og metoder gir et sterkere empirisk grunnlag når funnene er sammenfallende, og innebærer samtidig at en mer systematisk kan undersøke i hvilken forstand funnene er formet av teori og metoder, og dermed initiere spørsmål knyttet til kunnskapsområdets grenser.

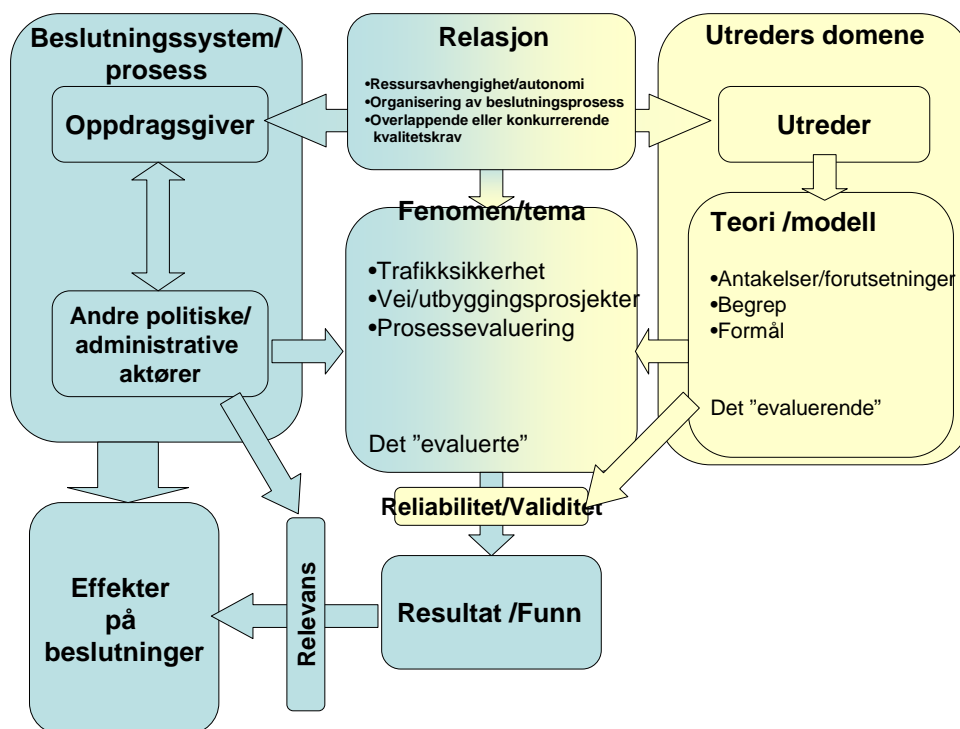
Når vi har tatt utgangspunkt i Vedungs arbeid innen den statsvitenskaplig forankrede systematiseringen av evalueringer, er det dels fordi den er bredt anlagt og søker å være uttømmende, og slik står som pionerarbeid i Norden. Dels er det også fordi sontringen mellom meta-analyse (forstått som vår konseptualisering og systematisering av analyser) og analyse er klarere hos Vedung enn i mange andre statsvitenskaplig forankrede arbeider. I den sammenheng finner vi grunn til å skille mellom meta-analyser av evalueringer og analyser av evalueringsprosesser.

Nærmere om sentrale dimensjoner ved evalueringer

Skillet mellom meta-analyser av evalueringer og analyser av evalueringsprosesser kan utdypes gjennom figuren nedenfor. I denne figuren gjør vi et skille mellom hhv utreders domene og det politisk-administrative domene, dvs. beslutningsprosesser. Kjernen i statsvitenskaplig virksomhet og analyse er den venstre siden i denne figuren, dvs. analyse av beslutningsprosesser og strukturer. De spørsmål en da stiller er av typen: Påvirker evalueringer beslutninger? Da blir kjennetegn ved beslutningssystemet, viktig. Er det for eksempel preget av mål-middelrasjonalitet, eller har bruken av evalueringer primært en symbolsk-legitimerende side (se for eksempel Dahler-Larsen 2005:25 osv.). Støtter evalueringenes funn mot etablerte interesser innen det politisk-administrative systemet?

Dette er interessante og høyst relevante spørsmål, men kjernen i denne rapporten er høyre side i figuren, dvs. hva som kjennetegner evaluers virksomhet og hvilke kvalitetskriterier disse tar i bruk. I vår sammenheng er meta-analyser bare av interesse for evalueringer og utvikling av kvalitetskrav i evalueringer. Vi er ikke, i denne rapporten, primært interessert i om de får effekt i beslutningsprosessen. Ei heller er vi interessert i statsvitenskaplig eller sosiologisk forankrede undersøkelser om hvorfor evalueringer ikke får effekt for politisk beslutningsprosess per se. Unntaket er i tilfeller der beslutningsprosessen er undersøkelsesobjektet for evalueringen. Det vil det også være en del eksempler på når en ser på bestemte typer prosess- og organisasjonsevaluering.

Figuren under viser mulige systemsammenhenger i en evalueringsprosess.



Figur 3. Systemsammenhenger i en evalueringsprosess

Kilde: TØI-rapport 929/2007

Oppsummert henter vi fra det statsvitenskapelig forankrede meta-perspektivet følgende. Vi ekskluderer ex ante undersøkelser, og avgrenser oss til underveis- og etterundersøkelse. Vi tar feste i et pluralistisk perspektiv som rommer ulike fagtradisjoner og metoder, inkludert metoder som inkluderer kvalitative data.

1.2.4 Det profesjons- og kunnskapssosiologiske perspektivet

Vår primære interesse i denne sammenhengen er innholdet i figurens høyre side, hva er kriteriene som er i bruk av ved evalueringer og hvordan forbedre kvaliteten ved evalueringer?

Vi legger til grunn at variasjon mellom evalueringer er knyttet til:

- Det de omhandler, dvs. deres formål eller tema
- Teori/modell som de benytter
- Metode som blir benyttet

Vi legger her til grunn det etablerte kunnskapssynet etter Kuhns banebrytende arbeid (1962), som innebærer at sannhet i en viss forstand er relativ til vitenskaplige paradigmer. Paradigmer er konstituert av teorier i form av begreper som nettopp griper spesielle aspekter ved virkeligheten. Normalvitenskaplig aktivitet innebærer å innordne funn og anomalier i den modell en benytter, mens dersom disse blir for store, vil det oppstå revisjon av hele paradigmet, og et nytt oppstår. Vitenskapssosiologisk forskning viser at det er et fremtredende trekk ved vitenskaplig aktivitet at en ikke tester de grunnforutsetninger som ligger til grunn

for teorien, men nettopp kartlegger sammenhenger innen det etablerte paradigmet til og med innen naturvitenskapen (se for eksempel Latour og Woolgar, 1979). Det er grunn til å anta at dette ikke er mindre framtreddende innen samfunns- og menneskevitenskapene, som jo gir enda større grunnlag for tolkingsmangfold, blant annet fordi studieobjektet er, som Charles Taylor (1995) har formulert det, ”det selvtolkende dyr” som dermed påvirkes av forskning på det og begreper om det.

Med andre ord, gitt vår vitenskapssosiologiske kunnskap, er de funn som gjøres og presenteres innen evalueringer, ikke bare et resultat av trekk ved studieobjektet; studieobjektet blir ”formet” av teorier og modeller. Vi legger imidlertid her ikke til grunn en radikal kunnskapsrelativistisk posisjon, slik vi bla finner den i den såkalte Edinburg-skolen (Barnes and Bloor 1982), om at det studerte objektet ikke setter grenser for hvilke tolkninger vi kan gjøre. Vi antar at vitenskaplige eller evaluerende utsagn er mer eller mindre gyldige eller adekvate, men vi antar også at deres gyldighet og ikke minst gyldighetsområde er definert av det teoretiske utgangspunktet.

Dermed åpner vi også for, i større grad enn det vi finner hos Vedung, at det er forskjeller mellom ulike disipliner/fag i hvordan bestemte, også likeartede fenomen blir definert eller fortolket. Det bryter også med ”lærebokforståelsen” av evaluering, slik vi finner den blant annet hos Sverdrup (2002) Variabelen *utreders teori/modell* i modellen over, viser til at det er variasjon mellom disipliner, når det gjelder hvordan en begriper et fenomen. Dette har også andre implikasjoner. Det er en veletablert antagelse innen statsvitenskap at all organisering er mobilisering av ”skjevhet eller partiskhet” (Schattschneider, 1960): Noen interesser og verdier blir definerte inn – andre ut. Gitt en kunnskaps sosiologisk tilnærming, kan en anta at også variasjon i disiplin kan være mobilisering av ”partiskhet”. Ulike profesjoner/disipliner tar i bruk ulike begreper som er ”briller” som de bruker til å betrakte virkeligheten, og disse begrepene er sjelden verdinøytrale, de har større affinitet til noen verdier enn andre. Dermed fremmer de også interessene til de som har oppmerksomhet om disse verdiene.

Slik sett åpner denne modellen for sterkere fokus på forholdet mellom profesjoner/disipliner og evalueringstradisjoner. En profesjon er kjennetegnet av å ha et bestemt kunnskapsgrunnlag, som er esoterisk (Abbott 1988), dvs. ikke lett tilgjengelig. Det krever lang trening, utøverne er valgt ut gjennom klare seleksjonsprosesser. En profesjon er i tillegg til felles, sertifisert utdanning kjennetegnet av stillingsmonopol på bestemte stillingstyper, slik at de utøver jurisdiksjon innen offentlig sektor. Profesjoner vil gjerne være representert i offentlig sektor, innen forskning og evaluering og i det private næringsliv, og representere kunnskapssystem som for medlemmene er gjenkjennbart på tvers av sektorer.

I et slikt perspektiv blir det viktig å spesifisere hvilke antagelser og forutsetninger ulike evalueringer er basert på, og hvordan de teoretisk begriper det de studerer.

Det følger av denne forståelsesrammen at alle utredninger bygger på en ”modell”, dvs. et stilisert bilde av virkeligheten som snitter ut noen aspekter, mens de utelukker andre. Spørsmålet er da dels hvor bevisst evaluererne er denne modellen og de trekk den har, og dels hvor formalisert den er. Her kan en grovt skille

mellom modeller med høyt abstraksjonsnivå som skal fungere på tvers av sektorer og modeller som er sektor- eller substansforankret.

Selv om vår interesse i den rapporten primært er utreders domene, har visse sider ved eller i alle fall relasjoner i den venstre siden av figuren også relevans, dette gjelder særlig når en skal evaluere evalueringer.

Dermed kommer *relasjoner mellom oppdragsgiver og evalueringer* i sentrum: Hvordan er rollefordelingen i utformingen av ulike deler av evalueringsprosessen; valg av problemstilling, metode, analyse og tolking av data? Slike prosessuelle variabler må videre knyttes til de mer strukturelle: Hvordan er ressurs- og maktfordelingen, mellom evalueringer og oppdragsgiver. Sentralt her er evalueringens autonomi.

Spørsmålet om overlappende eller konkurrerende kvalitetskrav viser til graden av innforståthet i kvalitetskrav mellom oppdragsgiver og evalueringer. Homogenitet på dette området vil ofte være knyttet til "epistemic communities" (Haas, 1992) som går på tvers av administrasjon og forskning. Et epistemisk samfunn er et nettverk av kunnskapsbaserte eksperter eller grupper med et autoritativt krav på politisk relevant kunnskap innen deres kunnskapsområde. Som nevnt ovenfor er ulike profesjoner (som jurister, økonomer, ingeniører) gjerne lokalisert både innen administrasjon (som oppdragsgivere) og som forskere/eksterne utredere (evalueringer). Dersom det er forskjellige, profesjonsspesifikke kvalitetskrav, og dersom disse i ulik grad er formaliserte, kan vi anta det vil ha effekter. Men det er også grunn til å anta at det oppstår mindre "kunnskapssamfunn" av administrativt ansatte, utredere og eventuelt praktikere knyttet til mer avgrensede problemstillinger, og der disse deler felles forståelsesrammer om hvilke problemstillinger som er viktige og hva som er kunnskapsstatus på området.

Oppsummert tar vi, fra det profesjons- og kunnskapssosiologiske perspektivet, med en forståelse av at teorier og modeller som er i bruk i evaluering er med og former evalueringens objekt. Dette innebærer at både problemstillinger, og til dels også funn og konklusjoner er teoriinformerte – og dermed disiplininformerte. Implikasjonen er at ikke bare kan den konkrete utformingen av validitetskrav variere, men også at kriteriet er relativt til disiplinen. Ulike teorier og modeller kan dermed medvirke til ulike funn eller vurderinger – som kan medvirke til interdisiplinære kontroverser, eller kontroverser mellom ulike evalueringer. Vi legger imidlertid også til grunn at slike kontroverser kan diskuteres rasjonelt, dvs. at i de tilfeller der det ikke er klare empiriske funn som styrker eller svekker en teori på bekostning av en annen – så kan en, med utgangspunkt i alminnelige krav til logisk og konsistent vitenskaplig argumentasjon, lokalisere de punkter i ulike teoretiske posisjoner som er avgjørende for kontroversen.

1.3 Oppsummering av analytisk rammeverk og rapportens videre oppbygning

Formålet med denne rapporten er å kartlegge og sammenfatte ulike evalueringer innen samferdselssektoren, med det formålet å sammenligne og vurdere gyldigheten av ulike evalueringskriterier og ulike evalueringer. Sentrale spørsmål er dermed:

- Hvilke kvalitetskriterier tas i bruk?
- Hvilke likheter og forskjeller kan vi finne på tvers av evalueringstradisjoner og tema?
- Hvilke særlige utfordringer og dilemmaer står en overfor, når en skal realisere disse kriteriene?

Vi avgrenser oss til evalueringer som tar utgangspunkt i en målrasjonell tenking, dvs. som gjelder effektene av et tiltak på et bestemt fenomen utenfor forvaltningen. Vi avgrenser oss videre til etterundersøkelser som har et metodisk design for å måle eller vurdere endring som følge av tiltaket, men inkluderer prosess- eller organisasjonsanalyse som forholder seg til et slikt imperativ.

I det følgende skal vi derfor gå gjennom ulike tradisjoner, for å se i hvilken grad vi kan spesifisere hvilke typer kvalitetskrav som er operative, og søke å formulere hypoteser om hvordan disse kommer under press eller er underutviklet i bestemte tradisjoner og relasjoner.

Blant de sentrale temaene som vil bli tatt opp i avslutningskapitlet, er hvilke likheter og forskjeller vi kan finne på tvers av forskningstradisjoner og tema, og hvilke særlige utfordringer og dilemmaer en står overfor når en skal realisere kvalitetskriterier i ulike fagtradisjoner.

For dette formålet har vi, basert på de fire perspektivene på evaluering som vi har skissert ovenfor, laget en enkel modell som skisserer et knippe av dimensjoner eller variabler som kan utgjøre en matrise for sammenfatningen av ulike evalueringer og evalueringstradisjoner.

- Hva er *temaet, problemstillingen eller formålet* med evalueringen?
- Hva kjennetegner *metode og datagrunnlag* (kvalitative eller kvantitative data, eventuelt type underkategori)? Hvor omfattende er undersøkelsen, hvor mange variabler og typer variabler har den?
- Hva kjennetegner *teorien* som blir benyttet, hvilke *forutsetninger* bygger den på, hvilke *antagelser* søker den å undersøke, hva er de sentrale *begrepene*, har den en klar *disiplinforankring*, hva er dens *gyldighetsområde*?
- Er mulige/sannsynlige utilsiktede effekter godt dekket?
- Kumulativitet: Hvordan antas evalueringen å kunne føre til ny kunnskap (substansielt/metodisk/teoretisk)?

Kvalitetskriterier: Hvilke kvalitetskriterier tar de ulike evalueringene i bruk? Utgangspunktet vårt, målestokken for en slik vurdering, er Elvik sine kriterier vedrørende *statistisk validitet, teoretisk validitet, intern og ekstern validitet*. I den grad de ulike evalueringene avviker fra Elviks liste, hva er begrunnelsene?

Om utvalg – eller ”forskere sitter aldri i glasshus”

De fleste av evalueringene som blir gjennomgått i dette dokumentet er gjennomført på TØI. Det mest framtreddende unntaket er en gjennomgang av forskningen om mega-prosjekter. Bakgrunnen for at denne forskningen blir løftet fram er dels at store prosjekter bør underlegges evaluering, fordi de omfatter store økonomiske ressurser og dels at denne forskningen representerer et nyttig korrektiv til norske evalueringer på området, som har fokusert på forbedring av metodikken.

De evalueringene som ellers blir tatt fram, er eksempler på evalueringer som representanter for (de daværende) avdelingene på TØI tok fram som sentrale: Evaluering av tiltakspakkene, evaluering av TP 10 og nyttekostnadsetterundersøkelser. En annen avdeling trekker fram Elviks arbeid innen evaluering av trafikksikkerhetstiltak og meta-analyser, slik denne ikke minst kommer fram i Elvik og Vaa (2004). Vi har her valgt å ta utgangspunkt i Elviks tilnærming som et generelt rammeverk, og ikke å gå inn på den type evaluering – metaanalyser – han selv har gjennomført med bakgrunn i dette rammeverket. Metaanalysen representerer en metode som stiller krav til evalueringene som i utgangspunktet utdefinerer rene kvalitative evalueringer, som vi inkluderer i denne rapporten.

Fordelen med dette utvalget har vært at det skal representere stor bredde langs de dimensjoner vi har skissert ovenfor og slik gir et godt grunnlag for å diskutere relevansen av ulike kvalitetskriterier. De representerer også evalueringer som har hatt og har konsekvens for forståelsen av de feltene som de omhandler. Det kan være ulemper med å analysere evalueringer foretatt av den organisasjon en selv inngår i – en kan bli ukritisk. På den andre siden ville et utvalg fra andre forskningsmiljøer der en påpekte kritiske punkter i denne type evaluering, kunne oppfattes som strategiske utvalg: Evalueringsfeltet er et felt preget av konkurranse og strategisk agering. Desto viktigere er normen om at i forskning og evaluering ”sitter ingen i glasshus” – de normer og kriterier som vi søker å opprettholde kan ikke rensoneres på i kritikk av andre fordi de vil ramme en selv. Tvert om er enhver type kritikk også en implisitt selvkritikk. Få eller ingen rapporter er perfekte, spørsmålet er mer om de er åpne om sine svakheter. Dette til tross, det er grunn til å anta at det oppfattes som mindre problematisk dersom en kritisk gjennomgår evalueringer fra egen forskningsinstitusjon.

I valget mellom disse alternativene har vi vurdert det som riktig å gjennomgå evalueringer gjennomført på TØI. Gjennomgangen er ment å være kritisk-konstruktiv. Vi prøver å kartlegge hvilke kriterier som blir benyttet i evalueringer, grad av bevissthet knyttet til dette, og dermed kritiske punkter for denne evalueringen. Slike kritiske punkter er framhevet fordi de kan sees som vesentlig for den typen evaluering eller evalueringstradisjon denne inngår i. Det følger av dette at vi ikke i detalj går gjennom alle aspekter ved de ulike evalueringene, som reliabilitet ved data, men vi drøfter primært hvordan dette blir adressert i evalueringsdesignet.

Den videre oppbygningen av rapporten er som følger. I kapittel 2 går vi gjennom evalueringen av de såkalte tiltakspakkene for kollektivtransport. I kapittel 3 går vi gjennom to evalueringstradisjoner knyttet til store samferdselsprosjekter, ex post nyttekostnadsanalyse, og Aalborg-miljøets undersøkelse av samme tema, som også tar opp organisatoriske og psykologiske forhold. I det fjerde kapitlet går vi

gjennom ulike varianter av prosess- eller organisasjonsanalyse, med basis i TP 10-prosessevalueringen og en evaluering som også inkluderte effektvariabler, ”Friskbuss”-evalueringen. I det femte og siste kapitlet oppsummerer vi hovedfunn og drøfter forholdet mellom de kriterier som er skissert med bakgrunn i Elvik og kriterier som faktisk er i bruk, kritiske punkter i ulike tradisjoner og implikasjoner for videre arbeid på dette feltet.

2 Evaluering av tiltakspakker for kollektivtransport

I dette kapittelet skal vi gå gjennom det evalueringsopplegget som ble gjennomført på elleve tiltakspakker i norske byområder. Det vil bli sammenfattet og klassifisert i samsvar med det analytiske skjemaet vi presenterte i innledningskapitlet: Hva er tema, problemstilling og formål med evalueringen? Hva kjennetegner teori, metode og datagrunnlag? Hvordan antas evalueringen å føre til ny kunnskap? I hvilken grad er designet i samsvar med kriteriene om teoretisk, statistisk, intern og ekstern validitet?

2.1 Bakgrunn

I 1991 ble "Forsøksordningen for utvikling av rasjonell og miljøvennlig transport" opprettet av Samferdselsdepartementet. Hensikten var å styrke kollektivtransporten. Fra å innledningsvis være rettet mot enkeltprosjekter, ble innsatsen i 1996 mer dreid mot at pakker av tiltak skulle kunne motta tilskudd. Ordningen "Tilskudd til utvikling av rasjonell og miljøvennlig transport" (populært kalt "Tiltakspakker") medførte at det fra 1996 til avslutningen av prosjektet i 2000 ble bevilget ca 86 millioner kroner fra Samferdselsdepartementet. Dette utgjorde halvparten av de økonomiske bidragene til de aktuelle prosjektene, mens lokale aktører finansierte den andre halvparten. Fylkeskommunene var ansvarlige for søknadene, mens det ikke var krav til hvordan den lokale organiseringen skulle utformes.

18 tiltakspakker fikk støtte innenfor denne ordningen, og av disse var 11 tiltak gjennomført i byområder. Samferdselsdepartementet ga Transportøkonomisk institutt oppdraget om å utarbeide og gjennomføre evaluering av ordningen. Det var imidlertid forutsatt av deler av evalueringen skulle gjennomføres lokalt. Av den grunn ble 1/3 av midlene fra departementet først utløst når lokal evaluering var gjennomført, rapportert og godkjent av departementet.

Vi begrenser oss til å se på evalueringen av de 11 tiltakspakkene i byområdene. Disse fant sted i Tønsberg og omegn (Vestfold), Hundvåg (Rogaland), Drammensregionen (Buskerud), Nedre Glomma (Østfold), Larvik (Vestfold), Ålesund og Giske (Møre og Romsdal), Grenland (Telemark), Lillehammer og Gjøvik (Oppland), Tromsø (Troms), Trondheim (Sør-Trøndelag), Kristiansand (Vest-Agder). Tiltakene som fikk støtte var av forskjellig karakter. Det var tale om ruteendringer, introduksjon av nye tilbud, infrastrukturutvikling eller tilpasning og informasjon/markedsføring.

"Disse 11 tiltakspakkene er forskjellige med hensyn til hvilke typer tiltak som er gjennomført og også med hensyn til omfang: På Hundvåg, i Grenland, i Trondheim og i Kristiansand er det satset målrettet på opprydding i rutestrukturen, med økt frekvens på trafikkfaste strekninger og opprusting av holdeplasser. I de andre områdene er tiltakene noe mer spredt. Det er gjennomført noe frekvensendring på enkelte ruter, opprusting av terminaler og holdeplasser mv. I

alle tiltakspakkene er det gjennomført markedsførings- og informasjonstiltak. I Kristiansand og Tromsø er det innført sanntidsinformasjon på deler av rutenettet (Kjørstad og Norheim 2005).”

”Tiltakspakkene er forskjellige med hensyn til hvilke typer tiltak som er gjennomført, og omfanget av tiltakene. Noen tiltakspakker har gjort få endringer i selve rutetilbudet (rute- og frekvensendringer), mens andre har lagt hovedvekten på dette. Enkelte pakker har hovedsakelig gjennomført tiltak på infrastrukturens side, blant annet i form av holdeplassutbedringer, etablering av nye leskur eller tiltak på knutepunkt og terminaler (Kjørstad og Norheim 2005:1).”

TØI hadde flere roller i forbindelse med evalueringssopplegget: Å utarbeide felles opplegg for basisevaluering, løpende veiledning og bistand om evalueringen, samle lokale grunnlagsdata i felles databaser og foreta samlede analyser av tiltakspakkene.

Evalueringene er rapportert i ca 10 forskjellige rapporter, datert fra 1998-2005. Selv om disse kan oppsummeres hver for seg, skal vi her gi dette benevnelsen evalueringen, og primært rette hovedoppmerksomheten mot den samlede analysen og sammenfatte de mer overordnede funnene. Sammenfatning av evalueringen er hovedsakelig basert på følgende dokumenter:

1. Renolen, Heidi (1998): Kollektivtransport. Praktisk evalueringsveileder for tiltakspakker i by. TØI rapport 388/1998
2. Kjørstad, Katrine Næss og Bård Norheim (2005): Hva tiltakspakkene for kollektivtransport har lært oss. TØI rapport 810/2005
3. Kjørstad, Katrine Næss og Bård Norheim (2005): Tiltakspakker for kollektivtransport 1996 – 2000. Befolkningens vurdering av tiltakene og effekter på reisemiddelvalg. TØI rapport 794/2005
4. Norheim, Bård og Katrine Næss Kjørstad (2004): Tiltakspakker for kollektivtransport 1996 – 2000. Kollektivtrafikantenes vurdering av tiltakene og endret bruk av buss. TØI rapport 736/2004
5. Fearnley, Nils og Åse Nossun: Tiltakspakker for kollektivtransport 1996-2000. Samfunnsøkonomiske analyser

Den første beskriver evalueringssopplegget, den andre oppsummerer hovedfunn fra evalueringen, samt erfaringene med evalueringssopplegget. Den tredje gir en grundigere beskrivelse av gjennomføring av RVU/panelundersøkelser og den fjerde gjør det samme for brukerundersøkelsene (BRU). RVU og BRU er de viktigste og mest brukte undersøkelsene i den samlede studien. Den femte gjennomfører en nyttekostnadsanalyse av tiltakspakkene.

2.2 Evalueringens tema, teori og metodikk

Temaet for evalueringen er virkninger av tiltakspakker for forbedret kollektivtransport (et sett av virkemidler gjennomført parallelt innenfor et begrenset geografisk område). Målsettingen for de fleste tiltakspakkene for kollektivtransport i byområder er et mer effektivt kollektivtilbud både for trafikantene og for selskapene, samt å få flere til å reise kollektivt. To hovedproblemstillinger knyttet til evalueringen vil dermed være å besvare følgende spørsmål (736/2004):

- I hvilken grad tiltakene har bidratt til et bedre tilbud?
- I hvilken grad de har ført til endret reisemiddelvalg?

I 810/2005 (side 8) sies det at "Målet med **den samlede** evalueringen er å finne de isolerte effektene av tiltakspakkene, det vil si de effektene som skyldes tiltakene" (vår utheving).

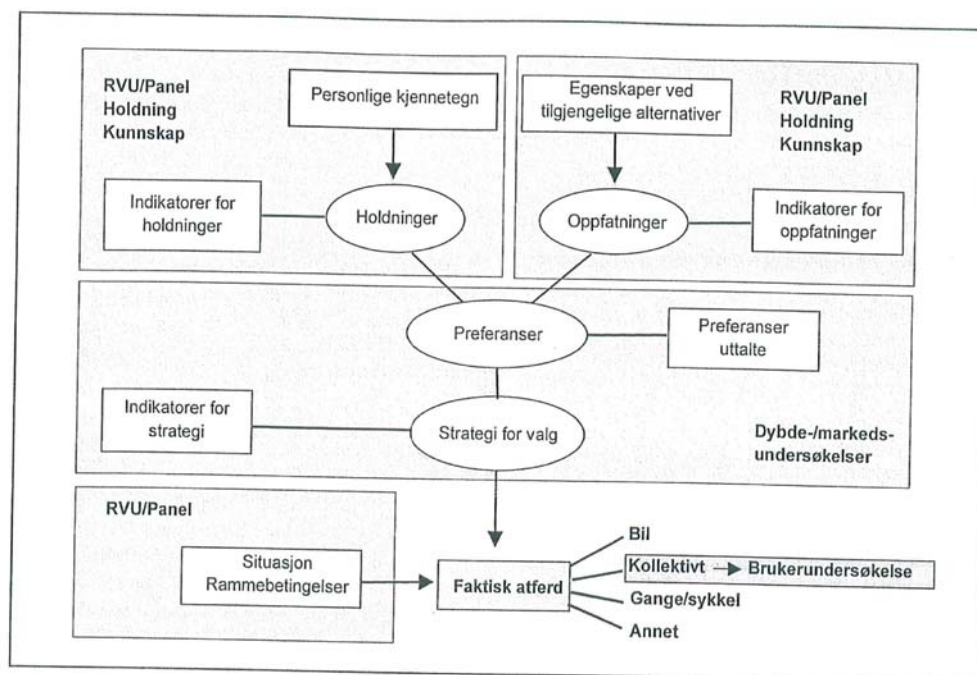
2.2.1 Problemformulering/forskningsspørsmål

I det felles evalueringsopplegget for bytiltakspakkene (TØI-rapport 810/2005) er det valgt å sette fokus på følgende problemstillinger:

1. I hvilken grad har tiltakspakkene ført til at trafikantene har endret reisemiddelvalg og/eller totalt reiseomfang?
2. Hvordan vurderer trafikantene de ulike tiltakene, og hva er deres vurdering av de ulike standardforbedringene (reisetid, bytte, frekvens, pris etc.)?
3. I hvilken grad har tiltakene trukket bilister over til kollektivtransport, og hva er en ev. netto miljøgevinst av tiltakspakkene?
4. Hva er den samfunnsøkonomiske gevinsten av satsingen?
5. I hvilken grad vil ulike barrierer (fysiske, psykologiske eller informative) begrense effekten av tiltakene?
6. I hvilken grad har rammebetingelsene for reisen (både egenskapene ved alternative transportmidler og mulighetene for å benytte disse som et alternativ til kollektivtransport) påvirket effekten av tiltakene?
7. I hvilken grad er det synergieffekter som gjør at en samlet pakke av tiltak vil ha større effekt enn summen av hvert enkelt tiltak isolert?
8. I hvilken grad har planlegging, organisering og utvikling av tiltakspakkene påvirket resultatene og den planløsning som er valgt?

2.2.2 Teoretisk utgangspunkt

Det er (i TØI-rapport 388/1988 og 794/2005) beskrevet et teoretisk grunnlag som ligger til grunn for design, metodevalg, valg av variabler med mer. Dette framgår av figuren på neste side.



Figur 2.2: Panelundersøkelser kartlegger endringer i reisevaner, rammebetingelser, holdninger og kunnskap (jfr kap 3), brukerundersøkelser kartlegger kollektivtrafikanteres vurderinger av tiltakene og alternative transportmåter (jfr kap 4). En dybde- markedsundersøkelse er en av flere mulige tilleggundersøkelser (jfr kap 5).

Figur 4. Teoretisk modell for evalueringen av tiltakspakkene

Kilde: TØI-rapport 736/2004

Denne figuren representerer en teoretisk modell der forfatterne spesifiserer hvilke knipper av variabler forfatterne antar at faktisk reisemiddelvalg er avhengig av (736/2004:24). Knippene er:

- Trafikantenes preferanser når det gjelder reisetid, kostnader med mer
- Trafikantenes holdninger til ulike transportmidler
- Trafikantenes oppfatninger om det konkrete tilbudet
- Kjennetegn ved de ulike transportalternativene
- Rammebetingelser for den konkrete reisen, som tilgang til bil, parkering med mer

De *sentrale begrepene* som konstituerer de variablene som blir benyttet i dette designet er dermed: Holdninger, oppfatninger/kunnskap, preferanser, personlige kjennetegn, egenskaper ved tilgjengelige alternativer, strategi for valg, situasjon/rammebetingelser, faktisk atferd.

Evalueringdesignet hviler dermed på den *forutsetningen* at dette er adekvate begreper for å forklare reisemiddelvalg, mens de *antagelsene* som gjøres er at det er bestemte kausal- sammenhenger mellom disse knippene av variabler, som illustreres gjennom piler i modellen. Modellen formulerer ikke klare hypoteser om hvilke variable som har sterkest effekt. Evalueringdesignet er rettet mot hvordan holdninger og oppfatninger av kollektivtilbudet endrer, og dermed deres preferanser og strategier for valg. Strukturelle variabler som følger av bystruktur og bosettingsmønster, inngår i kategorien rammebetingelser.

Disse variablene og måten de er begrepsfestet på, kan ikke sies å ha en bestemt *disiplinforankring*: De er variabler på et relativt lavt abstraksjonsnivå, som kan knyttes til flere ulike mer omfattende teoretiske disipliner, som sosiologi og sosialpsykologi. De er interdisiplinært forankret.

Gyldighetsområdet er ikke videre spesifisert. Hovedinntrykket er at det er en evaluering som skal gi kunnskap om de konkrete effektene av ulike tiltak, og dermed kunne generaliseres i det minste til byområder i Norge. Det kan altså tolkes å være hva som påvirker reisemiddelvalg i norske byområder.

2.2.3 Metoder, datagrunnlag og devalueringer

Evalueringsdesignet som skisseres (Renolen 1998) innebærer å ta i bruk en panelundersøkelse og en brukerundersøkelse. Panelundersøkelsen skal være en før- og etterundersøkelse (spørreundersøkelse) blant et tilfeldig utvalg av befolkningen, og den skal i hovedsak gi kunnskap om befolkningens reisevaner, i hvilken grad de har mulighet til å reise på ulike måter og hva som virker inn på deres transportmiddelvalg. I tillegg er det lagt inn spørsmål om holdninger, oppfatninger og kunnskap. Samme panel deltar i før- og etterundersøkelsen.

Brukerundersøkelsen er også før- og etterundersøkelse (spørreundersøkelse), den gjennomføres blant kollektivpassasjerer, og skal gi svar på hvordan trafikantene vurderer tiltakene som er gjennomført, hvem brukerne er, hvilke alternative reisemuligheter de har og hvordan tiltakene virker enkeltvis og sammen. Figuren foran illustrerer hvordan de to undersøkelsene knyttes til teorien som er beskrevet tidligere (se figur 5).

I både brukerundersøkelsen (blant busspassasjerer) og i panelundersøkelsen (blant "alle") registrerte man sone- og nettverksdata som kan brukes til å stedfeste eller rutefeste innsamlede data, og på den måten kunne sammenligne en bussrute (der det er gjort forbedringer) med en annen (hvor det ikke er gjort forbedringer), eller områder med og uten forbedringer. Vi kan ikke se at disse mulighetene er brukt analytisk i det videre arbeidet.

Undersøkelsene som brukes er hovedsakelig kvantitative. (Det er også gjort en prosessevaluering som ikke tas opp i vår rapport.) Det gjennomføres to forskjellige spørreundersøkelser, samt registrering av passasjerutvikling (metoder som er brukt er billettsalgsstatistikk og tellinger). I tillegg gjøres det kvalitative undersøkelser/registreringer/beskrivelser av endringer i kollektivtilbudet.

I TØI-rapport 388/1998 er det beskrevet hvordan man kan gjennomføre brukerundersøkelsene og panelundersøkelsene, og hvordan man skal analysere dataene (dette skal gjøres lokalt), i tillegg er det beskrevet hvordan man kan koble data. Det er laget skjema til bruk i hver av de to datainnsamlingene.

Forenklet er endringer i kollektivtilbudet (tiltaksprogrammene) uavhengig variabel, mens endringer i reisemiddelvalg eller i antall kollektivreisende er avhengig variabel. Hovedtiltakene var endringer i informasjon og markedsføring, infrastrukturtiltak i form av holdeplasser, terminaler og knutepunkter og fremkommelighet, og noen steder også økt frekvens. (Se vedlegg 4 for oversikt og tiltakene i den samlede evalueringen av tiltaksprogrammene.)

Resultatene av tiltakene, slik de framkommer i oppsummering av lokale resultater (TØI-rapport 735/2004) , er presentert i tabellen nedenfor. De viser at mens tiltakspakkene har hatt ingen eller uavklart effekt på transportmiddelfordeling, finner en passasjervekst med kollektiv i noen byområder.

Tabell 2. Resultater av tiltakspakkene perioden (avhengig variabel)

Byområde	Passasjervekst (z)	Reiseomfang	Transportmiddel- fordeling	Holdning til tilbudet (blant passasjerer og/eller befolkningen i regionen)
	-: Færre passasjerer 0: Uendret +: Positiv (z ≤ 5%) ++: Positiv (5% < z ≤ 10%) +++: Positiv (z > 10%) x: Ikke data	-: Mindre 0: Uendret +: Større x: Ikke data	-: Økt bilandel 0: Uendret +: Økt kollektivandel x: Ikke data	-: Negativ 0: Uendret +: Positiv x: Ikke data
Hundvågpakken i Stavanger	+++	0	0	+
Tønsberg-området	+	-	0	-
Drammens-regionen	+	0	+	0
Nedre Glomma	x	+	0	+
Larviksområdet	x	-	0	-
Ålesunds-pakken	+	0	0	?
Grenland	+++	x	X	+
Lillehammer og Gjøvik	++	x	X	0
"Buss 2000" i Tromsø	+	x	X	+
Trondheim	++	+	-	+
Kristiansand	++	0	0	+

Kilde: TØI-rapport 735/2004

2.4 Aggregert analyse av isolerte tiltakseffekter

Som teksten og tabellene ovenfor viser, er det gjennomført og sammenfattet en rekke devalueringer som tar for seg effektene av hver enkelt pakke. Samtidig har forfatterne søkt å få mer systematisk kunnskap utover den som går fram av oppsummeringer av hver enkelt pakke, og slik gjennomføre en samlet evaluering. Målet med den samlede evalueringen er å finne de isolerte effektene av tiltakspakkene, det vil si de effektene som skyldes tiltakene.

I 794/2005 er den analytiske strukturen for en slik aggregert analyse beskrevet. Det er valgt en hovedstruktur i tre steg.

1. Analyse av sammenhengen mellom endringer i kvaliteten på tilbudet og endringer i tilfredshet med tilbudet, inkludert andre faktorer som kan påvirke denne vurderingen.
2. Deretter analyseres om det er noen sammenhenger mellom graden av endring i tilfredshet og endring i bruk av buss i de ulike områdene, korrigert for andre forhold som kan påvirke bruk.
3. Til slutt brukes resultatene fra 1 og 2 til å beregne den isolerte etterspørseffekten av de ulike tiltakene.

De metodiske utfordringene med et slikt evalueringsopplegg diskuteres i rapport 736/2004 (side 23-31). Særlig blir usikkerhet ved passasjerstatistikken framhevet (se også rapport 738/2004). Det er en svakhet ved drøftingen at den er generell, og ikke søker å avklare nærmere i hvilken grad måten de metodiske utfordringene er løst, påvirker de substansielle funnene, dvs. om noen av disse kan ha metodologiske forklaringer.

I en diskusjon om interne sammenhenger/eksogene variable påpekes det at forklaringsfaktorene betraktes som uavhengige variable i analysene, selv om flere av faktorene med rimelighet også kan antas å være avhengige med et annet analytisk perspektiv. Det er gjennomført enkle krysskjøringer av par av variable. Det er også gjennomført regresjonsanalyser. Det er en viss diskusjon om hvordan man kan forklare endringer i kollektivbruken som analysene ikke forklarer, der man fokuserer på ikke lineære sammenhenger og utenforliggende faktorer, men evaluererne går ikke inn på andre forståelser av årsak-virkningsforhold enn de som er beskrevet og brukt i analysen.

2.4.1 Noen substansielle hovedfunn

Kjørstad og Norheim (2005) mener at tiltakspakkene har gitt trafikantene et *bedre tilbud*. I gjennomsnitt har det gitt trafikantene forbedret tilbud på 6 prosent, målt i vektet reisetid. I Grenland har kollektivtrafikantene fått 24 prosent forbedring i tilbudet, mens det i Trondheim og Hundvåg har vært en forbedring på 14 prosent og Kristiansand på 9 prosent. For de andre områdene har det vært mindre endringer, med unntak av Tønsberg som fikk 7 prosent dårligere tilbud.

Forfatterne mener også at tiltakene har gitt *passasjervekst*, men dette er begrenset til byområder hvor de mest omfattende og målrettede tiltakene er satt i gang. Tiltakspakkene har primært hatt suksess hvor en omfordeling og prioritering av tilbudet mot større trafikantgrupper. Korrigert for takstendringen ville passasjerutviklingen ha vært positiv i alle observasjonene. Kjørstad og Norheim (2005) mener at takstøkningen har dempet passasjerveksten.

De argumenterer også for at det største potensialet for økning i kollektivtransporten er blant "av og til"-brukerne. Det er de som i størst grad har økt sin bruk av buss. Tiltakene har med andre ord gitt de som allerede reiser kollektivt et bedre tilbud. Samtidig svarer halvparten av kollektivtrafikantene som har førerkort og bil i husstanden at de kunne ha brukt bil på den aktuelle reisen.

Passasjerer som bruker buss til jobb reiser så ofte at de ikke kan øke antall bussreiser med mindre de begynner å benytte buss også til andre reisemål. ”Dette er en utfordring for kollektivtransporten fordi dette vil kreve andre typer tilbud enn sentrumsrettede tilbud i rushtiden (Kjørstad og Norheim 2005).”

Et annet funn er at trafikantenes oppfatninger av om tilbudet er blitt bedre/dårligere påvirker antall reiser med kollektivt transportmiddel. Forfatterne finner ”en overraskende klar sammenheng mellom netto tilfredshet og reiseaktivitet” og oppgir følgende tommelfingerregel: 10 prosentpoeng i netto økt tilfredshet vil kunne gi 3,7 prosentpoeng i netto økt reiseaktivitet. De mener også at det er lettere å miste passasjerer enn å skaffe nye.

Den viktigste faktoren som forklarer trafikantenes endrede reisehyppighet er frekvens, i følge forfatterne. I tillegg finner de en asymmetri i at et dårligere tilbud gir større bortfall i reisende enn forbedringer gir passasjervekst. ”Det betyr at det er lettere å miste trafikanter enn å skaffe nye (Kjørstad og Norheim 2005).” Grunnen er passasjerenes muligheter til å endre reisefrekvensen, ikke at de har ulik vurdering av forbedringer og forverringer. ”Det er større barrierer mot økt bruk enn mot å redusere bruken av kollektivtransport (Kjørstad og Norheim 2005).”

De antar at det er synergieffekter av å øke satsingen på tilbudet sammen med kunnskapsformidling. Gevinstene ved kombinerte tiltakspakker avhenger av nivået på tiltakene og trafikkgrunnet i utgangspunktet.

2.4.2 Samfunnsøkonomiske analyser

Fearnley og Nossum (2004) gjennomfører også en nyttekostnadsanalyse av tiltakene. Den samfunnsøkonomiske nytten deles grovt inn i tre hovedkategorier, brukernytte, ikke-brukernytte og produsentnytte. Alle disse kan være negative eller positive.

Brukernytten representerer verdien av alle forbedringer som nye og eksisterende passasjerer nyter godt av, som tids- og kostnadsbesparelser. Til grunn for denne verdsettingen av ulike kvalitetsendringer ligger ”stated preference”-analyser (SP-analyser), hvis formål er å finne fram til hvilke standardfaktorer/-forbedringer ved en reise kollektivtrafikantene foretrekker og hvordan de verdsetter disse forbedringene. Dette danner grunnlaget for å beregne trafikantenes nytte av de forskjellige kollektivtiltakene ved å multiplisere antall kollektivtrafikanter som berøres av tiltaket (her basert på passasjertellinger), med faktisk endring i tilbudet (her med sonedata som kilde) som multipliseres med verdsetting av disse endringene basert på SP-analyser.

Trafikantenes nytte omtales gjerne som reduserte generaliserte reisekostnader. Ikke-brukernytte (eksterne effekter) omhandler slikt som effekter for andre trafikanter, miljø og tilgrensede boligområder. Produsentnytte er nytten for selskaper. Forfatterne understreker at begge to ofte er vanskelig å beregne.

Rapporten går så gjennom beregninger av passasjertall, overført trafikk og brukernytte i form av redusert ventetid osv. Vi skal ikke gå gjennom denne i detalj, men kommer tilbake til nyttekostnadsanalyse som evalueringsmetode i neste kapittel. Vi vil kort understreke to forhold: Metodisk gjør også Fearnley og

Nossum klare reservasjoner når det gjelder data fra passasjertellingene. Den samlede analysen viser at tiltakspakkene i sum har vært nyttige, samfunnsøkonomisk sett, med en Nyttekostnadsbrøk som estimeres å være mellom 1,05 og 1,17 samlet. Samtidig er det variasjon mellom tiltaksområdene.

2.5 Imøtekommer evalueringen validitetskriteriene?

I det følgende skal vi gjennomgå evalueringen med bakgrunn i kriteriesettet vi etablerte i innledningskapitlet. Det er gjennomført mange devalueringer, men vi retter oppmerksomheten mot den aggregerte analysen. Imøtekommer evalueringensdesignet til tiltakspakkene validitetskriteriene slik de er formulert av Elvik?

Teoretisk validitet

Det er formulert et eksplisitt teorigrunnlag for evaluering, som inneholder hypoteser som skal testes. Disse teoretiske begrepene kan lett operasjonaliseres, men det er noen klare svakheter ved operasjonaliseringen (se nedenfor). Den teoretiske modellen er slik formulert at det i prinsippet bør være mulig å inkludere de ulike tiltakene som blir forsøkt ut, som variabler i modellen. Modellen spesifiserer ikke klare hypoteser om hvilke variabler som har størst effekt, noe som også innebærer at det er vanskelig å teste om teorien finner støtte i resultatene av undersøkelsen eller ikke. Rett nok foreligger det tester av hvor mye enkelte modelleringer forklarer av variasjon, for eksempel viser en simultan modell for endret reisefrekvens og vurdering av om tilbudet er bedre eller dårlig en forklart varians på mellom 20 og 50 (736/2004:79).

Statistisk validitet omhandler som nevnt innledningsvis graden av tallmessig nøyaktighet, feilfrihet og representativitet i resultatene av en undersøkelse eller et sett av undersøkelser.

Det synes særlig å være fem problemstillinger knyttet til utvalg og reliabilitet.

For det første er det, som beskrevet ovenfor, særlig usikkerhet knyttet til en av indikatorene på den avhengige variabelen: Passasjerstatistikken. Dette medførte at en i mange analyser måtte ta i bruk andre, mindre valide mål, som opplevd reisehyppighet. Man undersøker derfor sammenhenger mellom *opplevd* endring i tilbud og *opplevd* endring i reisevaner, heller enn sammenhenger mellom faktiske endringer i tilbud og faktiske endringer i reisevaner.

For det andre er ikke tiltakene bare valgt ut basert på rent evalueringsmessig interesse. De er også et resultat av politisk-administrative prosesser som kravet om lokal medfinansiering. Medfinansiering utgjør et viktig utvalgs-kriterium. Det er en problemstilling for mange evalueringer, at utvalget av tiltak ikke er rent faglig begrunnet, slik at en eventuell ambisjon om å etablere en generell kunnskapsstatus, skjer på tross av at en ikke har hatt et systematisk opplegg som har vært basert på kun faglige, evalueringsmessige hensyn.

For det tredje kan medfinansiering også påvirke reliabiliteten: I tiltakspakkene var undersøkelsesopplegget lagt opp slik at det skulle gjøres lokal datainnsamling og analyse av de enkelte tiltakspakkene, mens TØI skulle gjøre analysene av

undersøkelsene/dataene. I TØI-rapport 810/2005 er problemer med at datainnsamlingen ble gjort lokalt beskrevet. Problemene kan hovedsakelig oppsummeres med at en del spørsmål ikke ble stilt og at en del ble stilt på en ukorrekt måte – man fikk altså datafrafall og unøyaktigheter i innsamlede data.

For det fjerde er tiltakene primært utført på mindre områder/strekninger, mens datainnsamling gjennomføres for større del av området. RVU gjennomføres de fleste steder for hele byområdet (i Drammen omfatter den for eksempel fire kommuner), mens ett sted (Hundvåg) gjennomføres det kun i det området hvor det er gjennomført tiltak. Hovedtyngden av endringer som kan bidra til å endre reisefrekvens eller transportmiddelfordeling gjøres på enkeltruter/områder. RVU synes å være gjennomført slik at det skal være mulig å skille ut områder hvor tiltakspakkene burde ha stor effekt og områder (innenfor byområdet) hvor man ikke kunne forvente slik effekt, men den avsluttende analysen framstår imidlertid i liten grad å være på rutenivå, men mer på områdenivå. Problemet er at rapportene, slik de nå foreligger, i liten grad er etterprøvbare for slike spørsmål.

For det femte er det særegne problemer knyttet til den aggregerte data-analyse der en sammenfatter alle undersøkelsene. For å få fram mer systematisk og generell kunnskap, er dette et nødvendig grep. Det oppstår imidlertid bestemte problemer knyttet til forholdet mellom lokal kontekst, reliabiliteten ved dataene og aggregerte analyser. problemet her er dels valg av variable, der tilfredshet og bruk, de to variablene der årsaksforholdet er minst avklart, blir sentralt i analysen. Videre får vi ikke noe klart bilde av i hvilken grad den aggregerte analysen baserer seg på ulike tiltak som er gjennomført i noen områder, som kan få relativt stor vekt i den samlede analysen. Igjen er altså etterprøvbareheten for en alminnelig interessert og orientert leser et problem.

Intern validitet gjelder spørsmålet om i hvilken grad evalueringene gir grunnlag for å hevde at det er en årsakssammenheng mellom det undersøkte tiltaket og de endringene som kan påvises i den eller de avhengige variablene. Det viktigste å understreke her, er at selve det metodiske opplegget med før- og etterundersøkelse, og innsamling av data for å undersøke om eller i hvilken grad det hadde foregått endringer, er det avgjørende grepet for å foreta en slik effektevaluering. Dermed har en også, i utgangspunktet et design for å analysere det første og avgjørende spørsmålet ved denne type evaluering: Om det er en statistisk sammenheng mellom årsaksvariabelen og virkningsvariabelen. Men forutsetningen for høy intern validitet er de metodiske utfordringer vi beskrev under statistisk validitet er tilfredsstillende løst. Og som vi har pekt på ovenfor, er det klare svakheter i det konkrete designet på dette området. Ikke minst er det et problem at indikatorene oppfatninger om endrede reisevaner, neppe er en god indikator for den teoretiske variabelen reiseatferd.

Dette er også et design som i utgangspunktet imøtekommer kriteriet om å kunne fastsette årsaksretningen relativt entydig, med et særlig unntak for det "klassiske" problemet når det gjelder forholdet mellom holdning og handling. Konkret gjelder dette blant annet sammenhenger mellom kunnskap om busstilbudet og frekvens i bruk av dette, der det ikke er gitt at årsaksretningen er at kunnskap påvirker handling, men like gjerne kan være vice versa.

Ekstern validitet betegner muligheten for å generalisere resultatene av en undersøkelse utover den spesifikke konteksten den er utført i over tid og på tvers av trekk ved konteksten undersøkelsene er utført i. Formålet med den aggregerte analysen er å få fram kunnskap utover den substansielle kunnskapen om effekten av et gitt tiltak i et gitt byområde, og å kunne si noe mer generelt av effekten av bestemte tiltak og kombinasjoner av tiltak. Vi tolker ambisjonene om ekstern validitet her til å omfatte kollektivtransport i norske byområdet, også at denne kan diskuteres med bakgrunn i funn fra undersøkelser i andre land, eller fra tidligere norske undersøkelser og mer teoretisk arbeider, men dette blir i liten grad gjort her.

2.6 Oppsummering

Det evalueringsdesignet som opprinnelig ble utformet var på mange områder i tråd med de kriterier vi har beskrevet innledningsvis: En søkte å gjennomføre et relativt omfattende analyseopplegg basert på data fra før- og etter innføring av tiltakene, å få mer allmenn systematisk kunnskap utover effekten av noen tiltak i et gitt byområde, dels gjennom sammenfattende beskrivelser, dels gjennom aggregerte analyser. Dette grepet gjør det også mulig å avklare eventuelle inkonsistenser mellom funn på ulike analysenivå. Endelig avslutter en med en samfunnsøkonomisk analyse av tiltakspakkene.

Evalueringen av tiltakspakkene illustrerer imidlertid en bestemt problemstilling som ikke vil være uvanlig ved evalueringer: Utvalget av observasjonene baserer seg ikke bare på rent faglige, teoretiske begrunnelser. Selve designet innebærer forsøk på å utvikle mer systematisk kunnskap trass i denne typen begrensninger i utvalgsprosedyrer. De metodiske utfordringene er knyttet til denne spenningen. På den ene siden politisk-administrativt vedtatte men faglig vurderte tiltak, gjerne helt nede på nivå med leskur, og lokalt innsamlede data. På den andre siden er et analytisk design som tar sikte på å framskaffe mer systematisk kunnskap om hvilke tiltak som påvirker reisevaner, basert på aggregerte analyser av de ulike tiltakene og deres effekter.

Evalueringsdesignet lar seg i all hovedsak analysere med bakgrunn i de kvalitetskriteriene vi har skissert innledningsvis.

Statistisk validitet: Rapportene drøfter reliabilitetsproblemet med noen typer data, mest klart ved passasjertellinger. Men når passasjertall er lite reliable, og en i stedet etablerer opplevd reisehyppighet som indikator for avhengig variabel, er dette et mindre valid mål, noe som får konsekvenser for interne og teoretisk validitet. Evalueringen er generelt preget av et relativt omfattende datamateriale, som i utgangspunktet åpner for kontroll av reliabilitet ved ulike målinger. Minst klart formulert er betingelser knyttet til den aggregerte analysen, og koblinger mellom devalueringer og mer samlede evalueringer.

Intern validitet: Vurderingen av denne evalueringen lar seg i all hovedsak gjennomføre med bakgrunn i de kriterier vi presenterte i kapittel 1 om intern validitet. Problemet er imidlertid at når en opererer med opplevd endret tilbud og opplevd endret reisehyppighet som indikatorer, så svekker dette grunnlaget for å anta at en snakker om en årsakssammenheng og ikke bare en korrelasjon. Videre

kan årsakssammenhengene mellom kunnskap om kollektivtilbudet og bruk av det like gjerne være slik at bruk skaper kunnskap som omvendt.

Slike svakheter ved intern validitet svekker den teoretiske validiteten. I utgangspunktet var den teoretiske modellen som blir benyttet er preget av knipper av variabler på operasjonaliserbart nivå, som ikke er direkte knyttet til spesielle disipliner (som sosiologi, psykologi eller økonomisk teori). Sammenhengen mellom variablene (holdninger, alternativer, preferanser, rammebetingelser) er beskrevet som årsakssammenhenger, og samtidig er variabellisten relativt uttømmende, men noe underspesifisert. Dette innebærer også at en må svare "både og" på spørsmålet om teorien som er formulert kan forklare hvordan det undersøkte tiltaket kan virke på det problem det er ment å løse. En undersøker og søker å sammenfatte noen isolerte effekter av variable innen modellen og beskriver forklart varians av disse variablene, men denne analysen er svekker av de innvendinger vi har reist ovenfor. Uansett ville ikke den generelle, overordnede modellen som introduseres forut for designet, testet.

Ekstern validitet: Dette lar seg i prinsippet gjennomføre med bakgrunn i dette designet. Det er imidlertid ikke gjort systematiske sammenligninger med andre analyser av andre byområder i disse rapportene

Denne evalueringen har følgende generelle trekk: Den er basert på kvantitative data, som er samlet inn både før og etter tiltaket er gjennomført, noe som muliggjør måling av endring. Variablene dekker en rekke dimensjoner, fra kontekstuelle variabler (rammebetingelser), til klassifisering av tiltak som skal påvirke trafikantene, deres personlige kjennetegn og reisevaner. Evalueringen er i liten grad teoretisk informert i betydning klar spesifisering av aktørforutsetninger og andre forhold som vi normalt finner innen mer disiplinforankret forskning, men har i større grad karakter av å kartlegge empirisk sammenhenger mellom variabler. Den er også i liten grad fokusert på ekstern validitet, men til forskjell fra evalueringer vi skal gå inn på senere i denne rapporten, vil det her være mulig å undersøke den eksterne validiteten. Vår antagelser her er at disse generelle trekkene deler evalueringen med en rekke andre evalueringer, slik at vi skal her benevne denne tradisjonen; multidimensjonal variabelanalyse. Den representerer i første rekke en metodisk tradisjon som utøves av kvantitativt skolerte forskere fra ulike disipliner (som psykologi, sosiologi og økonomi), men liten eksplisitt teoretisk forankring. I noen tilfeller vil den også ha ambisjoner om å inkludere "ikke-sosiale" variabler, som variabler knyttet til teknologi, miljøeffekter (støy, forurensing osv.). Se for eksempel Klæboe (2006) sin avhandling for en oversikt over artikler som representerer denne tradisjonen, men som her har gjennomgått den kritikk som følger av publisering i tidsskrift og forsvar gjennom avhandling – ulikt den evalueringen vi har gjennomgått her.

3. Store samferdselsinvesteringer. To evalueringstradisjoner – to hovedutfordringer

Dette kapitlet har et dobbelt formål: For det første skal det introdusere et bestemt tema: evaluering av store samferdselsinvesteringer. Begrunnelsen for at det er særlig relevant er at slike investeringer er vesentlige i økonomisk betydning, og at det er risiko og usikkerhet knyttet til slike investeringer, både i form av kostnader og forventede virkninger. I kapitlets første del, presenteres noen hovedfunn når det gjelder omfang av feilbudsjettering. Disse funnene belyser en etablert problemstilling innen sektoren; at mange infrastrukturinvesteringer i ettetid viser seg å være underbudsjettet.

I andre del skal vi presentere og diskutere to hovedtilnærminger til evaluering av slike feilbudsjetteringer. Disse to tilnærmingene tar utgangspunkt i at vi her grovt kan skille mellom to typer hovedforklaringer på slik feilbudsjettering, nemlig *den metodiske forklaringen* og *den organisatoriske forklaringen*. Den metodiske tar utgangspunkt i at det er feil knyttet til modellverktøyet, mens den organisatoriske tar utgangspunkt i at det er trekk ved måten selve den faglig-administrative utredningsprosessen er organisert på, som påvirker det faglige resultatet. Slike forklaringer gir opphav til ulike evalueringstilnærminger. Vi konsentrerer oss her om å presentere og diskutere to av dem: Ex post nyttekostnadsanalyse (etteranalyser) og komparative analyser av mega-prosjekter. Disse er til en viss grad komplementære, men har ulike kritiske punkt når det gjelder de kvalitetskriteriene vi kan stille til evalueringer, det vil si teoretisk, statistisk, intern og eksternt validitet.

3.1. Kostnadsoverskridelser og feilbudsjettering og trafikkprognoser – noen funn og observasjoner

Kostnadsoverskridelser og feilbudsjettering har vært et sentralt tema innen samferdselssektoren, både nasjonalt og internasjonalt, i en årrekke. Her skal vi kort sammenfatte noen hovedfunn og problemstillinger.

Sentralt i *den internasjonale debatten* på dette området, har vært det danske fagmiljøet ved Aalborg Universitet. ¹ Første fase i programmet ”Planlægning, implementering og evaluering af store transportinvesteringer” startet opp ved Aalborg Universitet i 1994 og ble avsluttet i 2000. Arbeidet i annen fase begynte sommeren 2001, der intensjonen var å gå i dybden på bl a gjennomførbarhet, risiko og årsaksforklaringer. Fokuset er på lønnsomhetsvurderinger i vid forstand

¹ Denne gjennomgangen baserer seg på Strand (2006), som har gjennomgått sentrale bidrag innen dette prosjektet (se særlig Skamris Holm 2000 og Flyvbjerg et al 2003).

– perspektivet utvides til å omfatte også sikkerhets-, miljømessige og regionale virkninger – og på bakenforliggende årsaker til budsjettoverskridelser og lite treffsikre trafikkprognoser.

I første fase ble det gjennomført før-etterundersøkelser av i alt 36 prosjekter lokalisert i Danmark, Sverige, Frankrike, Tyskland og England, og nær halvparten i Danmark. Utover dette er det systematisert informasjon om økonomi og trafikkutvikling for ca 250 prosjekter i 20 land om oppstart- og avslutningstidspunkt, eierskap, budsjetterte og faktiske anleggskostnader, og prognostisert og faktisk trafikk.

Etterprøvingen omfatter store vegprosjekter (herunder både bruer og tunneler) og delstrekninger, og høyhastighetstog, bybaner og jernbaneprosjekter for øvrig. De eldste prosjektene stammer fra 1927, og er tatt med for å analysere om prognosene er blitt mer presise med tiden.

Spennet i anleggskostnader for de undersøkte prosjektene er fra 350 millioner kroner til 6 milliarder. Dette illustrerer hva som i denne sammenheng karakteriseres som store og små investeringer. Når spennet er så stort, gir det også grunnlag for å undersøke hvorvidt avvik mellom faktiske og planlagte kostnader varierer med investeringens størrelse.

Det overordnede funnet var at anleggsbudsjettene ble overskredet i 90 prosent av prosjektene. Den største gjennomsnittlige overskridelsen på 45 prosent var knyttet til baneprosjekter. "Fastforbindelser" som vegbruere og vegtunneler hadde i gjennomsnitt en budsjettsprikk på 35 prosent og 20 prosent for vegprosjektene. De danske forskerne fant ingen sammenheng mellom prosjektstørrelse og relativ prosjektoverskridelse.

Kostnadsprognosene er ikke blitt mer treffsikre over tid, budsjettoverskridelsene er av samme størrelsesorden som for 10, 20 og 70 år siden. Kostnadsutviklingen varierer med eierskap, det vil si om prosjektet er organisert som offentlig aksjeselskap, rent offentlig eie eller privat foretak: Private prosjekter synes å være vesentlig bedre til å holde budsjettene enn statseide aksjeselskaper, men mindre effektive enn vanlig offentlige.

Den første hovedkonklusjonen når det gjelder trafikkprognoser, var at det i gjennomsnitt for alle vegprosjektene som ble undersøkt, viser seg å ha blitt mer trafikk enn forutsatt. Det mest karakteriserende stikkordet er *variasjon*, slik at gjennomsnittsavviket på +10 prosent var resultatet av ekstremalverdier på -74 og +178 prosent. Med noen få unntak viser trafikkprognosens treffsikkerhet seg uforandret over tid. Heller ikke lengden på implementeringsfasen eller eierforhold har noen signifikant påvirkningskraft på prognosenøyaktighet.

I den *nasjonale debatten* har Riksrevisjonens undersøkelse av store vegbudsjettoverskridelser og Stortingets behandling av denne i 2003, hatt en sentral plass. Formålet med denne forvaltningsrevisjonen var å undersøke i hvilken grad Statens vegvesens systemer for planlegging, oppfølging og finansiering av vegprosjekter bidrar til å holde tids- og kostnadsplaner. De konkrete prosjektene som ble gjennomgått var Trekantsambandet, Rv. 159 Knatten-Vigernes, Ev16 Aurland-Lærdal, Ev 134 Drammen-Mjøndalen og Rv. 555 Nygårdstangen-Gyldenpris, og til dels hele 'Vegpakke Drammen', de fleste ferdigstilt i perioden 2000-02.

Utvalgskriteriet var økonomisk størrelse og prosentvis overskridelse i forhold til opprinnelig kostnadsoverslag (kostnadssprekk på over 20 prosent). Riksrevisjonen skriver at ”de fem prosjektene som er undersøkt ikke nødvendigvis er representative for den samlede prosjektporteføljen i Statens vegvesen. Det kan derfor ikke uten videre trekkes generelle konklusjoner om alle prosjektene i Statens vegvesen”.

Riksrevisjonens vurdering av årsaker oppsummeres slik:

”Når det gjelder årsakene til kostnadsoverskridelsene kan dette i stor grad tilbakeføres til mangelfull planlegging og mangelfulle kostnadsoverslag. I tillegg har vesentlige omprosjekteringer og manglende systemer for prosjektoppfølgning bidratt til at det har vært vanskelig å styre prosjektene.”

I Stortingets høring om Riksrevisjonens rapport var fokuset nesten ensidig på kostnadssiden, og i løpet av høringen kom det fram flere mulige forklaringer på overskridelsene. Disse som vi skal komme tilbake til senere i kapitlet. Her avslutter vi med å poengtere at behovet for videre arbeid med problemstillingene blir tydelige når Innstilling 193 til Stortinget eksplisitt og med referanse til forprosjektene 2002-04 ”slutter seg til TØIs anbefalinger om en integrert kunnskapsbase [...] slik at en for seinere tid vinner erfaringer gjennom de prosjekter som er gjennomført.”

Når det gjelder sammenfattende analyser av norske vegprosjekter, har Odeck (2004) og Odeck og Brudeseth (2003) gjennomført to makroanalyser. Odeck (2004) gjennomførte en økonometrisk analyse av kostnadsoverskridelser for 620 norske vegprosjekter ferdigstilt i perioden 1992-95. Hensikten var å belyse i hvilken grad det forekom underbudsjettering. Konsekvensene av underbudsjettering vil kunne være ineffektiv ressursallokering, fordi samfunnsøkonomisk ulønnsomme prosjekter kan bli realiserte.

Den økonomiske størrelsen av prosjektene i undersøkelsen varierer fra små (på under 15 millioner kroner) til store (på over 350 millioner). Modellen kontrollerer statistisk for faktorer som region, prosjektperiode, prosjekttype (veg, tunnel, bru) og kostnad. Også i denne undersøkelsen er *variasjon* et hovedstikkord. Gjennomsnitts-overskridelsen er åtte prosent, men som et resultat av en spredning fra -59 prosent til +183 prosent. Mens Odeck og Brudeseth i tidligere undersøkelser (2003) ikke fant noen sammenheng mellom prosjektstørrelse og overskridelse i sitt (nyere) materiale, fremhever Odeck i 2004-undersøkelsen at overskridelsene både absolutt og relativt synes å være dominante i mindre prosjekter. Han mener at en mulig årsak til dette er at større prosjekter har en mer kompetent prosjektledelse.

Odeck finner at modellen forklarer 20 prosent av variasjonene i kostnadsoverskridelsene. Han refererer til tidligere studier, som det for øvrig er få av, og som alle konkluderer med en hypotese om at endringer i design, forsinkelser og teknologi ikke kan forklare alle overskridelsene. Han referer også til Wachs (1990) som allerede i 1990 antydte at manipulerede prognoser var én sannsynlig årsaksforklaring.

Oppsummering av undersøkelser av kostnadsoverskridelser, feilbudsjettering og trafikkprognoser

Både den danske komparative undersøkelsen, flere norske undersøkelser og Riksevisjonen og andre forvaltningsorganers virksomhet understreker problemet med feilbudsjettering og følgelig kostnadsoverskridelser ved samferdselinvesteringer. Den danske komparative undersøkelsen indikerer at overskridelsene generelt er størst for jernbane, deretter vegtunneler og vegbroer og til sist vegutbygging. Når det gjelder trafikkprognosene, er disse også preget av feilvurderinger, men her er det ikke entydig slik at trafikkveksten overvurderes, men like ofte at de undervurderes, særlig for veginvesteringer.

Slike feilvurderinger kan ha ulike forklaringer. Vi skal grovt skille mellom to hovedtyper forklaringer: *Den metodiske* og *den organisatoriske*. Den metodiske forklaringen tar utgangspunkt i at det finnes feil og svakheter ved datagrunnlag og metode. Vi skal her først gå nærmere inn på en type evaluering som både søker å lokalisere disse svakhetene og redusere dem, nemlig *ex post* nyttekostnadsanalyse (også kalt etteranalyser og/eller etterprøving).

3.2 Ex post nyttekostnadsanalyse av store samferdselsinvesteringer

Tradisjonen med evaluering av nyttekostnadsanalyser er vesentlig av to grunner. For det første angår den evaluering av omfattende offentlig ressursbruk. For det andre er *ex ante* nyttekostnadsanalyse sentrale innen samferdselssektoren, og vegsektoren er den offentlige sektoren med lengst tradisjon for systematisk bruk av Nyttekostnadsanalyser i Norge (Grøvdal og Hjelle, 1998: 233, boken gir for øvrig en grei lærebokframstilling av nyttekostnadsanalyse). Formålet med denne type etteranalyse (og/eller etterprøving) etter denne malen/metoden, er å forbedre *ex ante* metodikken gjennom å kartlegge avvik mellom de beregnede samfunnsøkonomiske virkningene i for- og etterundersøkelsen. Det er dessuten en hensikt å lokalisere hvilke av disse som skyldes ulik metode eller metodebruk, ytre forutsetninger og/eller prosjektavhengige avvik.

3.2.1 Kort om Nyttekostnadsanalyse

Nyttekostnadsanalyse er teoretisk forankret i velferdsøkonomisk teori. Hensikten er å vurdere hvorvidt et tiltak kan gi et positivt eller negativt samfunnsøkonomisk resultat. Slike analyser utføres ofte for å gi grunnlag for å vurdere om det er nyttig for samfunnet å bruke økonomiske midler til å gjennomføre et tiltak, i tilfeller der det ikke kan observeres direkte økonomiske effekter, altså hvor det finnes positive virkninger som ikke umiddelbart kan oppgis eller registreres med kroneverdi. Nyttekostnadsanalysen er en sammenstilling av virkninger som så langt som mulig tilstrebes å gi en felles benevnelse, en pengeverdi. Tiltak kan i prinsippet vurderes på flere nivåer: Effektene av et enkelttiltak, sammenligning av alternative tiltak/løsninger for dette samme tiltaket, sammenligninger av ulike tiltak i samme sektor (for eksempel et veg- og et jernbaneinvesteringstiltak) og sammenligning av tiltak i ulike sektorer. Utstrekningen og anvendelsen av analyse materialet og resultatet i disse ulike sammenhengene avhenger av

ambisjonene om hvorvidt metodetilnærmingen og verdsettingen antas å kunne gi et tilstrekkelig bilde av tiltakets samlede samfunnsøkonomiske effekt.

Utgangspunktet for slike analyser er begrepene om konsument- og produsentoverskudd. Et prosjekt hvor den samlede neddiskonterte nytten (målt i kroneverdi) er større enn de samlede neddiskonterte kostnadene (også i kroneverdi) kan sies å være samfunnsøkonomisk lønnsomt. Dette kan representere akseptabel ressursbruk for samfunnet.

En nyttekostnadsanalyse skal vise tiltakets effekt for ulike aktører i samfunnet. Her skiller det gjerne mellom noen hovedkategorier av interessenter: Det offentlige, brukerne og samfunnet for øvrig. I samferdselssektoren framkommer ofte effektene for det offentlige i form av investeringskostnader, endringer i drifts- og vedlikeholdskostnader og offentlig kjøp av transporttjenester. Brukerne er både operatører og trafikanter som kan oppleve endret tidsbruk (reisetid, ventetid, gangtid) og endring i direkte økonomiske størrelser som driftskostnader og utlegg (som for eksempel billett-kostnader). Samfunnet for øvrig uttrykker det som også kalles eksterne effekter, som for eksempel endringer i lokale, regionale og globale utslipp, støy og naturinngrep.

Forskjellen mellom de to analysetypene er i hovedsak at en ex ante-metode forutsetter ulike utfall som vil følge av prosjektet, mens i et ex-post-tilfelle har forskeren/utrederen/planleggeren anledning til å vurdere tiltakets samfunnsøkonomiske virkning ut fra observerbare resultater for enkeltfaktorer (som for eksempel trafikkutvikling, kostnader og tilbudsending). I så måte kan ex-post-analysen gi informasjon om metodetilnærmingen synes hensiktsmessig, i hvilken grad datagrunnlaget som var brukt i analysen var tilstrekkelig detaljert eller begrunnet og hvorvidt faktorene/verdiene som er brukt virker hensiktsmessig for å uttrykke samfunnets nytte av et tiltak.

3.2.2 Nærmere om ex post nyttekostnadsanalyser av samferdselsprosjekter

Hva kjennetegner ex post nyttekostnadsanalyser? I hvilken grad tilfredsstillende denne type analyser de kravene vi har stilt til evalueringer? Hvilke særegne metodiske problemstillinger står en overfor i slike analyser?

For å belyse disse spørsmålene tar vi utgangspunkt i en teoretisk gjennomgang og konkrete etterstudier slik disse er foretatt av Eriksen (2003) og Eriksen og Minken (2004 a og b). Den foreliggende teksten er en kortfattet versjon og til dels direkte gjengivelse av disse arbeidene. Disse undersøkelsene representerer bare et knippe bidrag om Nytttekostnads-etteranalyser, men vi legger her til grunn at de kan sees på som representative for særegne trekk og problemstillinger ved nyttekostnadsanalyser sammenlignet med de andre typer evalueringer vi diskuterer i denne rapporten.

Problemstilling og teoretisk utgangspunkt

Formålet med slike etteranalyser er gjerne å forbedre nyttekostnadsverktøyet for etteranalyser – i dette tilfellet å forbedre samferdselsetatenes arbeid med nyttekostnadsanalyser av store samferdselsprosjekter.

De sentrale elementene i en slik analyse er:

- Prosjektkostnader, gjerne operasjonalisert som de kostnader som er underlagt utbyggers budsjett
- Prosjektgevinster. Sentrale variabler i nytten av et prosjekt vil være tidsgevinster i form av forkortet reisetid, bedret trafikkikkerhet eller sparte miljøkostnader. I mange prosjekter er tidsgevinster det største nytteelementet. Metoden for pengeverdering av nyttegevinster som ikke kjøpes eller selges i et marked, tar utgangspunkt i folks betalingsvillighet for å oppnå gevinsten
- Ordinære eksternaliteter. Miljøvirkninger og andre virkninger som påføres omgivelsene og som ikke omsettes i et marked må også tallfestes ved hjelp av betalingsvillighetsstudier. De virkningene som vanligvis regnes med i nytte-/kostnadsanalyser er miljøutslipp, støy, ulykker, kø og trengsel og slitasje på infrastruktur. Disse ulike virkningene er ikke alle like enkle å tallfeste
- Andre tallfestbare virkninger er knyttet til positive nettoringvirkninger som for eksempel *nettverkseffekter*: En forbedring i form av en ny lenke, har ikke bare effekter for de som benytter seg av denne forbedringen, men avlaste også hele nettverket eller gjøre nettverket større. Ved å *redusere kø* kan den totale nyttegevinsten bli større enn gevinsten for dem som benytter den konkrete forbedringen. Dette er en tilleggsgevinst for samfunnet som ikke fanges opp av brukernytten for det enkelte prosjektet

Det er en del forutsetninger knyttet til datagrunnlaget for etteranalyser:

- En trenger detaljerte data og dokumentasjon fra den opprinnelige nytte-/kostnadsanalysen, om kostnadene ved prosjektet, trafikkutviklingen og eksterne kostnader. I tillegg må selve metoden for beregning av konsumentoverskudd være dokumentert
- Samme data må foreligge for virkelig utvikling. Som for den opprinnelige Nyttekostnadsanalysen er det særegne utfordringer ved beregning av eksterne kostnader som miljø- og ulykkeskostnadene. Særlig sistnevnte er et spesielt problem siden ett prosjekt som regel ikke gir mange nok observasjoner til statistisk utsagnskraft om vi står overfor en endring som er signifikant

Kontrafaktisitetsproblemet: Hypotetisk utvikling uten prosjektet vurdert i etterkant

En kontrafaktisitetsanalyse ved denne type evalueringer innebærer å beregne den "hypotetiske" trafikkutviklingen for det tilfellet at prosjektet ikke blir gjennomført. En måte å gjennomføre slike analyser er å ta utgangspunkt i samme beregningsopplegg som i nyttekostnadsanalysen. Ved å sette inn realisert verdi for de parametrene som er benyttet i det opprinnelige nullalternativet, fås en virkelighetskorrigert "prognose på etterskudd" for det ikke realiserede alternativet. Alternativt kan en gjøre en beregning av den hypotetiske utviklingen uten prosjektet uavhengig av den opprinnelige modellen, for eksempel basert på økonometriske analyser av tilgjengelige data for områder som ligner på det analyserte området.

3.2.3 Nærmere om to konkrete jernbaneprosjekter – Sandeparsellene og Nationaltheatret stasjon

I arbeidet med etteranalyse i praksis er det gjennomført flere analyser. Vi skal sammenfatte noen trekk ved slike analyser ved å presentere hovedfunn fra to arbeidsdokumenter som omhandler etteranalyse av henholdsvis Sandeparsellene og Nationaltheatret stasjon.

En kan skille mellom to elementer i etteranalysen. For det første, selve gjennomføringen av nyttekostnadsanalysen, og for det andre, presentasjonen av før- og etterundersøkelsen, i form av tabeller som oppsummerer avvik mellom foranalysen og etteranalysen.

Slike avvik forklares normalt med ulik metode, ytre forutsetninger og prosjektavhengige avvik:

- *Ulik metode* kan bety avvikende, men i og for seg korrekte beregningsmåter, ulik detaljeringsgrad, ulike satser (kalkulasjonspriser) og ikke minst direkte metodefeil
- *Ytre forutsetninger* er generelle forhold som er uavhengige av prosjektet, men som påvirker resultatet; generelle prognoser for trafikkutviklingen i landet eller en region; økonomiske prognoser som påvirker trafikkutviklingen; og ytre forhold som påvirker kostnadsutviklingen for prosjektet
- *Prosjektavhengige avvik* kan være avvikende investeringskostnader eller driftskostnader for prosjektet. Det kan også være avvikende trafikkutvikling som kan tilskrives tiltaket

Formålet med disse undersøkelsene var å teste ut metodeverktøyet, sammenligne jernbaneverkets nytte-/kostnadsverktøy og gjennomgå metodehåndboka.

Oppsummering av funn

Formålet med disse undersøkelsene var å teste ut metodeverktøyet, sammenligne Jernbaneverkets Nyttekostnadsverktøy på to ulike tidspunkter (ved planlegging og ved ferdigstilling av prosjektene) og gjennomgå den nyeste versjonen av metodehåndboka.

En konkluderte med at en sammenlikning av metodehåndboka av 2001 med den fra 1992 viste betydelige forskjeller og høy grad av forbedring. Tross enkelte mindre feil og svakheter var dagens metodehåndbok, JD 205, etter etterprøvernes vurdering et godt grunnlag for utarbeidelse av riktige nyttekostnadsanalyser.

Et hovedfunn i gjennomgangen av Sandeparsellene var at nyttekostnadsanalysen som en legger opp til i etteranalysen, bygger på sviktende forutsetninger og bryter med JD 205. En forutsetter implisitt at nullalternativet er riktig beregnet i føranalysen, og at tiltaksalternativet kan beregnes med hjelp av noen få aggregerte tall. Et like stort problem er manglende dokumentasjon av det *faktisk gjennomførte* nullalternativet. Siden det ikke forelå noe reelt nullalternativ, var det ikke mulig å gjennomføre etteranalyse. De analysene som ble foretatt hadde bare illustrativ verdi. Vi finner likevel grunn til å nevne at den konkrete etteranalysen av avvik arter seg nokså likt for de to parsellene Bergsenga-Åshaugen og Åshaugen-Holm. For begge parseller er det betydelig reduksjon i nåverdien, *henholdsvis* fra Nytttekostnadsbrøk 1,29 til 0,72 og 1,31, til 0,71.

Det negative avviket mellom beregningene i før- og etteranalysen var størst når det gjaldt nytten for offentlig sektor. Operatørens nytte var også betydelig redusert i forhold til foranalysene, mens trafikantnyten var noe redusert. Av partene var det kun tredjepart eller "samfunnet for øvrig" som kommer bedre ut nyttemessig i etteranalysen enn i foranalysen.

Av de prosjektavhengige forklaringsfaktorene bidro underkalkulering av investeringskostnadene og metodefeil til en stor nyttereduksjon for den offentlige sektoren. Investeringene var betydelig underkalkulert i begge prosjektene med overskridelser i størrelsesorden 70 til 85 prosent. I tillegg hadde man glemt å nedskrive restverdien av prosjektene, noe som gir bidrag i samme størrelse.

Tiltakene hadde ikke ført til noen identifiserbar trafikkvekst, noe som gjorde at inntektsvikten for operatøren blir betydelig i forhold til forventningene. Dette bidro også til redusert nytte for trafikantene ved at færre personer enn forventet drar nytte av innkortet reisetid.

Heller ikke for "Nye Nationaltheatret Stasjon", viste det seg mulig å gjennomføre en reell etteranalyse, særlig fordi de opprinnelige nytteberegningene var dårlig dokumentert. Investeringskostnadene var rimelig godt dokumentert, selv om tidsfordelingen av kostnadene kunne vært klarere gjort rede for. Disse svakhetene tilsa at etteranalysen av de samfunnsøkonomiske virkningene av Nye Nationaltheatret Stasjon nødvendigvis ville bli summariske.

Andre bidrag til forbedring av nyttekostnadsanalyse i samferdselssektoren.

Det er ikke bare etteranalyser av denne type en har tatt i bruk for å bedre modellverktøyet. Transportetatene har også, i forbindelse med arbeidet med Nasjonal transportplan, i fellesskap søkt å standardisere modellene og de enkelte etatenes metodeverktøy. I dette arbeidet har en også støttet seg på eksterne vurderinger (Minken og Holmleid, 2005) Hensikten er at samfunnsøkonomiske analyser av ulike infrastrukturprosjekter, i hovedsak veg- og baneprosjekter, skal kunne sammenliknes.

Et annet sentralt arbeid angår beregning av kalkulasjonsrenta. (se for eksempel Minken 2005) Begrunnelsen for bruk av kalkulasjonsrente i slike analyser er at virkningene (positive og negative) av prosjektene vil inntreffe på ulike tidspunkt, og en slik årlig diskontering vil gjøre det mulig å sammenligne disse størrelsene med samme kroneverdi for et gitt referanseår. Generelt vil en anta at en virkning som oppstår om kort tid, bør tillegges større vekt enn en som oppstår lenger fram i tid. Det valgte nivået for kalkulasjonsrenten har stor innvirkning på resultatet av analysene, idet det uttrykker en forventning om avkastning på ressursbruken i prosjektet over en relativt lang tidsperiode (i hovedsak 25 år).

3.2.4. Oppsummering av nyttekostnadsetteranalyse som metodikk for kontinuerlig læring

Etteranalyser er sentrale for videreutvikling av ex ante nyttekostnadsanalyser, som er sentral faglig input for ressurstildeling innen samferdselssektoren. Det er også sentralt for standardisering og gir muligheter for å sammenligne for eksempel jernbane og veginvesteringer. Denne type metodeutvikling er eksemplarisk for det arbeidet som i økende grad søkes institusjonalisert for alle sektorer gjennom SSØs arbeid. Mens nyttekostnadsanalyse har vært et etablert verktøy i samferdselssektoren i mange år, er det lite utbredt på andre offentlige sektorer preget av omfattende ressursbruk.

Dersom vi skal summere opp trekk ved denne type evaluering, basert på de kriteriene vi beskrev i kapittel 1, er følgende sentralt: Det teoretiske rammeverket for nyttekostnadsetteranalyser er preget av at antagelser og forutsetninger er eksplisitt og klart formulerte. Undersøkelsen tilfredsstillt kravene til et klart teorigrunnlag ved at begreper er operasjonaliserbare. Det mest kritiske punktet i operasjonaliseringsprosessen er usikkerhet ved operasjonalisering av komponenter som i utgangspunktet ikke er beskrevet i økonomiske termer: Tidskomponenter og eksterne kostnader som miljø- og ulykkeskostnader. Gitt disse forutsetningene tilfredsstillt metodikken i stor grad kriteriene for teoretisk validitet.

Nyttekostnads etteranalyse opererer med to typer statistiske sammenhenger. Den statistiske validiteten er avhengig av disse sammenhengene. For det første operer den med sammenhenger som uttrykkes gjennom selve nyttekostnadsanalysen, som krever høy spesialisering og er lite transparent, og for det andre framstillingen av avvikstabeller som gir en oversiktlig avviksanalyse.

Undersøkelsesmetoden gir også klare og presise forutsetninger for å kunne undersøke og drøfte intern validitet, og forutsetningene for å hevde at det er årsakssammenheng mellom et tiltak og observerte endringer. Nettopp denne typen forutsetninger var ikke tilstede i de etterundersøkelsene vi beskrev ovenfor. Klare spesifiseringer av forutsetningene førte til en like klar konklusjon om at etteranalyse ikke kunne gjennomføres. Slik sett viser dette hvordan klare formuleringer om forutsetninger for evalueringen, kan få konsekvenser der evaluerer sier at oppdraget om etteranalyse ikke kan gjennomføres.

Den eksterne validiteten og dermed generaliserbarheten er primært knyttet til utvikling av metodikken, dels som gjennomgang og forbedring av selve metodehåndbøkene og ensartet metodikk for vei og jernbane. Det er også i denne

forstand denne type undersøkelser er kumulative, som utvikling av metodeverktøyet. De refererte evalueringene til Eriksen og Minken viser også klare feil og mangler ved de metodehåndbøkene som har vært benyttet. De kommer med en rekke konkrete forslag til forbedringer for å redusere feil i fremtiden.

Oppsummert er nyttekostnads-etteranalyser en evalueringstradisjon preget av høy grad av formalisering som imøtekommer de grunnleggende kravene til evalueringsmetodikk. Formålet er å utvikle og forbedre metodeverktøyet slik at en gjør færre feil i framtidige Nyttekostnadsanalyser. Særlig en type prosjektgevinster og -kostnader blir ikke empirisk etterprøvd gjennom et slikt opplegg, og det er usikkerheten knyttet til pengevurdering av tids-, miljø- og ulykkeskostnader. Innenfor dette fagfeltet, blant økonomer som jobber med nyttekostnadsanalyse, foregår det imidlertid et relativt systematisk og kontinuerlig arbeid med tidsverdsettingsundersøkelser for å forbedre kunnskapen om tidsverdsetting basert på individenes egne verdsettinger, eventuelt utvidet til kartlegging av ulike trafikantsegmenters verdsettinger.

3.2.5 Interdisiplinær kritikk mot nyttekostnadsanalyser

Etter det vi kjenner til er det ikke reist vesentlig kritikk spesielt mot nyttekostnadsanalyser som etteranalyser. Det er imidlertid en omfattende kritisk litteratur om nyttekostnadsanalyser generelt og innen samferdselssektoren mer spesielt. Deler av kritikken mot nyttekostnadsanalyse innen samferdselssektoren, tar feste i den mer allmenne diskusjonen, som har fulgt i kjølvannet av at nyttekostnadsanalyse har fått økt utbredelse. Denne omhandler dels det normative fundamentet for nyttekostnadsanalyse som grunnlag for kollektive beslutninger (for en oversikt av denne type kritikk, se for eksempel Gillroy og Wade 1992), dels problemer knyttet til verdsetting av naturen og miljøgoder (se for eksempel Foster 1997).

Vi skal her sammenfatte hovedpoeng i den kritikk Næss (2006 a og b) har fremmet om nyttekostnadsanalyse innen samferdselssektoren, men som tar utgangspunkt i den mer generelle kritikken. Vi kan sammenfatte Næss sin kritikk i følgende i tre generelle argumenter – og to mer samferdselsspesifikke. De tre generelle argumentene er:

Argumentet om sviktende aktørforutsetninger: Den teoretiske aktørforutsetningen som ligger til grunn for nyttekostnadsanalysen – at aktørene er fullt informerte forbrukere som er nyttemaksimerende – er ikke gyldig. Næss støtter seg på den vanlige sosiologiske antagelse om at mennesket også er norm- og vanestyrkt. De kan ha verdier som går mot deres preferanser som forbrukere på et gitt område. Implikasjonen, slik Næss ser det, er at nyttekostnadsanalyse fremmer offentlig deregulering under dekke av vitenskaplig argumentasjon.

Argumentet om neddiskontering og manglende bærekraft: Næss hevder at nyttekostnadsanalyse bygger på en antagelse av at naturen helt eller delvis kan

erstattes av menneskeskapte verdier, og at naturen har ubegrenset kapasitet til å bære forbruks- og produksjonsvekst.

”Uden denne antagelse ville det ikke være mulig at diskontere negative langtidskonsekvenser på miljøet ned til negative nutidsverdier, som er kun en lille brøkdel af de oprindeligt udregnede negative værdier. Underdrivelsen af de fremtidige konsekvenser ved hjælp av diskontering baseres på den antagelse at forbrugsniveauet stadig stiger (Næss 2006b, 205).”

Næss mener at dette argumentet illustreres ved de kontraintuitive konklusjoner en kommer til ved neddiskonterte omkostninger av klimagasser. Med en rente på sju prosent har en ”klimakatastrofe” som skjer om 150 år en neddiskontert omkostning som er 26.000 ganger mindre enn en ikke-diskontert omkostning. Det generelle argumentet er at nyttekostnadsanalysen impliserer en forutsetning om dekobling av økonomisk vekst og ikke-bærekraftig forbruk.

Argumentet om manglende transparens. Nyttetekostnadsmodeller er matematiserte og lite transparente. Næss mener mangelen på transparens åpner for manipulasjon med data. Næss sitt alternativ er mer transparente modeller. Han foreslår å bytte ut matematiske trafikkmodeller med mer kvalitative vurderinger av de faktorer som kan bidra til å skape mer eller mindre trafikk. Heller en nyttekostnadsbrøker argumenterer han for en grov karakterskala som dekker de fleste dimensjoner, og slik ivaretar krav til sammenlignbarhet.

Så langt de generelle argumentene, i tillegg til dette er det to typer argumenter som er knyttet mer direkte til samferdselssektoren:

Argumentet om vegfavoriserende transportmodeller: Næss trekker her også fram den kritikken som er blitt reist mot transportmodeller blant annet av Tennøy (2004). Hun mener at ”typiske matematiske trafikkmodeller” har en tendens til å undervurdere at nye veger skaper vekst i trafikken og at andre virkemidler kan påvirke denne veksten.

Argumentet med problemet med aggregerte tidsverdier: Næss sin argumentasjon på dette punktet, samsvarer med den vi finner hos Sverre Strand (2006 og 2007). Han peker på at tidsverdien gjerne vil være den største verdien i mange nyttekostnadsanalyser, og peker på visse grunnlagsproblemstillinger knyttet til verdsetting av tid. Han skiller mellom tre trinn i tidsanalyser: 1) Måling av tid og tidsbesparelser, 2) aggregering av tidsbesparelser og 3) pengevurdering av tidsbesparelser. Det første problemet omhandler temaer om hvor mye større enn null enkeltbesparelsen må være for at den skal kunne omtales som signifikant nok til at tiden kan anvendes alternativt. Det andre problemet omhandler under hvilke forutsetninger tidsbesparelser kan aggregeres, mens det tredje problemet handler om konvertering av tidsverdier til pekuniære verdier. Strand hevder at alle tre er problematiske, men at standardtilnærmingen til problemet vil være at en antar at en gjennom tidsverdsetting løser det første og andre problemet.

Kritikk av kritikken

Vi har ikke forutsetninger for å drøfte gyldigheten av alle argumentene som Næss og andre utenfor transportøkonomiprofesjonen framfører, men vil komme med noen kommentarer. Argumentet om aktørforutsetningene er det grunn til å anta at

det er bred enighet om, i betydningen: Dette er de teoretiske forutsetningene for modellen – og disse er i varierende grad gyldige empirisk. Derimot vil det være mer omstridt om dette impliserer at nyttekostnadsanalyse fremmer deregulering forkledd som vitenskap. Det synes rimelig å si at nyttekostnadsanalyse vurdert som beslutningsmodell er individualistisk i den betydningen at den betrakter beslutninger som summen av individers preferanser, og ikke som resultat av kollektive beslutninger om hva som er rett og hva som er galt. Dermed blir gyldighetsområdet for modellen viktig: Skal den forstås som et alternativ til og en erstatning for politiske beslutningsprosesser, eller som et supplement som gir et bestemt faglig grunnlagsmateriale for politiske beslutninger? Dersom det første er tilfellet, at samfunnsøkonomiske analyser skulle være avgjørende for vurdering av hvilke tiltak som skulle realiseres, kan det være problematisk. Det er for eksempel presentert samfunnsøkonomiske analyser som indikerer at tilskuddsnivået til kollektivtransport i norske byområder ligger langt under det som er samfunnsøkonomisk optimalt (Bekken og Norheim, 2005). Å avgjøre hvilke konsekvenser dette skal få er i Norge et politisk spørsmål, blant annet fordi vi har lagt til grunn at politikk skal handle om realisering av de verdier og holdninger som utvikles gjennom offentlig debatt – og politiske forhandlinger. Et annet spørsmål er hvor grensedragingen mellom politikk og ulike faglige vurderinger (nyttekostnadsanalyse, miljøkonsekvensutredninger skal gjøres)

Argumentet mot å benytte kalkulasjonsrente i forhold til svært omfattende og langsiktige miljøproblemer er omstridt. Denne uenigheten er et stykke på vei interdisiplinær, i det mange økonomer vil forsvare prinsippet om at avkastningen av klimatiltak må kunne sammenlignes (se bla Grøvdahl og Hjelle, op cit.) om enn være åpne for betydelig usikkerhet (se Aaheim 2003). Politikken på dette området i dag i Norge er da også, ifølge Næss, preget av at myndighetene er forsiktig med å gi langsiktige miljøproblemer pengeverdi, i kontrast til i USA og til dels også Danmark der nyttekostnadsanalyser av klimagasstiltak har fått betydelig politisk aksept.

Argumentet om manglende transparens er mer problematisk. Det problematiske illustreres for så vidt av at Næss sin overordnede argumentasjon på dette punktet er inkonsistent i følgende betydning. Han antar, etter det vi forstår, at klimaendringer er en følge av utslipp av drivhusgasser. Dette er vurderinger som etter det vi kjenner til bygger på kunnskap produsert gjennom modeller som er lite transparente, komplekse og matematiske. Samtidig oppfatter han det som et argument mot transportmodeller at de har slike kjennetegn. Det forholdet at mange modeller, som for eksempel transport- og klimamodeller, er matematiske og komplekse og dermed lite transparente og kan manipuleres av brukerne, kan neppe sees på som et valid argument mot dem eller de funn de presenterer (trafikkvekst eller klimaendringer). Spørsmålet om manipulasjon innen evaluering av alle typer må møtes gjennom gode prosedyrer for etterprøvbarehet og kritikk, i de tilfeller der modellene er komplekse og lite transparente, av kollegaer som har den nødvendige kunnskap. Alternativt og som supplement er det et argument for ulike forsknings- og utredningsmiljøer som konkurrerer, og som gjerne gjør analyser på samme type tiltak.

Spørsmålet om kompleksitet bør ikke skilles fra spørsmålet om treffsikkerhet: Gitt at to modeller gir samme resultat eller samme treffsikkerhet, har vi grunner til å velge den som er enklest for flest mulig å sette seg inn i. Her peker Næss i sin

kritikk også i retning av alternativer, og den måten å imøtekomme noe av problemet knyttet til pengevurdering av i utgangspunktet ikke økonomiske og til dels også kvalitative data som støy, ulykker og tidsbruk, som flermålsanalyse representerer. Flermålsanalyse har som formål å assistere beslutningstakere som står overfor et problem der ulike verdier er i konflikt. Det overordnede siktemålet og tilnærmingen er å gjennomføre en systematisk alternativanalyse knyttet til ulike evalueringskriterier som uttrykkes på en objektiv og konsistent måte. Ikke alle variablene kan entydig kvantifiseres og rangordnes etter ikke-numeriske skalaer, men man forutsetter likevel at de ulike alternativene kan gis en score. Metoden framstår som et alternativ til nyttekostnadsanalyse i situasjoner med prioritering av mål som ikke er entydig kvantifiserbare, men begrenses ved at det primært framstår som en analyse for vurdering av et begrenset antall alternativer. I Norge har dette tatt blitt anvendt blant annet innen Concept-programmet (se for eksempel Jordange, Malerud, Minken og Strand 2007).

Argumentet om vegfavoriserende transportmodeller, dvs. at modellene undervurderer generert trafikkvekst som følge av veiutbygging, er i beste fall omstridt. Tennøy (2004) sin kritikk, som Næss delvis bygger på, har blitt avvist av økonomer (Johansen og Minken 2004) som viser til misforståelser av vurderingene og at transportmodellene har utviklet seg. I deres motargumentasjon viser de til modeller som en må kjenne til for å vurdere gyldigheten av argumentet. Alternativt må en, som mange gjør i tilfellet med klimamodeller, basere seg på tillit til prosedyrene og kontrollen i de vitenskapelige miljøene som produserer og utvikler slike modeller. Denne diskusjonen illustrerer noe av vanskene med å gjennomføre interdisiplinære debatter, og behovet for kontinuerlig utprøving av ulike, konkurrerende modeller. På den andre siden er dette også et spørsmål om hvilke modeller som er i bruk. For eksempel har en av Danmarks fremste eksperter på transportmodeller, argumentert for at de mest brukte modellene undervurderer generert trafikkvekst som følge av veiutbygging (Nielsen og Fosgerau, 2005).

De argumentene vi har diskutert her er generelle argumenter mot nyttekostnadsanalyse i samferdselssektoren, og de har ikke relevans for forholdet mellom før- og etteranalyser. De viser derimot at den teoretiske validiteten i en etteranalyse er annerledes dersom denne stilles innen et gitt paradigme, som for eksempel Nytttekostnadsanalyse enn når disse teoriene vurderes uten i fra, på andre premisser.

Oppsummering

Det er rettet kritikk mot nyttekostnadsanalyse på en rekke punkter. Mange av disse kritiske innvendingene er knyttet til at ulike disipliners teorier bygger på ulike forutsetninger, og må sees på som interdisiplinære kontroverser. Denne kritikken har minst to implikasjoner. For det første illustrerer den betydningen av å institusjonalisere interdisiplinær debatt, for både å få utøveren innen en gitt disiplin til å klargjøre ulike disiplinærs gyldighetsområde, og for å skjerpe kritikerne og kravene til saklig argumentasjon.

For det andre peker den på et sentralt punkt der det er grunn til å anta personer fra ulike disipliner vil være enige om er sentralt: Det er særegne problemer knyttet til økonomisk verdsetting av verdier som ikke umiddelbart kan gis numeriske og dermed heller ikke økonomiske termer. Miljøeffekter er et eksempel på slike

verdier. Et annet kritisk punkt er transformering av verdier som i moderne samfunn er gitt numeriske termer, som for eksempel tid, til økonomiske kategorier. Det er betydelig usikkerhet knyttet til slike operasjoner, en usikkerhet som overskrider den som blir presentert gjennom et konfidensintervall. Dette blir også illustrert av den variasjon som kalkulasjonsrenta har hatt i Norge, og den internasjonale variasjonen på dette feltet. Tilsvarende gjelder internasjonal variasjon i tidsverdier (Odgaard, Kelyy og Laird, 2005) Dette tilsier at en i en presentasjon av nytte- og kostnadskomponentene, bør uttrykke forutsetninger og usikkerhet eksplisitt.

Dette tilsier også at slik usikkerhet diskuteres åpent i det offentlige ordskiftet, et argument som styrkes av at nyttekostnadsmodeller er lite transparente. Som vi har pekt på ovenfor, er mangel på transparens neppe et selvstendig argument mot en modell, eller en type evaluering. Det er kombinasjonen av presisjon i fremstilling, usikkerhetsnivå ved prediksjoner og manglende transparens som tilsier at nyttekostnadsanalyse gis en framstilling i det offentlige ordskiftet som gjør politikere og andre i stand til å vurdere dem, hvilken informasjon de gir og hvilken informasjon de forenkler på hvilken måte.

3.3. Organisatoriske forklaringer på feilbudsjettering

Som det går fram av gjennomgangen i tidligere kapitler, er det en rekke andre forklaringer enn de rent metodiske som kan forklare budsjettoverskridelser og avvik. Vi skal i denne delen, først presentere noen forslag til organisatoriske eller institusjonelle årsaksvariabler, slik disse er blitt presentert nasjonalt og internasjonalt. Vi skal legge hovedvekten nettopp på Aalborg-miljøets systematisering av årsaker, og gi en framstilling av det forskningsdesignet og de funnene som dette miljøet har gjort når det gjelder årsaker, før vi diskuterer styrker og svakheter med dette designet.

3.3.1. Organisasjonsvariabler og strategisk budsjettering som tema

Flere forskningsmiljøer har vært opptatt av andre forklaringer enn den metodologiske. Mackie og Preston (1998), som i sin artikkel om emnet vektlegger mange metodologiske forklaringer, peker blant annet på politiske forpliktelser som at akkumulert politisk prestisje er avgjørende. De lanserer også forklaringer på at vi så ofte finner *ex ante* overoptimisme. De mener at det er "obvious incentives to bias the appraisal." Avbøtende tiltak kan være åpenhet og andre tiltak av organisatorisk art, men også "to spend a lot more on ex-post evaluation than is currently done."

Odeck og Brudeseth (2003) peker også på muligheten for at planleggere av strategiske årsaker underbudsjetterer, som en hovedforklaring på budsjettoverskridelser. Dette kommer i tillegg til at det er faglig vanskelig å anslå kostnader, det kommer inn uforutsigbare kostnader underveis, og det er for lite fokus på kostnadskontroll.

Strategisk underbudsjettering kom også opp under Stortingets høring om Riksrevisjonens undersøkelse av store vegprosjektoverskridelser. Noen momenter som kom fram i denne debatten (beskrevet i Strand 2004a og 2006), og sitatene er derifra), er:

- *Personlig og profesjonell kompetanse*
- *Plannivåer, ansvarsforhold og (kommunalt) press.* ”Jeg må også være så ærlig – etter tolv års fartstid her i Stortinget – å si at jeg kjenner igjen mange av disse prosjektene fra stortingsbehandlingen, og vet hvilket press det har vært fra stortingskomiteer, fra enkelte representanter og fra fylkesbenker for å forsere prosjekter som egentlig burde ha ventet på tur ”
- *Strategisk optimisme* og romslige overslag i seg selv som årsak. Vegdirektøren uttrykker om etatens usikkerhetsregel og mulige konsekvenser av den: ”Legg vekt på å oppgi det overslaget som vi meiner er mest sannsynleg [...] opererer ikkje med større reserver, for vi veit at eit romsleg overslag også kan vera kostnadsdrivande.” Annerledes formulert kunne dette føre til at entreprenøren kan utnytte slingringsmonnet i sin totalramme.
- *Irreversibilitet.* I hvilken grad er det mulig å stoppe et prosjekt dersom prosjektutviklingen viser at det bærer mot budsjettsprekk – eller er førstegangsbevilgning point-of-no-return?
- *Organisasjon og omorganiseringer.* Selve omorganiseringen kan ha effekter, for eksempel den tradisjonelle dobbeltrollen vegetaten har hatt som planlegger og utbygger. På den bakgrunn ble de i høringen pekt på at omorganiseringen av Vegvesenet fra nyttår 2003 kunne medvirke til å sikre større samsvar mellom plan og virkelighet. Dette er et tiltak som effekter burde være mulig å etterprøve systematisk fem år senere. I stortingsmeldingen påpeker komiteen dette helt eksplisitt når den ”vil bemerke at omorganisering i seg selv ikke løser noen problemer, men kun er et virkemiddel [...] ber Regjeringen på egnet måte orientere Stortinget om fremdriften [...] og om organiseringen har hatt en positiv effekt på prosjektstyringen i Statens vegvesen [...] herunder om omfanget av overskridelsene reduseres.”

Oppsummert har både de undersøkelsene vi har referert til over, og de synspunkter som ble diskutert under Stortingets høring, til felles at de opererer med både politisk press og strategisk underbudsjettering som forklaring på budsjettoverskridelser og feilaktige prognoser. Særlig strategisk underbudsjettering er en forklaring som det er vesentlige utfordringer ved å dokumentere empirisk. Vi skal nå rette oppmerksomheten mot en knippe med undersøkelser som har satt seg fore å gjennomføre studier som inkluderer bruken av organisatoriske og institusjonelle variabler, og aktørers strategiske tilpasning. Det danske forskningsmiljøet med Flyvbjerg i spissen har undersøkt en rekke undersøkelser med slike variabler.

3.3.2 Aalborg-miljøet og Flyvbjergs undersøkelse av budsjettoverskridelser

I Aalborg-miljøet har en operert med ulike varianter av mulige årsaker til budsjettoverskridelser og feilbudsjettering.²

Vi tar her utgangspunkt i Flyvbjerg (2006) skille mellom tre hovedmodeller for forklaring av problemet med desinformasjon:

- Tekniske modeller (det vil si feil i data og metoder)
- Psykologiske modeller, ”planning fallacy”, ”optimism bias”
- Politisk økonomiske modeller (strategisk misrepresentasjon, løgn)

Teknologiske modeller viser grovt sagt til den typen metodiske problemer som vi har beskrevet i forrige del av dette kapitlet: Upålitelige data, inadekvate modeller og personlige feilvurderinger.

Psykologiske forklaringer referer til at planleggere baserer seg på ubegrunnet optimisme snarere enn rasjonell kalkyle av nytte, kostnader og sannsynligheter. De fremmer initiativ som neppe kan løses inne budsjettammen. Overoptimisme er primært et kognitivt problem, en svikt i måten hjernen sorterer informasjon.

De politisk-økonomiske forklaringene betrakter planleggere og andre som fremmer prosjekter som aktører som bevisst og strategisk overvurderer nytte og undervurderer kostnader når de predikerer resultatene til et prosjekt. Dette gjør de for å øke sannsynligheten for at deres prosjekt vinner fram. Dette har sammenheng med insentivstrukturen – løgn lønner seg. Flyvbjerg spissformulerer en insentivstruktur for ”survival of the unfittest”. Ofte er det ikke de beste prosjektene som vinner fram, men de som er mest misrepresentert, de som ser best ut på papiret.

Flyvbjerg argumenterer for at det er den tredje forklaringsmodellen som har størst forklaringskraft. Dersom tekniske forklaringer var korrekte skulle feilprediksjonene i større grad ha vært normalfordelt rundt null. Dette gjelder imidlertid bare for veiinvesteringene. Videre burde en forvente bedring over tid, på grunn av læring. Det finner en heller ikke støtte for. De psykologiske forklaringene passer funnene noe bedre. De predikerer en undervurdering av kostnader og overvurdering av nytte, som samsvarer med funnene. Argumentet som Flyvbjerg anfører mot en slik tolkning er at dette ikke blir nevnt av

² I noen versjoner, slik vi finner dem hos Flyvbjerg, Skamris Holm og Buhl (2002, 2004) fordeles observasjoner og hypoteser på fire til dels overlappende hovedgrupper; tekniske, økonomiske, psykologiske og politiske forklaringer. Disse kan oppsummeres slik:

- Tekniske forklaringer: Uutviklet metodikk, usikre data, generelle problemer knyttet til framtidsutsagn, manglende erfaring og kompetanse hos prognosemakerne, osv.
- Økonomiske forklaringer: Bevisst undervurdering av kostnader og/eller overvurdering av nytte for å fremme privatøkonomiske eller offentlige interesser knyttet til realiseringen av prosjektet
- Psykologiske forklaringer: Overoptimisme, også hos prognosemakerne selv, og iboende motvilje og evne til å lære av erfaringer
- Politiske forklaringer: Finansieringsstimulerte politiske føringer bak for lave kostnadsanslag

planleggerne selv. Videre peker han på at læring burde føre til, om ikke eliminering så i det minste, reduksjon av ”optimism bias”. Det finnes det ikke støtte for i materialet.

Det er (dermed) politisk-økonomiske forklaringer og strategisk representasjon som i følge Flyvbjerg passer best. Det metodiske problemet er selvsagt å få løgnere til å innrømme løgn (og dersom de gjør det, kan man stole på dem!), men i tillegg til at de to andre forklaringene ikke finner støtte, mener Flyvbjerg å ha funnet dokumentasjon på løgn og missinformasjon i to studier.

Dette får Flyvbjerg til å fremme forslag om en rekke endringer i insentivstrukturen for mega-prosjekter:

- Blokktilskudd til infrastruktur i stedet for øremerkede midler
- Uavhengige ”peer-reviews” og ”benchmarking” av alle prognoser
- Større offentlig innsikt i prognoser, bruk av borgerjuryer til vurdering av prosjekter og prognoser
- Vurdering av prognoser på faglige konferanser
- Straff til personer og organisasjoner som fremmer villedende prognoser
- Lokal medfinansiering av prosjektet pluss kontrakter som gjør prognosemakere finansielt medansvarlig
- Bare én prosjektansvarlig

3.3.3 Kritikk av Aalborg-prosjektet og Flyvbjergs argumentasjon

I det følgende skal vi kort summere opp noen viktige svakheter ved Flyvbjergs undersøkelse. I første del, som omhandler metodisk design og grunnlaget for konklusjonene, baserer vi oss på Arvid Strands opposisjon mot Flyvbjergs avhandling³.

En hovedkritikk er at mens Flyvbjerg primært har data om virkninger og bivirkninger, det vil si prosjektutfall i forhold til prognose, gjør han slutninger om aktørers intensjoner. Han skriver seg slik sett inn i en konspirasjonsteoretisk posisjon, en antagelse om at det ikke fins utilsiktede konsekvenser. Strand mener at de kvalitative dataene om dette ikke er fyllestgjørende for å kunne trekke denne konklusjonen. Videre er tesen om ”the survival of the unfittest” udokumentert, fordi dette forutsetter et datamateriale som inkluderer også de prosjekter som ikke vant fram. Viktige problemstillinger som kunne kastet lys over, og eventuelt nyansert de generelle konklusjonene, blir ikke drøftet. Eksempler her er: Hva kjennetegner prosjekter som har vært vellykkede med hensyn til prognoser for kostnader og trafikk tall? Hvorfor ”lyger” ikke vegprosjekterende like mye som andre prosjekterende? Hva kjennetegner prosjekter som har falt ut og ikke er blitt realisert, sammenlignet med de som har fått gjennomslag?

Strands kritikk impliserer at hovedkonklusjonene som blir trukket i denne undersøkelsen mangler metodisk fundament. Kritikken peker også på de

³ Takk til Arvid Strand for å ha stilt opposisjonsnotatene til vår disposisjon.

omfattende problemene en står overfor, når en skal søke å kartlegge strategisk, opportunistisk atferd empirisk.

I tillegg til de punktene Strand framhever i sin kritikk, er det et annet, analytisk poeng knyttet til det som blir benevnt som politisk-økonomiske modeller. Denne modellen rommer en kombinasjon av antagelser om insentivstrukturer og atferdsforutsetninger. Denne atferdsforutsetningen bygger på, la oss kalle han, "the project man" – en aktør hvis overordnede ambisjon er å få aksept for og tilslutning til det prosjektet som vedkommende har utredningsansvar for. Den allmenne karakteristikken er at insentivstrukturen gjør at slik atferd lønner seg og dermed styrkes. Derimot er modellen underspesifisert i forhold til ulike kombinasjoner av aktørforutsetninger og institusjonelle betingelser. Det er nettopp slike varianter av institusjonelle betingelser som følger av Flyvbjergs forslag til reformer. Av dette følger at det er avgjørende å utvikle et design som klarere spesifiserer både ulike aktørforutsetninger som variable – og ulike institusjonelle forutsetninger som variabler.

To poeng kan dermed løftes fram. For det første har vi i Norge ikke gjort systematiske studier av effekten av organisatoriske og metodiske variabler for feilbudsjettering – og dermed for læring og forbedring. Her ligger et interessant område av relevans for norsk samferdselspolitikk. For det andre: De internasjonale undersøkelsene som har sin basis i Aalborg-miljøet, har klare teoretiske og metodiske svakheter. Dermed kunne et slikt norsk arbeid, dersom det overskred disse svakhetene, være et bidrag.

3.4.1 Oppsummering av organisatoriske og institusjonelle forklaringer på feilslåtte analyseforutsetninger

I Norge har få evalueringer søkt å inkludere organisatoriske eller institusjonelle variabler i forklaringer på underbudsjettering eller andre feilslåtte forutsetninger ved større samferdselsprosjekter. Rett nok har slike variabler vært fremhevet av Riksrevisjonen, både ved spesifikke vegutbyggingsprosjekt og bompengeprosjekter, men det er ikke blitt systematisk sammenlignet med metodiske problemstillinger.

Blant de mer framtrede internasjonale forskningsmiljøene på dette området har Aalborg-miljøet, og særlig Bengt Flyvbjerg vært. Flyvbjerg forstår overskridelser som resultat av desinformasjon, og han lokaliserer tre hovedmodeller for å forklare slik desinformasjon: Disse er tekniske modeller som refererer til feil i data og metoder; psykologiske modeller som refererer til kognitive feil; og politisk-økonomiske modeller i form av strategisk misrepresentasjon og løgn. Flyvbjerg argumenterer for at hans analyse viser at de to første modellene ikke kan forklare desinformasjon og feilbudsjettering, blant annet fordi en da skulle forvente læring og redusert feilbudsjettering over tid. Forklaringen blir da at insitamentstrukturen er slik at fordelene med å desinformere er større enn ulempene. Resultatet blir en "survival of the unfittest" – det er ikke det beste prosjektet som vinner men det som har best desinformasjon. Botemidlene på dette problemet er dels utvikling av bedre metoder for å estimere kostnader, nytte og risiko gjennom "reference class forecasting", dels endrede insitamentstrukturer som vil begrense desinformasjon og fremme informasjon.

Teoretisk er tilnærmingen til Flyvbjerg eklektisk uten en klar disiplinforankring. Fordelen med en slik tilnærming er at den spenner opp et relativt bredt analytisk rammeverk som inkluderer mange relevante variabler, både metodiske, psykologiske og organisatoriske. På den andre siden er aktørforutsetningene og antagelsene om institusjoners virkemåte noe underspesifiserte, men drøftingen indikerer at disse er forankret i institusjonell økonomi, og antagelser om opportunistisk atferd.

Det kritiske punktet i analysen synes å være den interne validiteten, det vil si hvordan sammenhengen mellom variablene fastlegges. Det er for eksempel få klare empiriske indikatorer på løgn, men denne forklaringsvariabelens kraft fastlegges gjennom et kontrafaktisk resonnement. Fordi de andre forklaringsmodellene ikke finner støtte gjenstår den siste. Det metodiske problemet Flyvbjerg står overfor er klassisk for de som forsker innen tradisjoner som omfattes av "mistankens hermeneutikk" (Skjervheim 1991), det er vanskelig å finne empirisk støtte for at løgnere lyver, enten de vet det selv eller ikke.

Tilnærmingen til Flyvbjerg og Aalborg-miljøet er imidlertid svært relevant, fordi den åpner for en systematisk evaluering av feilbudsjettering ved store samferdselsinvesteringer som ikke begrenser seg til forbedring av metodikken og dermed vekt på ex post nyttekostnadsanalyse, men også på de organisatoriske og institusjonelle variablene.

Her ligger også trolig et stort potensial for videreføring av evalueringer om store samferdselsinvesteringer som også inkluderer andre variable en bedret metodikk: En systematisk komparativ undersøkelse som analyserer ulike metoderegimer under ulike institusjonelle forhold/insentivsystemer, kombinert med ulike aktørforutsetninger.

Ulike metoderegimer: Det følger av det arbeidet i Norge som er gjort og er i ferd med å bli standardisert på metodesiden, at en skal utvikle mer ensartede analyser, som vil redusere muligheter for manipulasjon til produksjon av data som input til modellene. Dersom metodiske forklaringer, eller det som Aalborg-miljøet omtaler som tekniske forklaringer likevel har stor relevans, skulle en kunne observere endringer. Det er jo nettopp slike endringer den sterke vektleggingen på mer ensartet metodikk er ment å skulle føre til.

Ulike institusjonelle omgivelser/insentivsystemer: De reformer som Flyvbjerg og hans kollegaer har foreslått er dels knytte til økt transparens (uavhengige "peer review", "benchmarking" av alle prognoser, større offentlig åpenhet og innsikt i prognoser), økonomiske insentiver (rammetilskudd til infrastruktur heller en øremerking, og økt finansielt medansvar for mottakere av statlige midler), og ansvarliggjøring/rechtsliggjøring (straff for feilprognosering, bare en prosjektansvarlig). Men her foreligger det allerede i dag variasjon og endring over tid, både nasjonalt og internasjonalt, så en måte å videreføre analysene på, ville være å undersøke om vi finner systematisk variasjon i feilbudsjettering i situasjoner der de institusjonelle forholdene er endret i de retningene som Flyvbjerg har foreslått.

Her bør det også inkluderes forholdet mellom ulike faser av beslutningsprosessen; trekk ved organisering av utredningsarbeidet, organisering av beslutningene og av iverksettingen av ulike samferdselsinvesteringer. Det følger av de hypoteser som er formulert både hos Flyvbjerg og av andre, at politiske føringer påvirker

nyttekostnadsanalyser, som er noe annet enn at prosjekter vinner fram selv om disse har en lavere beregnet nytteverdi enn andre prosjekter. Det siste er en helt legitim politisk vurdering. Avgjørende blir dermed å kartlegge variasjon i om eller når politisk press påvirker resultatet. Tilsvarende kan i og for seg riktige vurderinger få andre resultater som følge av trekk ved iverksettende organisasjon.

I sum gir dette følgende variabelliste som minst ville være sentral for komparativ analyse av større samferdselsprosjekter, og som en ville undersøke om de forklarte variasjon i treffsikkerhet i analysen:

Utredningsfasen:

- Metodikk, det vil si etter hvilken metodehåndbok med mer
- Organisering og rolledeling: Hvem foretok utredning, var disse i organisasjon med økonomisk interesse i utfall? I hvilken grad var utredningsarbeidet transparent? Var andre aktører enn utrederne involvert i framlegging av samlet prioriteringsdokument?

Vedtaksfasen:

- Rolledeling og kartlegging av aktører som var involverte, hvilken plassering de med interesser i utfall hadde i beslutningsprosessen?

Iverksettings-/byggingsfasen:

- Rolledeling og risikofordeling av privat og offentlig kapital, kontraktsutforming med videre

Det er sentralt i en slik undersøkelse å kartlegge trekk ved de prosjekter som vant fram og de som ikke gjorde det. I forhold til norske undersøkelser er oppdeling av planlegging og utbygging i etatene et viktig skille, som skulle forventes å påvirke feilbudsjettering.

På alle disse dimensjonene - metodikk, institusjonelle omgivelser for utredninger, og trekk ved organisering av utredning, politiske beslutninger og iverksettende organisasjon - foreligger de nå omfattende variasjon og endring over tid i Norge. Her er det dels gjennomført tidligere analyser som en kan ta utgangspunkt i, og Sverre Strand (2006; 2007) har sammenfattet og pekt på mulige cases. Når det gjelder den siste variabelen, ulike atferdsforutsetninger, antar vi at en måtte utvidet til et internasjonalt, komparativt prosjekt.

3.4 Kort om felles implikasjon av to ulike tradisjoner

Vi har her gjennomgått to typer evalueringer, nyttekostnads-etteranalyser og den mer organisasjonsteoretisk forankrede forskningen som vi finner i Aalborg-miljøet. De har det til felles at de omhandler større samferdselsinvesteringer, men de skiller seg fra hverandre ved at den første har fokus på bedret metodikk, den andre på bedre institusjonelle rammer for metodikerne. De har hver sine kritiske punkter. Nyttekostnadsanalysens er gyldigheten av analyseforutsetningene og da spesielt i forhold til økonomisk verdsetting av tid, ulykker og miljø. Aalborg-tradisjonen bygger problematiske forutsetninger knyttet til metodikken for kartlegging av bevisst normbrytende handling i form av løgn og desinformasjon. En lærdom fra den eksterne kritikken mot

nyttekostnadsanalyse er at den synliggjør hvordan forståelsen av teoretisk validitet i utgangspunktet er teori-intern, knyttet til teorien og dermed disiplinens indre logikk. I tilfellet nyttekostnadsanalyse kan neppe selve teorien falsifiseres, men dens gyldighetsområde blir synliggjort gjennom ekstern kritikk.

Det er en felles lærdom fra begge tradisjoner, det avgjørende punkt der gjennomgangen av ex post nyttekostnadsanalyse (og usikkerhet knyttet til denne) og de komparative organisasjonsanalyser ut til å sammenfalle: Behovet for offentlig debatt og kritikk er til stede for begge. Nyttekostnadsanalysen er preget av et høyt presisjonsnivå i forhold til den nyttekostnadsbrøken som presenteres, noe som bør ha sitt tilsvarende i en åpen dialog om usikkerhet knyttet til forutsetninger for slik presisjon. Og nettopp offentliggjøringsprisnippet er en av de klare forslag som munner ut av Aalborg-prosjektet.

De to evalueringstilnærmingene skiller seg altså fra hverandre på det punktet at ex post nyttekostnadsanalyse begrenser seg til å undersøke og forbedre metodikken, mens metaprojekt-evalueringene inkluderer organisasjonsvariabler og dermed behovet for omstrukturering av organisatoriske og institusjonelle omgivelser. Denne gjennomgangen synliggjør også særegne problemer knyttet til denne type organisasjonsanalyse, som skal drøftes ytterligere i neste kapittel.

4. Evalueringer som organisasjonsanalyse

Vi avsluttet forrige kapittel med en type analyse som inkluderer organisasjonsvariabler som forklaringer på bestemte fenomen. Vi skal her gå videre på ulike varianter av organisasjonsevaluering, og ta for oss to ulike undersøkelser som har til felles at de omhandler organisatoriske variabler, men skiller seg på en rekke andre dimensjoner: Evaluering av arbeidsmiljøtiltak i transportsektoren, evaluering av TP 10 som prosess. Denne drøftingen blir utvidet med noen momenter fra en tredje evaluering om organisatoriske endringer og tilpasninger for å fremme internasjonal godstransport med tog.

4.1. Evaluering av arbeidsmiljøtiltak i bussbransjen

Syv organisasjoner fra arbeidsgiver- og arbeidstakersiden tok i 2001 initiativet til et felles prosjekt for å motvirke den negative utviklingen for ansatte innen norsk rutebiltransport. Bakgrunnen var et vedvarende høyt sykefravær, høy gjennomstrømning og tidlig avgangsalder blant ansatte i bransjen. Formålet var å utforme og prøve ut tiltak som kunne redusere belastende forhold for førere og bedre arbeidsmiljøet på kort og lang sikt. Mer konkret ønsket prosjektgruppen å prøve ut to tiltak i utvalgte bedrifter: (i) innføring av team- eller gruppearbeid i selskapene og (ii) endring i skiftordningene i mer helsevennlige retninger.

Hovedprosjektet ble etablert under navnet ”Friskbuss” i 2002 og Transportøkonomisk institutt (TØI) ble tildelt ansvaret for å gjennomføre den faglige evalueringen av tiltakene. Tre busselskaper meldte seg som frivillige til å prøve ut tiltak – Nettbuss Drammen, Sporveisbussene/Nexus og Schøyen Bilcentraler AS – og teamarbeid var det eneste tiltaket som ble forsøkt innført i alle tre deltakerbedriftene. Tiltakene ble startet opp i løpet av årene 2002/2003 i alle bedriftene og virket i tre år før siste undersøkelse ble gjennomført fra TØIs side.

Den overordnede problemstillingen i evalueringsrapporten kan formuleres som følger: I hvilken grad kan innføringen av teamarbeid i de tre deltakerbedriftene sies å ha ført til bedringer i førernes arbeidsmiljø? Mer konkret forsøker rapporten å svare på tre spørsmål:

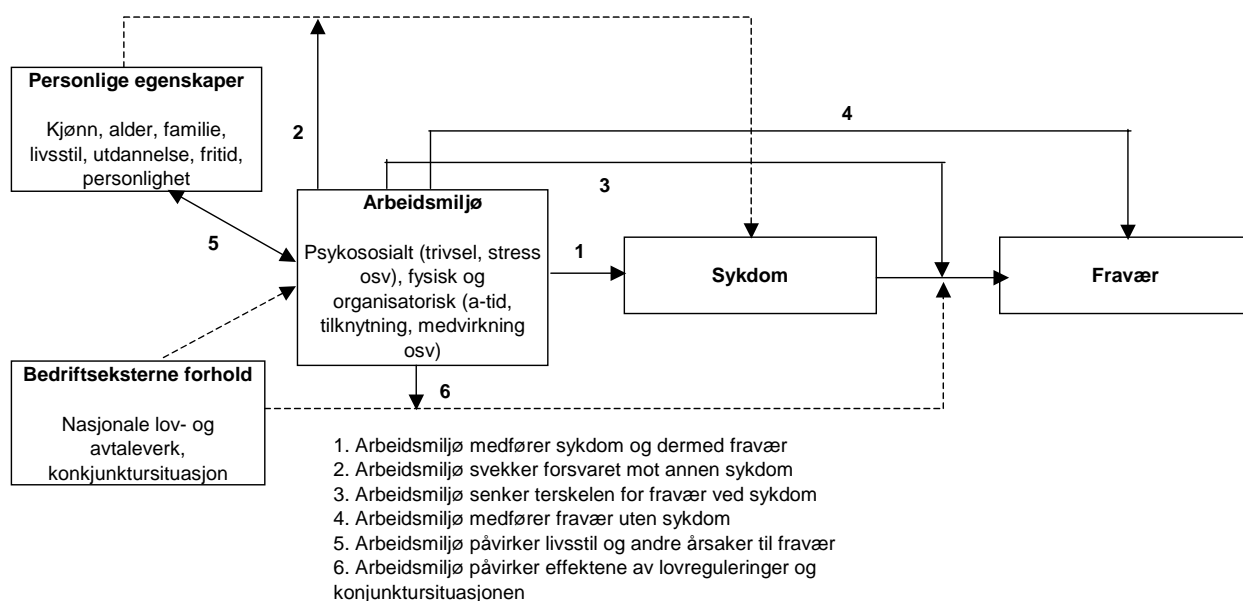
1. Hva påvirker jobbtilfredsheten hos førerne i de tre bedriftene?
2. Hva påvirker det jobbrelaterte fraværet?
3. Hvordan har innføringen av team påvirket årsakene til førernes tilfredshet og fravær slik det fremkom av spørsmål 1 og 2 ovenfor?

4.1.1. Teoretisk modell og avgrensning

Rapporten peker på at tiltakene skulle være rettet mot å forbedre arbeidsmiljøet blant bussførerne, som igjen skulle føre til redusert sykefravær for bedriftene. Arbeidsmiljø er imidlertid bare en av mange faktorer som påvirker fraværsnivået i en bedrift. Generelt sett er det vanlig å ta utgangspunkt i en av følgende årsaksfaktorer bak sykefravær:

1. Demografiske faktorer (alder og kjønn)
2. Situasjonen på arbeidsmarkedet (konjunktur- og arbeidsledighetsnivå)
3. Yrkes- og arbeidsmiljøfaktorer (fysisk, organisatorisk og psykososialt arbeidsmiljø)
4. Fraværskultur og holdninger (arbeidsmoral og normer for fravær)
5. Familiesituasjonen (omsorg og annen sykdom i hjemmet)
6. Nasjonale kompensasjonsordninger (nivået på sykepengen)

Figur 6 illustrerer ulike måter arbeidsmiljøet kan påvirke fraværet på.



Figur 5. Sammenhenger mellom arbeidsmiljø, sykdom og fravær

Kilde: Longva m fl 2007 og Longva og Ruud 2003

Selskapene kan med andre ord oppleve positive arbeidsmiljøeffekter av tiltakene uten at disse nødvendigvis slår ut i lavere fraværstall for førerne, ettersom andre faktorer kan dra i motsatt retning.

4.1.2. Metodisk opplegg og avgrensning

De sentrale variablene som en søker å kartlegge sammenhengen mellom i undersøkelsen er *jobbrelatert fravær, jobbtilfredshet og teamorganisering*. Dette er undersøkelsens tema. Undersøkelsen begrenser seg til sjåførere i en bestemt type bedrifter; *motiverte bedrifter i større byområder som ønsker å ta i bruk bevisste tiltak for å øke jobbtilfredsheten og redusere sykefraværet*. Det at de er motiverte for å gjøre noe med sykefraværet, og ønsker å ta i bruk tiltak, viser de ved å inngå i prosjektet.

Å undersøke hvordan teamorganisering, som er et organisatorisk tiltak, påvirker jobbtilfredshet og fravær, innebærer bestemte typer metodiske utfordringer som vi kort skal presentere og vise hvordan forskerne har håndtert:

Behovet for data om endring over tid: For å svare på en slik problemstilling, er det første kravet som må være oppfylt at en har data om jobbtilfredshet, på minst to tidspunkter: Før og etter introduksjonen av team

Data fra to ulike tidspunkter, såkalte diakrone data som har det til felles at vi kan undersøke endring over tid, kan være av ulik type (se for eksempel Hellevik, 1991: 389). For det første data som er samlet inn på ett tidspunkt, men som omhandler to tidspunkt, det vil si erindringsdata eller bruk av retrospektive/historiske spørsmål. Friskbussevalueringen har noen slike data i form av intervjuer som også inkluderer informantenes vurdering av fortidige hendinger.

For det andre data som er samlet inn på to tidspunkter, som vi deler inn i to typer:

- *Tidsseriedata* (der en spør ulike utvalg i samme populasjon på to ulike tidspunkt), der det spesielle problemet er om utvalgene er sammenlignbare og at en bare kan måle endring på aggregatnivå.
- *Panelstudie* (der en spør samme utvalg på ulike tidspunkt), der de spesielle problemene er eventuell kontrolleffekt og respondentbortfall.

I denne evalueringen har en tidsseriedata fra spørreundersøkelser gjennomført i de berørte bedriftene to ganger.

Kontroll for tredje variabler: Diakrone data er nødvendige for å kunne konkludere med om det har skjedd endring over tid, men de er ikke tilstrekkelige. En observert positiv endring i arbeidsmiljøet som i tid sammenfaller med introduksjon av teamorganisering kan ha sin årsak i andre forhold/variabler. Noen eksempler er: Utskiftning av personell, bedring av bedriftens rammebetingelser (for eksempel arbeidsmarkeds- og konkurransesituasjon). Dette krever ideelt sett et forskningsdesign som gjør at en har kontroll på alle andre variabler og bare manipulerer én variabel, den en ønsker å undersøke effekten av. I sosialpsykologiske undersøkelser blir et slikt forskningsdesign gjerne søkt etablert ved å lage en laboratorie-eksperimentsituasjon. En (eksperiment)gruppe påvirkes av en variabel, og en annen gruppe, kontrollgruppen, gjør ikke det, slik at en har kontroll på andre variabler. En svakere variant er å ha kontrollcase – eller bruke kvasi-eksperimentell metode – der en også studerer bedrifter eller deler av bedrifter som ikke utsettes for påvirkning fra Y, men som ligner de andre undersøkelsesbedriftene på andre områder.

I denne evalueringen gjør de til en viss grad bruk av slike kontrollgruppevurderinger, ettersom de har samlet inn data fra flere avdelinger i Sporveisbussene/Nexus enn den avdelingen hvor gruppearbeid ble innført. Disse fungerer til en viss grad som kontrollcase for bedriftseksterne påvirkningsfaktorer, som arbeidsmarkedssituasjon og endringer i sykelønnslovgivningen. Avdelingene er imidlertid vidt forskjellige i størrelse og rutestruktur, samtidig som de har gjennomgått til dels store endringer i løpet av tiltaksperioden, slik at nytten av en slik kvasi-eksperimentell tilnærming ikke må overdrives. Det vil i praksis være umulig å kontrollere alle eksterne faktorer som kan tenkes å påvirke en bedrifts arbeidsrelaterte sykefravær *samtidig*.

Utover delvis bruk av kontrollcase i tilfellet Sporveisbussene/Nexus, har de også forsøkt å oppveie svakhetene som ligger i mangel på kontroll av andre variabler gjennom omfattende datainnsamling i de ulike bedriftene, og ved kombinasjon av ulike metoder.

Metodetriangulering innebærer å kombinere ulike metoder, ofte kvalitative og kvantitative metoder, for å utnytte de ulike metodenes fortrinn og redusere deres svakheter.⁴ Kvantitative metoder er gjerne knyttet til ekstensive opplegg der en får kunnskap om fordelingen av relativt få, ferdigdefinerte variabler på mange enheter, de gir grunnlag for oversikt og statistiske analyser. Kvalitative metoder gir i større grad grunnlag for innsikt enn oversikt, de gir oss forståelse av kategoriers innhold snarere enn deres utbredelse, de er gjerne knyttet til intensive opplegg der en får kunnskap om mange variabler knyttet til få undersøkelsesobjekter. En antar gjerne at ulike typer data kan være komplementære, slik at det å kombinere metoder, såkalt metodetriangulering, gir oss bedre kunnskap enn bruken av en metode (Denzin 1978, Grønmo, 1995).

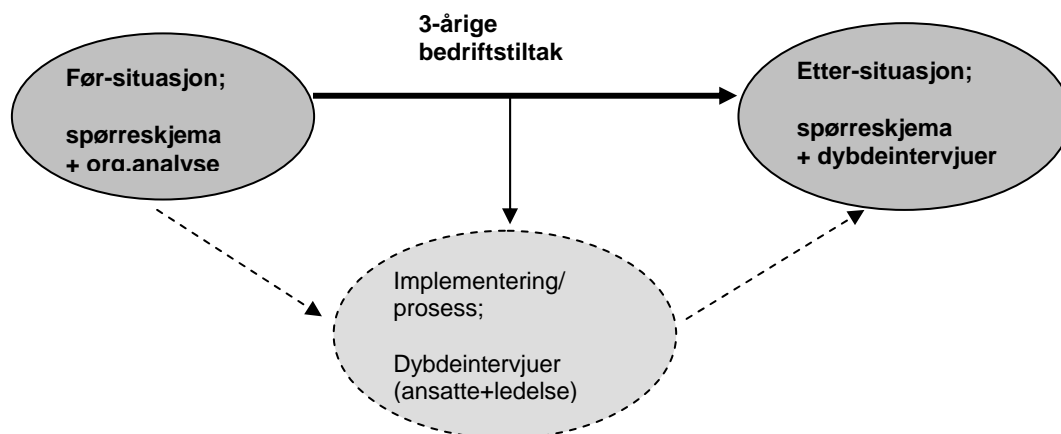
I Friskbuss-undersøkelsen ble det gjennomført både før- og etterundersøkelser, der det anvendes både kvalitative og kvantitative tilnæringsmåter. Følgende metoder er tatt i bruk i studien:

- Spørreskjema til alle arbeidstakere i de delene av bedriftene der det var iverksatt tiltak i henhold til prosjektets hypoteser om hva som har effekt. Skjema ble sendt både i før- og etterundersøkelsen.
- Observasjon av prosessen rundt iverksettingen av tiltakene. Dette gav samtidige data fra gjennomføringen av tiltakene.
- Intervjuer med representanter for toppledelse, mellomledelse og arbeidstakere i bedriften. Slike intervjuer ble også gjennomført både tidlig i prosessen og etter at tiltakene hadde vært i drift i bedriften en stund.

I tillegg til før- og etterundersøkelsen har de retrospektive data fra etterundersøkelsen. De gjennomførte også en kvalitativ prosess- og organisasjonsanalyse, med kvalitative data fra to tidspunkter. Dels var denne kvalitative undersøkelsen tenkt å gi utfyllende informasjon i forhold til de kvantitative dataene (hvorfor-spørsmål), de gav mulighet for kontroll for den

⁴ Benevnelsen viser til hvilken type data det er tale om. Kvantitative data er data som kan gis et numerisk uttrykk og forelegges som tall, mens kvalitative data foreligger som tekst. Kvantitative data har en forenklet form som gjør at de kan benyttes til komplekse regneoperasjoner og dermed modellering.

kvantitative undersøkelsen (stemmer de ansattes oppfatninger med organisasjonsforskjellene mellom de tre bedriftene?). Innsamlingen av kvalitative data hadde også en selvstendig funksjon (følge prosessen og ledelsesinformasjon). Den dokumenterer også hvordan selve teamutformingene i bedriftene varierer, slik at variabel x har flere verdier.



Figur 6. Illustrasjon av gangen i evalueringen og datainnsamlingen

Kilde: TØI-rapport 908/2007

I neste trinn i undersøkelsen ble det gjennomført en kartlegging av arbeidsmiljøet, basert på spørreskjema. Denne undersøkelsen viste at de viktigste faktorene ved arbeidsmiljøet som økte jobbtilfredshet var: At en sjelden var stresset, at en hadde et godt kollegialt miljø, at ledelsen var opptatt av og tok hensyn til de ansatte, at de kunne påvirke egen arbeidstid og ellers følte kollektiv medvirkning. De fleste av disse faktorene påvirket også jobbrelatert fravær.

Når selskapene i neste omgang etablerte team, så var hensikten nettopp at de skulle gripe fatt i disse faktorene. Midtveisevalueringen, fungerte som et informasjonsgrunnlag for å korrigere team- og tiltaksutformingene i bedriftene, men gav også data om organisasjonsutformingene i de respektive bedriftene. Den viste at teamene varierte langs flere variable, blant annet både med hensyn til størrelse og bakgrunn til teamleder.

Arbeidsmiljøet ble så kartlagt igjen tre år senere. Da hadde imidlertid Schøyen Bilcentraler tapt sin Oslo-avdeling gjennom anbud og det var nettopp denne avdelingen som var utgangspunktet for selskapets arbeid med ny teamorganisering. Det som da gjenstod var Nettbuss Drammen og Sporveisbussene, men her gikk det til gjengjeld an å trekke en del klare konklusjoner:

- Begge selskapene hadde opplevd redusert sykefravær i løpet av tiltaksperioden. Reduksjonen var større der tiltakene var blitt iverksatt enn ellers i selskapene og i landet forøvrig.
- Førerne var mer tilfreds med arbeidsmiljøet enn tidligere. Dette gjaldt særlig deres innflytelse over egen arbeidssituasjon.
- Opplevelsen av å ha en fast gruppe kolleger økte førernes kollegiale trivsel og følelse av innflytelse.

- Grupper hvor førerne selv inntok lederansvaret hadde like gode resultater som grupper hvor det ble innført et nytt administrativt lederledd.
- Tiltakene hadde hatt størst effekt i den bedriften hvor førerne var minst tilfreds med sin egen innflytelse i forkant av tiltakene.
- Innføring av gruppeorganisering er ikke og bør ikke være et enkeltstående og entydig definert tiltak. I begge bedriftene hadde flere forhold blitt tatt tak i parallelt med tiltakene, noe som hadde hatt avgjørende betydning for resultatene.

4.1.3. Trekk ved undersøkelsen

Det teoretiske grunnlaget for Friskbuss-evalueringen er spesifisering av en rekke variabler og antatte sammenhenger, der variablene er på ”mellomnivå”, dvs. at de er lett operasjonaliserbare variabler med lite klar disiplin- og dermed teoretisk forankring, slik som: Psykososialt arbeidsmiljø, fysisk arbeidsmiljø, sykdom, fravær osv. Sentrale begreper i undersøkelsen er jobbrelatert fravær, jobbtilfredshet og teamorganisering. En tok i bruk både kvalitative og kvantitative metoder, og har data fra både før- og ettersituasjonen. Undersøkelsen kan antas å være kumulativ, i det en har fått støtte for en hypotese om effekten av å redusere underorganisering i bestemte typer foretak, gitt bestemte forutsetninger. Undersøkelsen oppfyller kriteriet om statistisk validitet.

Undersøkelsen har et design som i utgangspunktet tilfredsstillende de fleste kravene en normalt vil stille til intern validitet, statistisk, og teoretisk validitet. Tilsvarende drøftes den ekstern validitet for foretak av samme karakter – og den tar utgangspunkt i en teoretisk modell som har en relativt etablert status i norsk arbeidsforskningsmiljø. Evalueringen er relativt unik fordi den har gått over en så vidt lang tidsperiode, slik at en faktisk har kunne måle endring over tid.4.2. Evaluering av TP 10 som prosess.

4.2.1. Bakgrunn

Vi skal i det følgende sammenfatte evalueringen av ”Transportarbeidet i de ti største byområdene” (TP 10) i Norge som prosess. Begrunnelsen for et slikt valg er todelt. Dels illustrerer denne evalueringen viktige trekk ved organisasjonsevalueringer av en bestemt type, dels er TP 10 et viktig skillemerke i norsk samferdselspolitikk, og den forståelse av ansvars- og maktfordeling innen sektoren som kom ut av denne evalueringen skulle komme til å få en sentral plass i senere arbeid på feltet.

TP-10 arbeidet ble satt i gang høsten 1989. Bakgrunnen var at det i de største byområdene var kapasitetsproblemer på vegnettet og høye kostnader knyttet til trafikkavviklingen. I tillegg hadde miljøvirkningene fått økt oppmerksomhet på slutten av 1980-årene. For å utvikle ”helhetlige transportsystemer som ivaretar hensynet til framkommelighet for alle trafikantgrupper, trafiksikkerhet samt helse og miljø, ble det satt i gang transportplanarbeid i de ti største byområdene i Norge: Fredrikstad/Sarpsborg (Nedre Glomma), Oslo/Akershus, Drammen, Tønsberg, Skien/Porsgrunn (Grenland), Kristiansand, Stavanger/Sandnes, Bergen, Trondheim og Tromsø.

I retningslinjene, fastsatt av Miljøverndepartementet og Samferdselsdepartementet, het det at en i arbeidet skulle:

- Samordne ulike forvaltningsnivåer og sektorinteresser slik at best mulige helhetsløsninger sikres
- Integre miljøhensyn i planarbeidet på en slik måte at det blir en premis for og ikke bare en konsekvens av de planer som i dag legges fram
- Vurdere arealplanlegging og transportsystem i sammenheng, og legge forholdene til rette for miljøvennlige transportformer

Det ble produsert en rekke evalueringsrapporter om trafikkberegninger, konsekvensutredninger, samfunnsøkonomiske kalkyler, luftforurensingsberegninger.

I den delen av TP 10-evalueringen som vi drøfter her, prosessevalueringen, fokuserer forfatterne av TP-10 evalueringen på hvordan sentrale signaler og konkrete virkemidler, sammen med lokale prosesser, har bidratt til lokale transportplaner. Evalueringen skal bidra til å belyse:

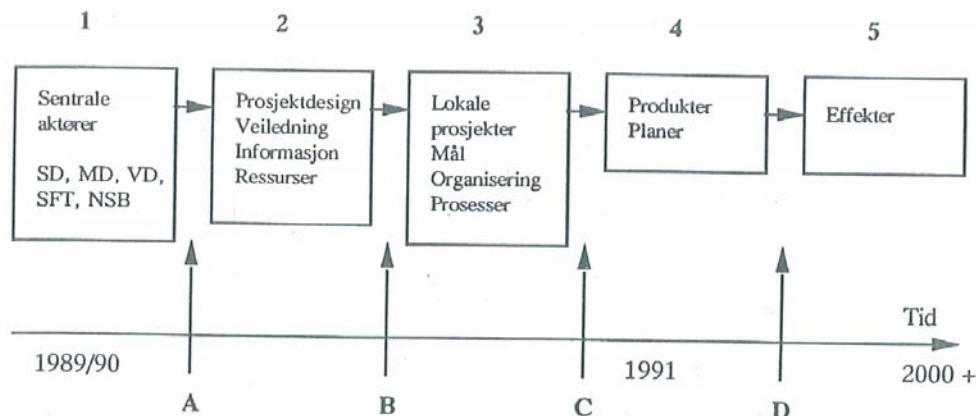
- Statens pådriverrolle og nødvendigheten av statlig initiativ for å få gjennomført lokalt areal- og transportarbeid
- Vurdering av opplegget på sentralt hold når det gjelder organisering, veiledning og lignende
- Vurdering av institusjonelle og forvaltningsmessige forhold, og spørsmålet om transportplanene vil bidra til å fremme helhetsløsninger uavhengig av formelle ansvarsgrenser mellom forvaltningsnivåene

4.2.2 Teoretisk og metodisk opplegg

Teamet som gjennomførte evalueringen var sentrale samferdselsforskere, forankret i statsvitenskap og planleggingssteori, to professorer, henholdsvis fra Universitetet i Oslo (UiO) og Norges teknisk-naturvitenskapelige universitet (NTNU) og forskere fra TØI og Norsk institutt for by- og regionforskning (NIBR).⁵

Figuren nedenfor illustrerer TP 10 som planleggings- og iverksettelsesprosess slik utredningsgruppen så prosessene.

⁵ Prosjektleder var professor Per Kristen Mydske fra UiO, de andre var professor Arvid Strand fra NTNU, Synnøve Lyssand Larsen og Per Gunnar Røe (NIBR), Tor Lerstang og Morten Stenstadvold (TØI)



Figur 7. Modell for evalueringen av TP 10 som prosess (NIBR og TØI 1992:11)

Kilde: TØI-rapport 929/2007

Det tematiske hovedfokuset for evalueringen var virkningen som den statlig initierte prosessen og valg av virkemidler hadde på lokale prosesser og planprodukter. Mer langsiktige effekter av slike prosesser ble oppfattet å ligge for langt fram i tid til at de kunne inkluderes innen en slik evaluering.

Utredningene skiller videre mellom målrasjonelle, effektorienterte og beslutningsorienterte tilnærminger: Ved målrasjonell tilnærming vurderer en grad av måloppnåelse og kostnadseffektivitet, noe som forutsetter relativt klare og entydige mål. I en effektorientert tilnærming søker en både intenderte og ikke intenderte effekter, mens innen en beslutningstilnærming vil evalueringen være konsentrert om styrbare faktorer som kan gi bedre grunnlag for beslutninger. De peker videre på at evalueringen har elementer av alle tre tilnærminger. De vurderer TP 10 ut fra sentrale mål og forutsetninger, og om disse er realisert i planinnhold, de søker også å kartlegge faktiske resultater eller mangel på resultater også utover de som det forelå sentrale føringer på. Evalueringen har også elementer av beslutningsorientering og skal gi grunnlag for å vurdere statens pådriverrolle, sentral organisering og institusjonelle og forvaltningsmessige forhold i virkemiddelbruk.

4.2.4 Design, datainnsamling og særlige utfordringer

Det var, som beskrevet ovenfor, et tverrfaglig og tverrinstitusjonelt team som sto bak evalueringen. Evalueringen bygde på omfattende dokumentstudier i form av offentlige dokumenter, møtoreferater. Det ble gjennomført to tidlige intervju-seminarer sentralt med veiledningsgruppen og med sekretærene fra alle byene. Etter dette ble det utarbeidet intervjuguide og forberedt case-studier i Bergen, Stavanger og Tromsø, basert på pladokumenter, møtoreferater og annen offentlig informasjon. Det ble utarbeidet en spørreskjemaundersøkelse rettet mot 233 deltakere i prosessene i alle 10 byene. I tillegg til lokale prosesser og TP 10 planer ble det gjennomført en beskrivelse av lokale vedtak og oppfølging i kommune og statlige etater. Det endelige rapportutkastet ble drøftet med oppdragsgiver Sd og gjennom et oppsummerende seminar med veiledningsgruppen.

Det er slik en bredt anlagt evaluering, men med en viktig metodisk utfordring som vi skal komme tilbake til nedenfor, manglende effektvariable stiller en overfor bestemte tolkningsutfordringer. Det oppstod også tolkningsutfordringer knyttet til ulike vurderinger mellom oppdragsgiver og evaluere. I den delen av evalueringen som omhandler bakgrunn og forventninger til TP 10, kom det fram informasjon om uklare mål og til dels også målkonflikter på statlig side, særlig mellom Samferdselsdepartementet (Sd) og Miljøverndepartementet. I den forbindelse ble det også, ifølge TØIs deltaker i utredningsteamet, uttrykt ønske fra oppdragsgiver om å omformulere og nedtone disse beskrivelsen i et omfang som evalueringsgruppen var uenige i. Når det var uproblematisk for utredningsteamet å ikke føye seg etter slike vurderinger, hadde det blant annet sammenheng med at teamet både var faglig sterkt, og at utredningsleder var fra universitetet og dermed stod helt fritt i forhold til oppdragsgiver

4.2.5 Hovedkonklusjoner

- Statlig initiativ var nødvendig for TP 10 som planleggingsreform, og rammebetingelser i form av økonomiske midler er avgjørende.
- Statens pådriverrolle og styring var moderat, TP10 var en desentralisert prosess med stor lokal fristilling, der staten har begrenset seg til faglig veiledning, tildeling av økonomiske ressurser til planlegging og tidsmessige rammer. De var uklare forventninger til hvordan de lokale planene skulle skje, bortsett fra at de skulle være input til Norsk veg- og vegtrafikkplan (NVVP) 1994-97. Det siste forholdet medførte sterk vektlegging av veibygging.

Når det gjaldt de institusjonelle og forvaltningsmessige forholdene, så var planarbeidet dominert av vegkontorene og byplankontorene, mens fylkeskommunene spilte en mindre sentral rolle. Sivilingeniører og ingeniører var de dominerende profesjonene. Disse vurderingene baserte seg i stor grad på spørreskjemaer der respondentene kom med sine vurderinger om hvem som hadde spilt den viktigste rollen.

Ser en på planene som kom som resultat av dette transportplanarbeidet, så skilte en mellom tre hovedstrategier: "Trend", "Kollektiv", som representerer en styrking av kollektivtransporten, og "Miljø" som også innebar tiltak for beregninger av biltrafikken. Arealbruksstrategier ble i liten grad konkretisert, sammenlignet med transportstrategiene, og ble lite synlig i avsluttende TP 10-dokumenter. Byene gjennomførte konsekvensanalyser i forhold til overordnede lokalt definerte mål. Selv om varianter av Miljø og Kollektiv ga best måloppnåelse, stod disse strategiene i motsetning til lokale politiske/administrative preferanser om å satse på hovedveibygging. I planene ble veibygging gitt større prioritet enn kollektivtransport og utnyttelse av areal. Utredningsgruppen konkluderer dermed med at intensjonen om at TP 10 som samordnet planlegging skulle føre til kursendring, slo feil.

4.2.6 Oppsummering med bakgrunn i analytisk skjema

TP 10-evalueringen var en prosessanalyse, der forfatterne undersøkte de sosiale prosessene der selve tiltaket ble utført, der ulike aktører fortolket, operasjonaliserte og gav retning til en sentralt initiert prosess. Begrunnelsen for denne typen evaluering er at selve organiseringen av TP 10-prosessen, hvem som har innflytelse på den og hvem som ikke har det, trolig påvirker utfallet.

Til forskjell fra Friskbuss-evalueringen mangler det i denne undersøkelsen data på effekter; hvordan TP 10 påvirket transportsystemet i retning av reduserte kø- og miljøproblemer som var formålet. Den metodiske løsningen på dette var dels å forholde seg til vedtatte planer og innspill til NVVP, og dels å evaluere slike planer med bakgrunn i kriteriene som ble formulert som grunnlag for TP10. Dette var også i samsvar med hvordan evalueringsmandatet ble formulert.

De sentrale begrepene i det teoretiske skjemaet er forhold som aktører, design, kjennetegn ved lokale prosesser i form av mål, organisering og prosesser, produkter, planer og effekter. Slike begreper har ingen klar disiplinforankring. De vil normalt kunne benyttes av forskere med bakgrunn i flere samfunns- og organisasjonsfag. Det går imidlertid fram av selve undersøkelsen og sammensetning av teamet at det er et sterkt innslag av statsvitere og planleggere. Det er uklart hvilket gyldighetsområde denne prosessevalueringen antas å ha. Den er i utgangspunktet formulert slik at den avgrenser seg til å undersøke akkurat dette tiltaket. Det er et framtreddende eksempel på det vi har omtalt som en substansiell evaluering. Den henter sin begrunnelse ikke i en formalisert modell, men i kunnskap om feltet. Den er i beste fall kumulativ i forhold til tilsvarende undersøkelser i disse byområdene.

Undersøkelsen er rent kvalitativ (med unntak av et spørreskjema som det gjøres bivariate analyser av). Den tilfredsstillende dermed ikke kravene til statistisk validitet. Når det gjelder teoretisk validitet, er denne vanskelig å vurdere fordi de teoretisk informerte drøftingene er lite eksplisitt formulert. Tilsvarende gjelder den interne validiteten. Denne illustrerer imidlertid et bestemt metodisk poeng.

Enkelt formulert er en konklusjon i denne evalueringen at de lokale vedtakene ikke er i samsvar verken med intensjonen til TP 10 eller for så vidt overordnede målsetninger lokalt. Den har altså ikke hatt tilsiktede effekter. Da blir neste spørsmål: Hvorfor ikke? Svaret som gis er at dette må knyttes til forhold i omgivelsene, nærmere bestemt NVVP som utløser ressurser dersom en vinner fram med sine prosjekter. Men dette er ikke en variabel i denne undersøkelsen, noe som innebærer at en ikke kan undersøke om dette er en viktig forklaringsfaktor. Validiteten til et slikt argument synes å basere seg på informantenes vurderinger, og dermed validering gjennom tilbakemelding fra informanter og offentliggjøring av rapporter sentralt. Det mer overordnede metodiske poenget er imidlertid at tolkingen i evalueringen er at når x ikke fører til y her, så er det fordi x (sentralt initiativ) er en nødvendig, men utilstrekkelig forutsetning for å føre til y (endrede planer). Dersom så er tilfelle, er det implisitte svaret at x + z (andre endringer i rammebetingelser), vil føre til y. "Svaret" blir dermed ikke at tiltaket – TP 10 – ikke hadde ønsket effekt og at en dermed kunne oppsummere at dette ikke var en god strategi, men snarere at tiltaket hadde feil utforming.

Når det gjelder ekstern validitet, er det ingen formuleringer i evalueringen som indikerer at forfatterne mener at funnene kan generaliseres utover de ti byene i Norge. Det er imidlertid blitt lagt til grunn i senere utredninger at for disse byområdene, og ved videreføring av disse trekkene ved omgivelsene, kan en anta samme effekter. Så utredningen er blitt oppfattet som mulig å generalisere i tid – om ikke i rom.

4.3. Manglende effektvariabler og ekstern validitet som metodisk utfordring

Det er noen generelle problemer knyttet til kombinasjonen av manglende effektvariabler og manglende ekstern validitet i organisasjonsevalueringer som TP 10-prosessevalueringen gir grunnlag for å diskutere. La oss innledningsvis slå fast at denne typen problemer ikke innebærer at konklusjonene i evalueringen er feil, ei heller at disse konklusjonene ikke er plausible og kan få konsekvenser for forståelsen av problemene en har evaluert. TP 10-evalueringen formulerte en generell konklusjon som skulle få virkning utover akkurat denne evalueringen: At selve de institusjonelle forholdene innen sektoren innebærer en makt-, rolle- og ressursfordeling som fører til systematisk opp-prioritering av veiinvesteringer og nedprioritering av andre tiltak. Denne forståelsen ga grunnlag for og ble reproduert i nye arbeid på feltet (Spangen 1995, Hompland 2001). Ti år etter TP 10 ble det satt i gang en utredning om organisatoriske hindringer for styrket kollektivtransport i byområder (Osland og Kråkenes, 1998 og 2000), der de to professorene som var med i TP 10-arbeidet, satt i referansegruppen. Disse utredningene la grunnlaget for forsøk med alternativ forvaltningsorganisering i by (2003-2007).

Denne forståelsen av trekk ved institusjonelle forhold som hindrer realisering av målene om økte kollektivandeler, er blitt reproduert under senere utredninger på feltet. I undersøkelse av organisering av transportpakker i norske byområder (Bekken og Osland, 2004), er nettopp rammebetingelsene i form av insentiver for infrastrukturensatsing og da særlig veiinfrastruktur vektlagt. Denne insentivstrukturen er koblet til en bestemt beslutningsstruktur. En vesentlig side ved organiseringen av beslutningsprosessene bak slike pakker er at de gir bestemte lokale aktører vetorett i prosessen, de besitter vetopunkter. Vetopunktene oppstår fordi det er krav om bred støtte lokalt, og dette representerer en barriere for å få innført tiltak som isse er med på å hindre for eksempel mer restriktive tiltak. Nå er forfatterne også i dette tilfellet forsiktige med å formulere at det er en slik klar sammenheng mellom organisering og effekter, men det svekker ikke grunnlaget for vårt spørsmål her: Hva er det empiriske grunnlaget for en slik vurdering? Denne baserer seg på 1) en sammenfatning av formelle bestemmelser som fastlegger prosedyrene og observasjoner om koalisjoner som foreligger lokalt, og 2) validering gjennom informantene. Men de har strengt tatt ikke gitt empirisk støtte for at endring av en slik barriere ville ha effekt på beslutninger.

En mulig strategi for å kartlegge den empiriske støtten for slike vurderinger, er en sammenlignende organisasjons- og effektanalyse, der en sammenstiller organisatoriske kjennetegn og effekter på beslutninger. I et senere arbeid (Osland

og Leiren, 2007) basert på innføringen av rushtidsavgift i Stockholm, var det mulig å gjøre en slik sammenligning. Denne sammenligningen gav støtte til hypotesen om at kravet om bred lokal konsensus var en faktor som hindrer denne typen restriktive tiltak. Dette illustrerer betydningen av ekstern validitet, men også hvordan det i denne type undersøkelser, ikke alltid foreligger grunnlag for å drøfte den eksterne validiteten. (Som er noe annet enn at en ikke drøfter at en ikke drøfter det.) En annen strategi enn å søke å klarlegge ekstern validitet gjennom undersøkelse av samme type substansområde ("barrierer mot innføring av restriktive tiltak i bytransportpolikk") vil være å knytte seg mer mot den allmenne litteraturen om temaet. I dette tilfellet kunne en knyttet diskusjonen til teorier om vetoforhandlinger, for eksempel har Cameron (2000) benytter rational choice-teori i sin undersøkelse av "veto forhandlinger" mellom to institusjoner som deler makt.

Manglende data om effektvariabler, innebærer også at vi ikke har de dataene vi trenger for å vurdere om fjerning av en organisatorisk barriere er en nødvendig og/eller en tilstrekkelig forutsetning for å få til ønskede effekter, og ønskede effekter vil i de fleste tilfeller si å realisere målsetningene for den offentlige politikken. Poenget kan illustreres med noen funn fra en annen undersøkelse som omhandlet effekten av organisatoriske endringer – det såkalte Reorient-prosjektet. Reorient-prosjektet var et omfattende EU-prosjekt som undersøkte betingelsene for sømløs godstransport i en korridor som inkluderte Norge, Sverige, Finland, Polen, Østerrike, Ungarn, Slovakia, Tsjekkia, Hellas, Romania, Bulgaria, Slovakia. Formålet var å lokalisere og å fjerne barrierer som kunne hindre effektiv godstransport med tog, for å styrke godstogenes konkurransekraft mot godstransport på vei. Slike barrierer kan være av ulik karakter. De kan være tekniske, organisatoriske, psykologiske, og de kan være knyttet til ulike typer institusjoner og organisasjoner; selskapene, nasjonale myndigheter (som lovgiver, ansvarlig for infrastruktur, eier av selskap, grensekontrollører) og overnasjonale myndigheter som ulike deler av EU-organisasjonen.

Som evaluering betraktet var et tema for Reorient barrierer på ulike nivåer. Vi skal rette oppmerksomheten mot en side av denne evalueringen, særlig knyttet til forholdet mellom EU-reguleringer, institusjonelle endringer på nasjonalt nivå (som skille mellom infrastruktur og drift ved oppdeling av de gamle nasjonale selskapene i en infrastrukturholder (som Jernbaneverket) og to operatørselskap, et for persontransport og et for gods, bestemmelser som sikrer like konkurransevilkår), og markedsutvikling i form av økt konkurranse mellom operatører på godssiden, og økte markedsandeler i forhold til godstransport på vei.

Den antatte kausale forbindelsen var knyttet til forståelsen til lovgiver (EU). Institusjonelle endringer på nasjonalt nivå, skulle føre til intramodal konkurranse som igjen ville føre til en bedret intermodal konkurransesituasjon. Resultatet ville altså bli bedret konkurransekraft mot godstransport på vei.

Grunnlaget for vurdering av status når det gjaldt implementering av EU-direktiv og lokalisering av barrierer var dels en sammenfatning av data om formell organisasjonsstruktur, dels evalueringer som allerede var foretatt på dette området, og ikke minst intervjuer med representanter fra alle relevante organisasjoner i de ulike landene. Samlet ga dette en oversikt over hvilke forhold de var enige i var et problem, og hvilke barrierer som var omstridt.

Analysen synliggjorde noen interessante paradokser (Ludvigsen og Osland 2008 a og b). Det var en klar forskjell på korridorlandene når det gjaldt implementering av EU-direktiv som skulle sikre lik markedsadgang ved etablering, formelt og reelt skille mellom infrastruktur og drift, transparent system når det gjaldt lisensiering, sporveisavgifter osv. På dette området var de nordiske landene kommet lengst, mens det stod mye igjen i flere av de sentral- og østeuropeiske landene, ikke minst når det gjaldt formelle og uformelle koblinger mellom aktørene som tidligere hadde vært i et integrert selskap som var både infrastrukturholder og godstransportør. I de sistnevnte landene innebar koblingen mellom infrastrukturholder og godstransportør at nye aktører ikke kunne være trygge på å ha like konkurransevilkår, og på noen områder var praksisen åpenbart diskriminerende.

Men dersom en så på konkurransesituasjonen, målt i antallet nye operatører, var bildet helt annerledes. Og dersom en så disse i sammenheng ble det første paradokset klart, det var ingen klar sammenheng mellom graden av implementering av EU-lovgivning for å fremme konkurranse og graden av konkurranse.

Tabell 3. Completeness of market deregulation and occurrence of intra-modal competition in the countries analysed (Ludvigsen og Osland, 2008)

		Deregulated Market Governance	
		Complete	Incomplete
Intra-rail competition	Yes	?	Poland, Romania, The Czech Republic,
	No	Norway, Finland, Sweden	Greece, Hungary, Austria

Kilde: TØI-rapport 929/2007

Noen av de landene som ikke hadde gjennomført mest omfattende endringer, Polen, Romania og Tsjekkia, hadde størst grad av konkurranse på sporet.

Dersom en derimot så på konkurranse med godstransport på vei, så var dette kun tilfelle i de tre nordiske landene, der de gamle monopolistene fortsatt hadde en dominerende posisjon. Det var her en hadde etablert intermodale løsninger som innebar at en konkurrerte med markedssegmenter som de en har på veisiden.

Tabell 4. Occurrence of intra-modal and intermodal competition in the countries analysed (Ludvigsen og Osland, 2008)

Road-rail competition			
		Yes	No
Intra-rail competition	Yes	?	Poland, Bulgaria, Romania
	No	Norway, Finland, Sweden	Austria, Hungary, Greece

Kilde: TØI-rapport 929/2007

Med andre ord holdt ikke de antagelsene om årsaksforbindelser som lå bak lovgivningen stikk. En komplett implementering av EU-lovgivning var ikke en nødvendig betingelse for økt konkurranse på sporet, det måtte være noen av elementene i omreguleringen måtte være tilstrekkelig. På den andre siden var det de landene som hadde implementert EU-reguleringene som hadde fått til konkurranse mellom godstransport på jernbane og godstransport på vei. Denne nyanseringen av effekten av ulike organisatoriske variabler ville ikke vært mulig uten å ha data på effektvariabelen; konkurransesituasjon i jernbane og mellom jernbane og veg.

4.4. Oppsummering

I dette kapitlet har vi sett på ulike typer organisasjonsevalueringer. En av disse, evalueringen av "Friskbuss", var kjennetegnet av en relativt klart fastlagt avhengig variabel, som en hadde datagrunnlag for å måle før og etter tiltaket, og visse muligheter for kontroll. Dermed hadde en et datagrunnlag for å kunne si om organisatoriske tiltak hadde effekter. Denne undersøkelsen tilfredstilte i stor grad kravene til teoretisk, metodisk, intern og ekstern validitet. Den dokumenterer også betydningen av organisasjonsvariable, slik Aalborg-undersøkelsen i forrige kapittel gjorde.

Når en ser på TP 10-evalueringen, hadde en der ikke diakrone data på effektvariabelen, og denne evalueringen tilfredstilte i liten grad kravene til både teoretisk, statistisk, intern og ekstern validitet.

Vi skal her løfte fram tre problemer som dels følger av disse svakhetene.

- Kvalitative data og ideologi/interesseinfiserte tolkinger. Kvalitative analyser er i en annen og til dels i større grad sårbare for manipulasjon enn kvantitative. En sammenstilling av tallmateriale som fører til en bestemt fordeling vil være preget av skjønnsvurderinger. Sentralt er reliabilitet i innsamling og registrering av data, og oppsettet av modellverktøyet. Kvalitetskontrollen ligger ekstern kontroll og diskusjon med fagfeller på utvikling av modellverktøy og økt nøyaktighet i innsamling og lagring av data, som vi pekte på under tilfellet nyttekostnadsanalyse. Men i større grad enn i kvalitative analyser er selve analysen da i stor grad en teknisk prosess, før en tolker resultatene. Når en foretar kvalitative analyser,

derimot, er analysen også en tolkning, et utvalg av systematiske trekk ved det en undersøker. Da vil nettopp sammenfallende verdier eller vurderinger kunne påvirke materialet i en annen grad. Dette er et kritisk punkt, ikke minst ved framvekst av epistemiske "communities" på tvers av forvaltning og evaluering. Det er viktig med tiltak for å avbøte dette. Et slikt tiltak som TP 10-evalueringen dokumenterer, er å sikre evaluereers autonomi.

- Problemet med manglende effektanalyse. Vår vurdering av TP 10-prosessevalueringen, er at kontrafaktisitetproblemet i dette tilfellet ikke er løst. Denne vurderingen fra vår side impliserer ikke at evaluereene ikke gjennomførte en evaluering i samsvar med det de sa de skulle gjøre, ei heller at det nødvendige datamaterialet kunne vært stilt til rådvelde. Tvert om virker antagelsen om at eventuelle effekter måtte sees over et lengre tidsrom, og at en kunne foreta en evaluering av om en observerte endringer i vedtak, rimelig. Derimot innebærer vår vurdering at mangel på diakrone data for effektvariable, bør implisere at forklaringer på mangelfulle effekter får status som hypoteser. Den neste videreføringen blir å søke å gjennomføre en type analyse som muliggjør å undersøke om en slik hypotese har støtte. Eventuelt måtte den bli utviklet til en internasjonal, komparativ undersøkelse for å kartlegge om funnet er valid.
- Problemet med ekspanderende policy-nettverk og -områder. La oss enkelt skille mellom to konklusjoner som følge av at en evaluering viser at et tiltak ikke har ønskede effekter. 1) Siden det ikke har effekt, foreligger det ikke grunner for ytterligere intervensjon av den type. 2) Tiltaket hadde ikke effekt, men der er fordi det var for svakt dosert, en trenger mer, ikke mindre av medisinen. La oss anta at det er etablert et politikk-område som blir holdt oppe av et kunnskapssamfunn som har verdimeslige og økonomiske interesser knyttet til kontinuerlig problemløsning om et fenomen (det kan være "kollektivtransport", "fattigdom", "innvandrere"). De vil da favorisere konklusjon 2. I noen tilfeller vil dette kunne medføre en utvikling der en utvikler et mer raffinert tiltaksapparat – i andre tilfeller kan det medføre kontinuerlig oppmerksomhet mot et knippe med tiltak overfor et problem som ikke lar seg løse, en faglig-politisk-administrativ evighetsmaskin. Poenget her er ikke at denne type unyttig eller feilslått aktivitet bare er et problem knyttet til organisasjonsevalueringer eller undersøkelser basert på kvalitative data. Tvert om har jo blant annet Kringlen, 2008, hevdet at den omfattende økningen av antidepressive midler introdusert fra 1990-tallet, begrunnet med kvantitative undersøkelser, har vært basert på feilinformasjon og båret fram av et epistemisk samfunn i skjæringsfeltet mellom legemiddelindustri og leger - i Norge finansiert gjennom omfattende offentlige midler. Men poenget her er at når Kringlen reiser denne debatten i en situasjon der det foreligger kvantitative analyser som inkluderer effektvariable, så har han et utgangspunkt for en slik debatt i form av analyser som er etterprøvbare. Alternativt kan han peke på at de data som skulle begrunne intervensjonen, ikke er etterprøvbare. Kontrasten til kvalitative evalueringer som ikke en gang foregir å ha data på effektvariabelen er tydelig.

5. Oppsummering og avsluttende diskusjon

Formålet med denne rapporten er å kartlegge og sammenfatte ulike evalueringer innen samferdselssektoren, med det formålet å sammenligne og vurdere gyldigheten av ulike evalueringskriterier og ulike evalueringer. Sentrale spørsmål var dermed:

- Hvilke kvalitetskriterier tas i bruk?
- Hvilke likheter og forskjeller kan vi finne på tvers av evalueringstradisjoner og tema?
- Hvilke særlige utfordringer og dilemmaer står en overfor, når en skal realisere disse kriteriene?

Vi avgrenset oss til evalueringer som tar utgangspunkt i en målrasjonell tenking, dvs. som gjelder effektene av et tiltak på et bestemt fenomen utenfor forvaltningen. Vi avgrenset oss videre til etterundersøkelser som har et metodisk design for å måle eller vurdere endring som følge av tiltaket, men inkluderer prosess- eller organisasjonsanalyse som forholder seg til et slikt imperativ.

I det følgende skal vi oppsummere ulike tradisjoner, for å se i hvilken grad vi kan spesifisere hvilke typer kvalitetskrav som er operative, og søke å formulere hypoteser om hvordan disse kommer under press eller er underutviklet i bestemte tradisjoner og relasjoner.

Blant de sentrale temaene som vil bli tatt opp i avslutningskapitlet, er hvilke likheter og forskjeller vi kan finne på tvers av forskningstradisjoner og tema, og hvilke særlige utfordringer og dilemmaer en står overfor når en skal realisere kvalitetskriterier i ulike fagtradisjoner.

For dette formålet har vi, basert på de fire perspektivene på evaluering som vi har skissert ovenfor, laget en enkel modell som skisserer et knippe av dimensjoner eller variabler som kan utgjøre en matrise for sammenfatningen av ulike evalueringer og evalueringstradisjoner.

- Hva er *temaet, problemstillingen eller formålet* med evalueringen?
- Hva kjennetegner *metode og datagrunnlag* (kvalitative, kvantitative data, eventuelt type underkategori)? Hvor omfattende er undersøkelsen, hvor mange variabler og typer variabler har den?
- Hva kjennetegner *teorien* som blir benyttet, hvilke *forutsetninger* bygger den på, hvilke *antagelser* søker den å undersøke, hva er de sentrale *begrepene*, har den en klar *disiplinforankring*, hva er dens *gyldighetsområde*?
- Er mulige/sannsynlige utilsiktede effekter godt dekket?
- Kumulativitet: Hvordan antas evalueringen å kunne føre til ny kunnskap (substansielt/metodisk/teoretisk)?

- **Kvalitetskriterier:** Hvilke kvalitetskriterier tar de ulike evalueringene i bruk? Utgangspunktet vårt, målestokken for en slik vurdering, er Elvik sine kriterier vedrørende *statistisk validitet, teoretisk validitet, intern og ekstern validitet*. I den grad de ulike evalueringene avviker fra Elviks liste, hva er begrunnelsene? I tillegg til disse evalueringsspesifikke kriteriene er selvsagt også de generelle forutsetninger knyttet til minimumskrav som stilles i forskning; logisk konsistens, etterrettelighet, etterprøvbarehet og generaliserbarhet

Tabell 5. Oversikt over prosjekter og deres kjennetegn

Prosjekt	Disiplinforankring	Tema/formål	Sentrale begreper	Metode	Type
Tiltakspakkene for kollektivtransport	Interdisiplinær	Undersøke endret reisemiddelvalg og reiseomfang på grunn av endret kollektivtransporttilbud	Holdninger, oppfatninger og kunnskap, preferanser, personlige kjennetegn, egenskaper ved tilgjengelige alternativer, strategi for valg, situasjon og rammebetingelser, faktisk atferd	Kvantitativ og kvalitativ: Reisevaneundersøkelser, reisemiddelfordeling, panel- og brukerundersøkelser	Multidimensjonal variabelanalyse, etterspørselsanalyse
Expost nyttekostnadsanalyse	Økonomi	Forbedre og utvikle metode-verktøyet	Prosjektgevinster, og kostnader, eksternaliteter, kalkulasjonspriser, avvik	Nyttekostnadsanalyse, verdsetting	Ex post Nyttekostnadsanalyse
Flyvbjerg et al 2003	Uklar/interdisiplinær	Årsaker til budsjettoverskridelser og lite treffsikre trafikkprognoser	Tekniske, psykologiske, politiske og økonomiske forklaringer	Kvantitativ og kvalitativ	Multidimensjonal variabelanalyse
Friskbuss	Sosiologi/interdisiplinær	Team-organiseringens innvirkning på sykefravær i bussbransjen	Sykefravær, teamorganisering, trivsel	Intervjuer og spørreundersøkelser, sykefraværs-statistikk	Organisasjons-evaluering med effektvariable
TP 10	Statsvitenskap	Hvordan organisering av lokale prosesser påvirker utfallet av samordnet areal- og transportpolitikk i norske byområder	Institusjonelle og forvaltningsmessige forhold (forvaltningsnivåer, sektorinteresser, miljøhensyn i planer)	Intervjuer, seminarer, spørreskjemaer	Organisasjons-evaluering
Reorient	Interdisiplinær	Effekter av organisatorisk og institusjonell reform i Europa innen jernbanens godsmarked på konkurranse og markedsandeler	Tekniske, organisatoriske, psykologiske og institusjonelle	Kvantitativ og kvalitativ	Organisasjonsanalyse med effektvariable

Kilde: TØI-rapport 929/2007

Vi skal i det følgende summere opp trekk ved de evalueringer vi har gjennomgått med bakgrunn i dette skjemaet. Et formål med en slik gjennomgang er å sammenfatte om eller i hvilken grad de ulike evalueringene tilfredsstillende de ulike kravene. Videre søker vi å lokalisere det vi kaller ”kritiske punkter”, dvs. punkter der de ulike evalueringene og de tradisjoner de kan sees som uttrykk for, står overfor særlige utfordringer for å sikre kvalitet ved evalueringer.

5.1. Oppsummering av kjennetegn ved ulike evalueringer

5.1.1 Tiltakspakkene – evaluering som multidimensjonal variabelanalyse

Kapittel 2 tar for seg evaluering av tiltakspakker for bedret kollektivtransport. Formålet med disse pakkene var altså å bedre kollektivtilbudet, og å få flere til å reise kollektivt. Evaluering av tiltakspakkene illustrerer en bestemt problemstilling som ikke vil være uvanlig ved evalueringer, det at tiltakene som blir valgt ut er gjort av i alt overveiende grad andre hensyn enn de rent faglige, noe som innebærer at de tiltak som er fremmet i varierende grad er teoretisk begrunnet. Selve designet innebærer forsøk på å utvikle mer systematisk kunnskap trass i denne type begrensninger i utvalg. Løsningen som er benyttet er en teoretisk modell som spesifiserer plausible sammenhenger mellom variable og et forskningsdesign med data fra før og etter tiltaket ble gjennomført. De metodiske utfordringene er knyttet til denne spenningen mellom politisk-administrativt vedtatte tiltak, og lokalt innsamlede data på den ene siden, og på den andre siden et analytisk design med ønske om å framskaffe mer systematisk kunnskap om hvilke tiltak som påvirker reisevaner, basert på aggregerte analyser av de ulike tiltakene og deres effekter. Evalueringsdesignet lar seg i all hovedsak analysere med bakgrunn i de kvalitetskriterier vi har skissert innledningsvis.

De **sentrale begrepene** som konstituerer de variable som blir benyttet i dette designet er: Holdninger, oppfatninger/kunnskap, preferanser, personlige kjennetegn, egenskaper ved tilgjengelige alternativer, strategi for valg, situasjon/rammebetingelser, faktisk atferd.

Evalueringsdesignet hviler dermed på den **forutsetning** at dette er adekvate begreper for å forklare reisemiddelvalg, mens de **antagelser** som gjøres i form av formulerte hypoteser, er at det er bestemte sammenhenger mellom disse knippene av variabler. Disse er formulerte som antatte kausalsammenhenger, men det er ikke spesifisert nærmere hvilken relativ betydning de antas å ha. Dette er altså et design som er rettet mot trafikantene, mens strukturelle variable som tilbudet, bystruktur, bosettingsmønster, inngår i mer generelle kategorier, som rammebetingelser.

Disse variablene og måten de er begrepsfestet på, kan ikke sies å ha en bestemt **disiplinforankring**. Det er variabler på et relativt lavt abstraksjonsnivå, som kan knyttet til flere ulike mer omfattende teoretiske posisjoner, særlig innen sosiologi og sosialpsykologi. De er altså interdisiplinært forankret.

Evalueringsdesignet lar seg i all hovedsak analysere med bakgrunn i de kvalitetskriteriene vi har skissert innledningsvis.

Statistisk validitet: Rapportene drøfter reliabilitetsproblemet med noen typer data, mest klart ved passasjertellinger. Men når passasjertall er lite reliable, og en i stedet etablerer opplevd reisehyppighet som indikator for avhengig variabel, er dette et mindre valid mål, noe som får konsekvenser for interne og teoretisk validitet (se under). Evalueringen er generelt preget av et relativt omfattende datamateriale, som i utgangspunktet åpner for kontroll av reliabilitet ved ulike

målinger. Minst klart formulert er betingelser knyttet til den aggregerte analysen, og koblinger mellom devalueringer og mer samlede evalueringer.

Intern validitet: Vurderingen av denne evalueringen lar seg i all hovedsak gjennomføre med bakgrunn i de kriterier vi presenterte om intern validitet. Problemet er imidlertid at når en opererer med opplevd endret tilbud og opplevd endret reisehyppighet som indikatorer, så svekker dette grunnlaget for å anta at en snakker om en årsakssammenheng og ikke bare en korrelasjon. Videre kan årsakssammenhengene mellom kunnskap om kollektivtilbudet og bruk av det like gjerne være slik at bruk skaper kunnskap som omvendt.

Slike svakheter ved intern validitet svekker den teoretiske validiteten. Et annet trekk ved teoretisk validitet, er at selv om den teoretiske modellen i utgangspunktet var formulert slik at sammenhengen mellom variablene (holdninger, alternativer, preferanser, rammebetingelser) er beskrevet som årsakssammenhenger, er disse underspesifisert som hypoteser. Dette innebærer at selv om operasjonalisering og datagrunnlag hadde vært preget av mindre svakheter, må en svare "både og" på spørsmålet om teorien som er formulert kan forklare hvordan det undersøkte tiltaket kan virke på det problem det er ment å løse. En undersøker og søker å sammenfatte noen isolerte effekter av variable innen modellen og beskriver forklart varians av disse variablene, men denne analysen er svekker av de innvendinger vi har reist ovenfor. Uansett ville ikke den generelle, overordnede modellen som introduseres forut for designet, vært testet.

Ekstern validitet lar seg i prinsippet gjennomføre med bakgrunn i dette designet. Det er imidlertid ikke gjort systematiske sammenligninger med andre analyser av andre byområder i disse rapportene

Gyldighetsområdet er ikke videre spesifisert. Gitt at den teoretiske modellen ikke er videre spesifisert i forhold til sammenhenger mellom variable, er det grunn til å anta at den dekker så vidt mange variable at forfatterne antar den representerer en relativt uttømmende liste med variable for denne type undersøkelse, dvs. kollektivtransport i situasjoner med konkurranseflater med personbil. Ekstern validitet for denne type analyser lar seg i prinsippet gjennomføre, blant annet er de sentrale variablene vi finner her de samme som i UITP-databasen. Det er imidlertid et framtrødende trekk at den eksterne validiteten ikke er undersøkt. Det er altså ikke gjort forsøk på å diskutere om funnene er i samsvar med tilsvarende undersøkelser, men antas å kunne generaliseres til andre byområder.

Når det gjelder kumulativitet, er evalueringsarbeidet preget av at det til en viss grad brukes tidligere utviklet kunnskap, blant annet om priselastisitet. Sentralt er det imidlertid at selve evalueringsdesignet er utviklet for å kunne få mer systematisk kunnskap om hvordan forskjellige endringer i kollektivsystemet påvirker folks holdninger, oppfatninger og reisevaner. Selv om tiltakene er mangeartede og til dels i varierende grad faglig fundert, er det etablert et design som skal kunne ivareta generelle konklusjoner. Resultatene fra disse evalueringene har også vært vesentlige som opparbeiding av kunnskapsgrunnlaget om kollektivtransport i Norge.

Oppsummert: Dette er en type evaluering som mer allment kan karakteriseres som flerdimensjonal variabelanalyse. Den omfatter variable på ulike dimensjoner; reisevaner, tiltak, rammebetingelser. Modellen er i liten grad spesifisert i form av

antatte variables forklaringskraft, og den har en tverrdisiplinær forankring. De betyr også at forutsetninger – for eksempel aktørforutsetninger – i liten grad er spesifisert. Ett kritisk punkt med denne type analyse er manglende transparens mht til selve analysen. I denne undersøkelsen er den eksterne validiteten i liten grad drøftet.

5.1.2 Ex post nyttekostnadsanalyse: Evaluering som systematisk læring – med ett kritisk unntak

I kapittel 3 gikk vi gjennom ex post nyttekostnadsanalyse av store samferdselsinvesteringer. Denne type analyse er klart disiplinforankret, den er forankret i velferdsøkonomisk teori, der formålet er å undersøke om et tiltak gir et positivt eller negativt samfunnsøkonomisk resultat. Dette skjer ved bla ved at en sammenstiller virkninger som så langt mulig tilstrebes gitt en felles pengeverdi.

Utgangspunktet for slike analyser er begrepene om konsument- og produsentoverskudd. Et prosjekt hvor den samlede neddiskonterte nytten (målt i kroneverdi) er større enn de samlede neddiskonterte kostnader (også i kroneverdi) kan sies å være samfunnsøkonomisk lønnsomt og representerer følgelig akseptabel ressursbruk for samfunnet.

En nyttekostnadsanalyse skal vise tiltakets effekt i forhold til ulike aktører i samfunnet. Her skiller det gjerne mellom noen hovedkategorier av interessenter: det offentlige, brukerne og samfunnet for øvrig. I samferdselssektoren framkommer ofte effektene for det offentlige i form av investeringskostnader, endringer i drifts- og vedlikeholdskostnader og offentlig kjøp av transporttjenester. Brukerne er her gjerne både operatører og trafikanter, som kan oppleve endret tidsbruk (reisetid, ventetid, gangtid) og endring i direkte økonomiske størrelser som driftskostnader og utlegg (som for eksempel billettkostnader). Samfunnet for øvrig uttrykker gjerne det som også kalles eksterne effekter, som for eksempel endringer i lokale, regionale og globale utslipp, støy og naturinngrep.

Gyldighetsområdet er nyttekostnadsanalyse av store samferdselsprosjekt. Evalueringstradisjonen er vesentlig av to grunner; dels fordi den angår evaluering av omfattende offentlig ressursbruk, dels fordi det er en dominerende evalueringstradisjon som ex ante undersøkelse. Ulikt mange andre sektorer har samferdselssektoren en lang tradisjon med nyttekostnadsanalyse forut for prosjekttildeling. Formålet med nyttekostnadsanalyse som etter-analyse er nettopp å forbedre ex ante metodikken, gjennom å kartlegge avvik mellom de beregnede samfunnsøkonomiske virkningene i forundersøkelsen og etterundersøkelsen og å lokalisere hvilke av disse som skyldes ulik metode, ytre forutsetninger, prosjektavhengige avvik.

Det teoretiske rammeverket er preget av at antagelser og forutsetninger er eksplisitt og klart formulert. Undersøkelsen tilfredsstillende kravene til et klart teorigrunnlag ved at begreper er operasjonaliserbare. Det mest kritiske punktet i operasjonaliseringsprosessen er usikkerhet ved operasjonalisering av komponenter som i utgangspunktet ikke er beskrevet i økonomiske termer: Tidskomponenter og eksterne kostnader som miljø- og ulykkeskostnader. Teorien gjør det også mulig å teste antatte sammenhenger mellom sentrale variable, men

selvsagt ikke den teoretiske forutsetningen at dette er de relevante variablene. I sum tilfredsstillende denne type evaluering godt kravene til teoretisk validitet.

Når det gjelder statistisk validitet, opererer denne type analyse med to typer statistiske sammenhenger, for det første de sammenhenger som uttrykkes gjennom selve Nyttekostnadsanalysen, som krever høy spesialisering og er lite transparent, og for det andre framstillingen av avvikstabeller som gir en oversikkelig avviksanalyse.

Selve det metodiske opplegget er designet for å ivareta de fleste sider ved intern validitet: Systematisk gjennomtenkning av årsaks-virkningsforhold. Ekstern validitet og generaliserbarhet er primært knyttet til utvikling av metodikken, dels som gjennomgang og forbedring av selve metodehåndbøkene og ensartet metodikk for vei og jernbane. Det er også i denne forstand denne type undersøkelser er kumulative, som utvikling av metodeverktøyet. De refererte evalueringene til Eriksen og Minken viser også på punkt etter punkt, klare feil og mangler ved de metodehåndbøker som har vært benyttet, og kommer med en rekke konkrete forslag til forbedringer slik at feil kan reduseres i fremtiden.

Oppsummert er dette en evalueringstradisjon preget av høy grad av formalisering, som imøtekommer de mest grunnleggende krav til evalueringsmetodikk som Elvik har formulert. Et sentralt formål hos Elvik er å forbedre metodikken. Formålet er å utvikle og forbedre metodeverktøyet slik at en gjør færre feil i framtidige nyttekostnadsanalyse.

Det er særlig en type prosjektgevinster og -kostnader som ikke blir empirisk etterprøvd gjennom et slikt opplegg, og det er usikkerheten knyttet til konvertering av tids, miljø og ulykkeskostnader til økonomiske kategorier.

Metodiske problemstillinger og alternativer; tidsverdsetting og flermålsanalyse

Spørsmålet om tidsverdsetting, og av verdsetting av kostnader knyttet til ulykker, støy osv er kontinuerlig behandlet innenfor den type fagmiljø som jobber med nyttekostnadsanalyse. Typisk i den sammenheng vil være å basere verdsettingen på individenes egne verdsettinger, eventuelt utvidet til kartlegging av ulike trafikantsegmenters verdsettinger.

Spørsmålet om tidsverdsetting er imidlertid også gjenstand for kritikk utenfra økonomiprofesjonen. Sverre Strand peker på at tidsverdien gjerne vil være den største verdien i mange nyttekostnadsanalyser, og peker på visse grunnlagsproblemstillinger knyttet til verdsetting av tid. Han skiller mellom tre trinn i tidsanalyser: 1) Måling av tid og tidsbesparelser, 2) aggregering av tidsbesparelser og 3) pengeverdering av tidsbesparelser.

Spørsmål 1 omhandler slike tema som hvor mye større enn null enkeltbesparelsen må være for at den skal kunne omtales som signifikant nok til at tiden kan anvendes alternativt. Problem 2 omhandler under hvilke forutsetninger tidsbesparelser kan aggregeres, mens det tredje punktet altså handler om konvertering av tidsverdier til pekuniære verdier. Strand hevder at alle tre er problematiske, men at standardtilnærmingen til problemet vil være at en gjennom tidsverdsetting løser 1 og 2.

Tilsvarende rettes det systematisk kritikk mot nyttekostnadsanalyse for en rekke andre antagelser som den bygger på, som aktørforutsetninger innen konsumentteori som konverterer de kritikerne oppfatter som moralske, kollektive problemstillinger til spørsmålet om individuelle preferanser, og problemer knyttet til bruk av neddiskonteringsraten på miljøspørsmål. Og også mot denne type modeller blir det rettet kritikk om manglende transparens.

Denne type spørsmål er relevante for nyttekostnadsanalyse per se, men neppe for forholdet mellom før- og etteranalyser, der noe av poenget er at en opererer med samme tidsverdsetting og verdsetting av eksternaliteter. Løsningen av denne type grunnlagsproblemer er derfor neppe av betydning for feilbudsjettering i den betydningen dette blir undersøkt gjennom etteranalyser. Eksempelvis: Gitt at en operer med en gitt tidsverdi vil det være faktoren antall trafikanter som er avgjørende, ikke tidsverdien en tilskriver disse trafikantene. Annerledes formulert: Den relativt permanente usikkerhet knyttet til verdsetting av tid i seg selv er et permanent problem ved Nyttekostnadsanalyse i denne sektoren, og i liten grad en forklaring på feilbudsjettering i konvensjonell forstand.

En måte å imøtekomme noe av problemet knyttet til pengevurdering av i utgangspunktet ikke økonomiske og til dels også kvalitative data som støy, ulykker og tidsbruk, er flermålsanalyse. Denne type analyse har som formål å assistere beslutningstakere som står overfor et problem der ulike verdier er i konflikt. Det overordnede siktemålet er å gjennomføre en systematisk alternativanalyse knyttet til ulike evalueringskriterier som uttrykkes på en objektiv og konsistent måte. De variablene som ikke entydig kan kvantifiseres rangordnes etter ikke-numeriske skalaer, men det forutsettes likevel at de ulike alternativene kan gies en score. Metoden framstår som et alternativ til nyttekostnadsanalyse i situasjoner med prioritering av mål som ikke er entydig kvantifiserbare, men begrenses ved at det primært framstår som en analyse for vurdering av et begrenset antall alternativer.

Oppsummert er ex post nyttekostnadsanalyse en evalueringstradisjon som imøtekommer de fleste krav som vi med Elviks modell kan stille til evalueringer, den har relativt klart spesifiserte forutsetninger og antagelser, den lar seg teste empirisk og opplegget er slik at det legger til rette for kontinuerlig forbedring. Den har det til felles med den evalueringstypen vi skisserte i forrige kapittel at den er lite transparent for andre enn de som er skolerte i metodikken. Men til forskjell fra den forrige er dette en metodikk knyttet til og dermed gjenstand for kritisk etterprøving innen en vitenskaplig disiplin. Det gjelder imidlertid for denne tradisjonen også at den er ikke falsifiserbar i følgende betydning: spørsmålet om de variable som benyttes er adekvate, lar seg ikke avkrefte empirisk. Sentralt for denne tradisjonen er de forutsetninger som gjøres om å konvertere sentrale ikke pekuniære og til dels også ikke-numeriske kategorier til pengeverdi.

5.1.3 Evaluering av mega-prosjekter som organisasjonsanalyse

Rasjonalet for ex post nytte kostnadsanalyse, er en etablert problemstilling innen sektoren: Mange investeringer viser seg i ettertid å være feilbudsjettert – og da oftest underbudsjettert. Relevansen av ex post nyttekostnadsanalyse bygger på en antagelse om at en viktig forklaring på slike feilbudsjetteringer er metodisk, at det er svakheter ved modellverktøyet. Grovt sagt kan vi skille mellom to andre

hovedforklaringer; uforutsette hendelser og organisatoriske forklaringer. Uforutsette hendelser kan en til en viss grad argumentere for også kan være metodisk, i det viser til manglende spesifisering av usikkerhet ved modellen (nyanser). Organisatoriske forklaringer tar utgangspunkt i at det er trekk ved måten selve den faglig-administrative utredningsprosessen er organisert på, som påvirker det faglige resultatet. Denne type forklaring har gitt opphav til en evalueringstradisjon som kan benevnes som komparative analyser av megaprojekter. Slike analyser inkluderer metodiske forklaringer på feilbudsjettering, men åpner også for en systematisk gjennomgang av de organisatoriske forholdene som kan forklare på feilbudsjettering.

I Norge har det få evalueringer som har søkt å inkludere organisatoriske eller institusjonelle variable i forklaringer på underbudsjettering ved større samferdselsprosjekter. Rett nok har slike variable vært fremhevet av Riksrevisjonen, både ved spesifikke veiutbyggingsprosjekt og bompengeprojekter, men det er ikke blitt systematisk sammenlignet med metodiske problemstillinger.

Blant de mer framtrepende forskningsmiljøene på dette området har Aalborg-miljøet, og særlig Bengt Flyvbjerg vært. Flyvbjerg forstår overskridelser som resultat av desinformasjon, og lokaliserer tre hovedmodeller for å forklare slik desinformasjon: Tekniske modeller som refererer til feil i data og metoder, psykologiske modeller som refererer til kognitive feil og politisk-økonomiske modeller i form av strategisk misrepresentasjon og løgn. Flyvbjerg argumenterer for at hans analyse viser at de to første modellene ikke kan forklare desinformasjon og feilbudsjettering, blant annet fordi en da skulle forvente læring og redusert feilbudsjettering over tid. Forklaringen blir da at insitamentstrukturen er slik at fordelene med å desinformere er større enn ulempene og resultatet blir "survival of the unfittest" – det er ikke det beste prosjektet som vinner men det som har best desinformasjon. Botemidlene på dette problemet er dels utvikling av bedre metoder for å estimere kostnader, nytte og risiko gjennom "reference class forecasting", dels endrede insitamentstrukturer som vil begrense desinformasjon og fremme riktig informasjon.

Teoretisk er tilnærmingen til Flyvbjerg eklektisk uten en klar disiplinforankring. Fordelen med en slik tilnærming er at den spenner opp et relativt bredt analytisk rammeverk som inkluderer mange relevante variable, i denne sammenheng både metodiske, psykologiske og organisatoriske. På den andre siden er aktørforutsetningene og antagelsene om institusjoners virkemåte noe underspesifiserte, men drøftingen indikerer at disse er forankret i institusjonell økonomi.

Det kritiske punktet i analysen synes å være den interne validiteten, hvordan sammenhengen mellom variablene fastlegges. Det er for eksempel få klare empiriske indikatorer på løgn, men denne forklaringsvariabelens kraft fastlegges gjennom et kontrafaktisk resonnement. Fordi de andre forklaringsmodellene ikke finner støtte gjenstår den siste. Det metodiske problemet Flyvbjerg står overfor er det klassiske for de som forsker innen tradisjoner som omfattes av mistankens hermeneutikk, de er vanskelig å finne empirisk støtte for at løgnere lyver, enten de vet de selv eller ikke.

Problemstillingen til Flyvbjerg er imidlertid svært relevant, fordi den åpner for en systematisk evaluering av feilbudsjettering ved store samferdselsinvesteringer som ikke begrenser seg til forbedring av metodikken og dermed vekt på ex post nyttekostnadsanalyse, men også på de organisatoriske og institusjonelle variable. En måte å videreføre denne type analyse på, ville være å undersøke om vi finner systematisk variasjon i feilbudsjettering i situasjoner der de institusjonelle forhold er endret i de retninger som Flyvbjerg har foreslått: Rammetilskudd til infrastruktur heller en øremerking, uavhengige "peer review" og "benchmarking" av alle prognoser, større offentlig åpenhet og innsikt i prognoser, og økt finansielt medansvar for mottaker av statlige midler. På flere av disse områdene kan en anta at en vil finne variasjon, også fordi noen av disse tiltakene etter hvert et blitt innført i noen land slik at deres forklaringskraft eventuelt kan falsifiseres.

På ett avgjørende punkt ser imidlertid gjennomgangen av ex post nytte kostnadsanalyse og usikkerhet knyttet til denne, og de komparative organisasjonsanalyser ut til å sammenfalle: Behovet for offentlig debatt og kritikk. Nyttekostnadsanalysen er preget av høyt presisjonsnivå i forhold til den nyttekostnadsbrøken som presenteres, noe som bør ha sitt tilsvarende i en åpen dialog om usikkerhet knyttet til forutsetninger for slik presisjon.

Men de skiller seg altså fra hverandre på det punktet at ex post nyttekostnadsanalyse begrenser seg til å undersøke og forbedre metodikken, mens metaprojektevalueringene inkluderer organisasjonsvariable og dermed behovet for omstrukturering av organisatoriske og institusjonelle omgivelser. De synliggjør også særegne problemer knyttet til organisasjonsanalyse, som blir ytterligere dokumentert i andre deler av denne rapporten.

Organisasjons- og prosessevalueringens svøpe – kontrafaktisitetens problemet

I kapittel tre gjennomgås tre evalueringer som kan rubriseres under vignetten organisasjons og prosessevaluering. Den første evalueringen – evalueringen av Friskbuss – er tatt med fordi den på mange måter representerer et evalueringsdesign som har hatt de ressurser og en slik varighet i tid at en kunne gjennomføre et systematisk, metodetrianglerende opplegg der en også hadde tidsseriedata. Caset var innføring av teamorganisering i bussbransjen for å få ned sykefraværet, både kvalitative og kvantitative data ble samlet inn på ulike tidspunkt: Kvantitative data ble samlet inn forut for etablering av forsøket, og kvalitative data underveis. I tillegg ble kvantitative data (identisk med forundersøkelsen) samlet inn til slutt. Den metodiske gjennomgangen dokumenterer hvordan et slikt evalueringsopplegg, der så vidt mange trekk ved omgivelsene endrer seg, aldri fullt ut kan tilfredsstille kravene til å dokumentere at et organisatorisk tiltak har effekter. Samtidig innebærer det at en har klare effektvariable (trivsel og sykefravær) at evalueringen har en klar målestokk, og der metodetriangleringen gir et grunnlag for, dersom alle data peker i samme retning, å trekke noenlunde robuste konklusjoner. Modellen er også så vidt klar at det er mulig å spesifisere hvilke variable en har kontroll på og manipulerer og hvilke en ikke kan påvirke.

Denne evalueringen tilfredsstillende de fleste krav til teoretisk, metodisk og intern validitet. I prinsippet skulle den også kunne tilfredsstillende krav til ekstern validitet, men slike spørsmål blir i liten grad drøftet.

Evalueringen av TP 10 som prosess, som tar for seg forsøk på samordnet areal- og transportpolitikk i ti større norske byområder, har status som en prosessevaluering. Den teoretiske modellen for evalueringen er lite formalisert, men avgrenser seg til å presentere en analytisk modell som beskriver sentrale aktører, prosjektdesign, mål for og organisering av lokale prosesser, og endelige produkter i form av planer og vedtak. Effektene av slike vedtak måles ikke, blant annet fordi de antas å være av så vidt langsiktig karakter. Det analytiske skjemaet impliserer ikke spesifisering av hvilke variable som kan antas å påvirke utfallet. Dataene baserer seg også på flere typer metoder; intervjuer, seminarer, spørreskjema. Basisen for tolkningen er slik mer omfattende beskrivelser der noen sentrale begreper (som sentrale aktører, føringer) har heuristisk karakter og strukturerer drøftingen. Analysen bygger slik ikke på et eksplisitt teoretisk grunnlag, men kan sees som å bygge på statsvitenskaplige og organisasjonsteoretisk fundament som ikke er eksplisitt. Designet bygger ikke på statistisk analyse og kan dermed heller ikke sies å tilfredsstillende kravene til statistisk validitet. Manglende eksplisitt teoretisk grunnlag gjør det vanskelig både å kartlegge den interne validiteten og den eksterne. Designet er slik lite formalisert og vurdering av evalueringen må basere seg på substansiell tolkning av feltet. Gyldighetsområdet og den eksterne validiteten avgrenser seg dermed til samordnet areal og transportpolitikk i norske byområder. På dette feltet har den imidlertid også hatt effekter; det fulgte en rekke utredninger om svakheter ved organiseringen av den lokale areal- og transportpolitikken i kjølvannet av denne evalueringen. Den har altså vært kumulativ i den forstand at den har vært oppfattet som basert på hovedfunn om kjennetegn ved sektoren, en forståelse som rett nok ikke så mye har vært drevet fram av fagetatene (som jo her delvis er blitt definert som problem) som av eksterne aktører, ikke minst politikere.

Kapitlet avsluttet med en kortfattet sammenstilling av funn fra Reorient. Undersøkelsen omhandler altså effekter av en større organisatorisk og institusjonell reform Europa - liberaliseringen av godstransportmarkedet med jernbane – og spørsmålet om dette har hatt de ønskede effekter i form av økt konkurranse og styrkede markedsandeler. Undersøkelsen illustrerer et trekk ved mange typer organisasjonsvariable; at en står framfor få cases (12 land), og en rekke variable som gjør statistiske analyser problematisk. Derimot kan slike analyser dokumentere interessante observasjoner når det gjelder manglende samvariasjon mellom ulike variable som kan gir grunnlag for problematisering av antagelsen om årsak-virkningsforhold som ofte ligger til grunn for den offentlige politikken.

Drøftingen avsluttet med en vurdering av om en kunne bruke begreper som barrierer eller veto-punkter som analytiske og metodiske grep i denne type analyse. Konklusjonen er imidlertid at uansett metodisk grep av denne type, er det avgjørende for denne type evalueringer å utvikle design for å styrke den eksterne validiteten, å gjøre det mulig å teste betydningen av organisasjonsvariabelen, tilsvarende slik det ble gjennomført i Friskbuss-prosjektet: Systematisk etterprøving av organisasjonsvariables effekter, tilsvarende det opplegget vi skisserte for mega-prosjekter.

Oppsummert er organisasjonsevaluering et sentralt område, fordi det er grunn til å anta at organiseringen av et tiltak i mange tilfeller kan påvirke resultatet. Samtidig trekk ved kvalitative evalueringer at de vanskelig håndterer kontrafaktisitet-problemet; strengt tatt ikke en målestokk for å vurdere relevansen av organisasjonsvariabelen. Videre er dette en form for evaluering der påvirkning på konklusjonene lettere kan gjennomføres enn i kvantitative undersøkelser – mer plastisk materiale og konklusjoner. Sammenlignet med de mer formaliserte og matematiserte modellene, har disse sin styrke i transparens – kan lettere valideres gjennom den kontroll som ligger i offentliggjøring.

5.2. Avsluttende drøfting – implikasjoner for videre arbeid

Noen konklusjoner:

- Det analytiske klassifikasjonsskjemaet og vektleggingen av ulike validitetskrav har vist seg fruktbar for den type gjennomgang vi har gjennomført her. De bør ha en sentral plass ved sammenfatning av evalueringer, også når disse ikke tar form av metaanalyser i den bestemte statistiske betydningen Elvik har introdusert dem for. Kravet til statistisk validitet trenger imidlertid ikke, når en har et mer kvalitativt sammenfattende siktemål, opprettholdes, dersom dette medfører at en utdefinerer kvalitative evalueringer som inkluderer organisasjons- eller prosessanalyse. Det er imidlertid særegne problemer knyttet til denne type evaluering når en ikke har effektvariable (se nedenfor)
- Variabel analysen og de øvrige disiplinfrakoblede og lite teoretiserte evalueringer, er preget av lav ekstern validitet. Denne type evalueringer kan benevnes som substansielle evalueringer, dvs. de utsier bare noe om et gitt saksområde, og har ikke som formål å generere mer generell kunnskap, utover saksområdet. Samtidig bør en ha ambisjon om å undersøke om konklusjonene som et minimum kan benyttes om det samme saksområdet. Dersom undersøkelsen omhandler evaluering av tiltak i bytransportpolitikken, bør ambisjonen være å undersøke om funnene samsvarer med undersøkelser på andre tidspunkt og i andre nasjonale kontekster. Dermed blir den viktigste mekanismen for å øke ekstern validitet oppfølging av evalueringer i form av systematisk forskning: Å følge opp hypoteser i form av komparative undersøkelser.
- Formaliserte, matematiserte modeller er lite transparente. Transparens kan økes gjennom to mekanismer; for det første kvalitetskontroll gjennom fagfellesskapet/disiplinen, fortrinnsvis gjennom peer-review. Dersom området ikke er utviklet som forskningsområde, er det viktig med konkurrerende fagmiljø og konkurrerende modeller for kontroll av funn. For det andre den kontroll som utøves gjennom offentligheten. Her er det også av betydning av å synliggjøre premisser og forutsetninger for konklusjoner, og usikkerhet knyttet til beregninger. Slik kan disiplinen og dens utøvere utvikle epistemisk refleksivitet, kunnskap om disiplinens gyldighetsområde og hvordan det oppfattes.

- Nyttekostnadsanalysen er på mange måter en eksemplarisk disiplin dersom en vurderer den med bakgrunn i de ulike validitetskrav. Den har særlig et kritisk punkt; omgjøring av slike forhold som tids- og miljøverdier til pengeverdsetting, men i denne sektoren også problemet med aggregerings av tidsverdier. Slik illustrerer den et poeng utover de validitetskrav vi har beskrevet innledningsvis: Å spesifisere et gyldighetsområde er ikke bare teoriinternt og omhandler ikke bare ”det forklarte” - hvilket saksområde en undersøkelse omhandler. Det er også knyttet til ”det forklarende”, den modell som benyttes for å evaluere dette saksområdet. Dette omhandler også grensene for en disiplin eller teoris gyldighetsområde. Dette blir synliggjort gjennom interdisiplinær kritikk og argumentasjon.
- Gjennomgangen viser at vi er kommet kort forskningsmessig på analysering av det som er et sentralt tema ut i fra vesentlighetsvurderinger; analyse av mega-prosjekter. Den mest framtrædende forskningen her internasjonalt er preget omfattende metodiske svakheter. Mange grunner, både teoretiske, metodiske og substansielle tilsier dette at det burde bli et sentralt evaluerings-/og forskningsområde framover.
- Organisasjonsanalyse uten data om effekter støter på rekke metodiske problemer. Vår gjennomgang av en av de mer kjente internasjonale undersøkelsene, Aalborg-undersøkelsen om megaprojekter, dokumenterer at her finnes det også omfattende svakheter med design som inkluderer effektvariable. Dersom det vi har sett på her er representativt, er en viktig problemstilling innen evalueringsforskning i samferdselssektoren å styrke metodisk kunnskap om organisasjonsevaluering, og kvalitative evalueringer generelt. Disse lider av et overordnet problem knyttet til å svare for kontrafaktisitet; som et minimum bør vi kreve at de gjør greie for at de ikke kan gjøre greie for dette. Dermed blir det problematisk dersom de som arbeider med kvalitativt funderte organisasjonsevalueringer er mer opptatt av å være advokater for fortrinnet ved denne type evaluering enn å diskutere dens svakheter. Problemet med manglende håndtering av kontrafaktisitet er spesielt kritisk dersom evalueringer konkluderer med at organisering eller utforming av tiltak grunnen til at dette ikke virker. En hypotese her vil være at vi allment, i den type struktur som vi nå har, med rekke konkurrerende kunnskapsfunn, vil det vi retorisk kan formulere som ”mer medisn- paradigmet” være dominerende. En tar for gitt at det problemet en retter sin oppmerksomhet mot lar seg løse, problemet er at en enda ikke har funnet løsningen – eller fått aksept på den.
- Det er et fellestrekk for alle de evalueringer vi har gjennomgått her, at de i en viss forstand er teoriimmune. De er preget av at en ikke undersøker om teorien som ligger til grunn for undersøkelsen støttes av resultatet av undersøkelsen. Dette innebærer at det er mulig og sannsynlig at det etableres epistemiske samfunn på tvers av evaluering og forskning, som baserer seg på common beliefs som ikke lar seg avkrefte. Dette øker betydningen av tverrfaglig og tverrsektoriell kommunikasjon og kritikk, og dermed betydningen av institusjonelle ordninger som medvirker til dette. Det impliserer også at det kunne være fruktbart og interessant å gjennomføre mer omfattende undersøkelser av forholdet mellom

evalueringstradisjoner, epistemiske samfunn og bruken av resultater fra evalueringer – ikke minst hvordan en forholder seg til 0-hypotesen. Det er ikke uvanlig at forskere uttrykker irritasjon over at resultater fra evalueringer ikke får virkning i den politiske beslutningsprosess, som at virkemidler en vet har effekter ikke tas i bruk av andre, utenforliggende grunner. Det er like interessant å undersøke om, i hvilken grad eller på hvilken måte mangel på dokumenterte effekter ikke fører til at en gir opp virkemidlene.

Referanser

- Abbott, Andrew 1988 *The System of Professions*. The University of Chicago Press.
- Baklien, Bergljot 2000. "Evalueringforskning for og om forvaltningen". i Olaf Foss og Jan Mønnesland (red.) *Evaluering av offentlig virksomhet. Metoder og vurderinger*. NIBRs pluss-serie 4-2000.
- Barnes, B og Bloor, D 1982 "Relativism, Rationalism and the Sociology of Knowledge" i Hollies and Luke (ed.) *Rationality and Relativism*. Basil Blackwell.
- Bekken, Jon-Terje og Oddgeir Osland 2004. *Transportpakker i by. Rammebetingelser, organisering og innhold - en oversikt*. TØI rapport 744/2004. Oslo: Transportøkonomisk institutt.
- Bruzelius, N, Flyvbjerg B og Rothengatter W 2002: *Big decisions, big risks: Improving accountability in mega projects*. *Transport Policy* 9:3, 143-54.
- Cameron, Charles M 2000. *Veto Bargaining. Presidents and the Politics of Negative Power*. Cambridge.
- Dahler-Larsen, Petter 2005 *Den rituelle refleksjon – om evaluering i organisasjoner*. Syddansk universitetsforlag.
- Elvik, R og T Vaa 2004: *The Handbook of Road Safety Measures*. Elsevier
- Eriksen, Knut S og Harald Minken 2004a. *Etterprøving av jernbaneprosjekter. Sandeparsellene*. Arbeidsdokument TØ/1699/2004.
- Eriksen, Knut S og Harald Minken 2004b. *Etterprøving av jernbaneprosjekter. Nationaltheatret stasjon*. Arbeidsdokument TØ/1681/2004
- Eriksen, Knut S. 2003. *Analyseskjema for etteranalyse av transportprosjekter. Arbeidsdokument*. TØI/1476/2003.
- Fearnley, Nils og Åse Nossum 2004. *Tiltakspakker for kollektivtransport 1996-2000. Samfunnsøkonomiske analyser*. Oslo: Transportøkonomisk institutt.
- Flyvbjerg, B 2005: [Measuring Inaccuracy in Travel Demand Forecasting: Methodological Considerations Regarding Ramp Up and Sampling](#). *Transportation Research A* 39:6, 522-530.
- Flyvbjerg, B, Bruzelius N og Rothengatter W 2003a: *Megaprojects and Risk: An Anatomy of Ambition*. Cambridge: Cambridge University Press, 207s.
- Flyvbjerg, B, Skamris Holm M K og Buhl S L 2002: *Underestimating costs in public works projects: Error or lie?* *APA Journal* 68:3, 279-95.
- Flyvbjerg, B, Skamris Holm M K og Buhl S L 2003b: *How common and how large are cost overruns in transport infrastructure projects?* *Transport Reviews* 23:1, 71-88.

Flyvbjerg, B, Skamris Holm M K og Buhl S L 2004: [What Causes Cost Overrun in Transport Infrastructure Projects?](#) *Transport Reviews* 24:1, 3-18.

Flyvbjerg, B, Skamris Holm M K og Buhl S L 2005: [How \(In\)accurate Are Demand Forecasts in Public Works Projects? The Case of Transportation.](#) *Journal of the American Planning Association* 71:2, 131-146.

Flyvbjerg, B, Skamris Holm M K og Buhl S L 2006: [Inaccuracy in Traffic Forecasts.](#) *Transport Reviews* 26:1, 1-24.

Flyvbjerg, B 2007: *Megaproject Policy and Planning: Problems, Causes, Cures.* Institut for Samfunnsutvikling og Planlægning, Aalborg Universitet

Foss, Olav 2000 Evaluering – samfunnsforskning i politikkenes tjeneste i Olaf Foss og Jan Mønnesland (red.) Evaluering av offentlig virksomhet. Metoder og vurderinger. NIBRs pluss-serie 4-2000.

Gilroy, JM and M Wade (ed) 1992: *The Moral Dimensions of Public Choice. Beyond the Market Paradigm.* University of Pittsburg Press.

Grønn, E 1988: *Nyttekostnadsanalyse.* Oslo: Bedriftsøkonomens forlag, 177s.

Grønmo, S 1996: "Forholdet mellom kvalitative og kvantitative tilnæringer i samfunnsforskningen" i Holter H og R Kalleberg (red.) *Kvalitative metoder i samfunnsforskning.* Universitetsforlaget, Oslo 1996

Grøvdal, Anker og Harald M. Hjelle 1998: *Innføring i transportøkonomi.* Bergen. Fagbokforlaget.

Haas, P.M 1992: "Introduction. Epistemic Communities and International Policy Coordination." i *International Organization.* Vol.46 No 1. pp 1-35

Hompland, A (red.) 2001: *Byens veier.* Lokal transport og arealpolitikk. Fagbokforlaget.

Jernbaneverkets metodebok JD 205

Jordange, I, S Malerud, H Minken, A Strand 2007. *Flermålsanalyser i store statlige investeringsprosjekt.* Concept rapport nr 9.

Killi, M 1999: *Anbefalte tidsverdier i persontransport.* TØI-rapport 459, 38s.

Kjørstad, Katrine Næss og Bård Norheim 2005. *Hva tiltakspakkene for kollektivtransport har lært oss.* TØI rapport 810/2005. Oslo: Transportøkonomisk institutt.

Kjørstad, Katrine Næss og Bård Norheim 2005. *Tiltakspakker for kollektivtransport 1996 – 2000. Befolkningens vurdering av tiltakene og effekter på reisemiddelvalg.* TØI rapport 794/2005. Oslo: Transportøkonomisk institutt.

Klæboe, R 2006: *Human reactions to transportation noise, vehicular air pollution, and vibrations. Simple, and not so simple, exposure--effect relationships.* Doctoral thesis for the degree of Doctor Philosophiae. Faculty of Social Sciences. University of Oslo 2006

- Kringlen, E: 2007 *Norsk psykiatri gjennom tidene*. Damm forlag.
- Kuhn, Thomas 1962. *The Structure of Scientific Revolutions*.
- Latour, B and Woolgar, S 1979: *Laboratory Life. The Social Construction of Scientific Facts*. Sage Publications.
- Ludvigsen, J og O Osland: 2008 forthcoming. "Variation in European Liberalisation of Rail Freight Markets", i EJTIR special issue.
- Longva, Frode, Oddgeir Osland og Trine Hagen 2007. *Arbeidsmiljøtiltak i bussbransjen – evaluering av forsøk med team*. TØI rapport 908/2007. Oslo: Transportøkonomisk institutt.
- Mackie, P og Preston J 1998: *Twenty-one Sources of Error and Bias in Transport Project Appraisal*. Transport Policy 5:1-7.
- Minken, H 2005: *Risikotillegget i kalkulasjonsrenta*. TØI-rapport 796, 52s.
- Minken, H og Johansen KW 2004: *Håpløse transportmodeller?* Samferdsel 10, 24-25.
- Minken, Harald 2005. *Nyttekostnadsanalyse og prioriteringer på tvers av sektorene*. Arbeidsdokument TØI/1748/2005.
- Minken, H og T Holmeid 2005: *Nyttekostnadsanalyse og prioriteringer på tvers av sektorene*. Arbeidsdokument TØI/1748/2005
- Nagarajan, N og Vanheukelen, M 1997: *Evaluating EU Expenditure Programmes, a Guide*. Luxembourg, Office for Official Publications of the European Communities.
- Nielsen, O. A. og Fosgerau, M 2005 *Overvurderes tidsbenefit af vejprojekter?* Paper til Trafikdage på Aalborg Universitet
- Norheim, Bård og Katrine Næss Kjørstad 2004. *Tiltakspakker for kollektivtransport 1996 – 2000. Kollektivtrafikantenes vurdering av tiltakene og endret bruk av buss*. TØI rapport 736/2004.
- Næss, Petter 2006a. Cost-benefit analyses of transportation investments: neither critical nor realistic. *Journal of Critical Realism*, Vol 5., No1, 2006, pp 32-60
- Næss, Petter 2007. Cost-benefit-analyser, tall og usikkerhet. *Økonomi og Samfund*, 2007.
- Odeck, J 1996: *Ranking of Regional Road Investment in Norway: Does Socioeconomic Analysis Matter?* Transportation 23:2, 123-40.
- Odeck, J 2004: *Cost Overruns in Road Construction: What Are Their Sizes and Determinants?* Transport Policy 11:43-53.
- Odeck, J 2005: *Nå blir samferdselsprosjekter mer lønnsomme*. Samferdsel 8: 12-13.
- Odeck, J og Brudeseth, Å 2003: *Statens vegvesen: Generelt god kostnadskontroll, men tunnelprosjektene sprekker*. Samferdsel 8: 20-23.
- Odeck, J og Brudeseth, Å. 2003. Statens vegvesen: Generelt god kostnadskontroll, men tunnelprosjektene sprekker. *Samferdsel* 8: 20-23.

Osland og Leiren 2007. *Institutional and political conditions for the establishment of congestion charging regimes. A comparison of Norwegian and Swedish experiences* (presentert på 10th International Conference on Competition and Ownership in Land Passenger Transport, Hamilton Island, 12-17 August 2007).

Osland, Oddgeir og Trond Kråkenes 1998. *New Deal eller nye spilleregler. Om kollektivtransportens forvaltningsorganisasjon og valg av politiske styringsinstrumenter for å fremme kollektivtransporten*. Rapport nr. 65 (1998). Institutt for administrasjon og organisasjonsvitenskap.

Osland, Oddgeir og Trond Kråkenes 2000. *Kollektivtransport og kollektive forpliktelser – innspill til endringer i transportpolitikken organisasjon i norske byområder*. Rapport nr. 72 (2000). Institutt for administrasjon og organisasjonsvitenskap.

Spangen, I 1995: *Institusjonelle forutsetninger for en miljørettet samferdselspolitikk. Mulighetene for å iverksette miljøstrategien i lokale transportplaner (TP 10)*. TØI-rapport 306/1995

Ravlum, I-A og Sørensen, C H 2005: *Styring, delegering og innflytelse?* TØI-rapport 783/2005, 64s.

Renolen, Heidi 1998. Kollektivtransport. *Praktisk evalueringsveileder for tiltakspakker i by*. TØI rapport 388/1998. Oslo: Transportøkonomisk institutt.

Riksrevisjonens undersøkelse av store vegbudsjettsoverskridelser
Innstilling 193 til Stortinget

Rose, Gerry 1982 *Deciphering Sociological Research*. MacMillan. 1982

Ruud, Alberte 2003. *Erfaringer med lave takster i kollektivtransporten. En litteraturstudie*. TØI rapport 673/2003. Oslo: Transportøkonomisk institutt.

Sager, T og Ravlum I-A 2005: *The Political Relevance of Planners' Analysis: The Case of a Parliamentary Standing Committee*. *Planning Theory* 4:1, 33-65.

Schattschneider, E.E, (1960). *The Semisovereign People. A Realist's View of Democracy in America*. The Dryden Press.

Skamris, Holm M K 2000: *Economic Appraisal of Large Scale Transport Infrastructure Investments*. Ph D-avhandling, Aalborg University, Department of Development and Planning, Serie no 273, 248s.

Skamris, M K 1994: *Large transport projects: forecast versus actual traffic and costs*. Department of Development and Planning, Aalborg University.

Skamris, M K og Flyvbjerg, B 1997: *Inaccuracy of traffic forecasts and cost estimates on large transport projects*. *Transport Policy* 4:3, 141-46.

Skjervheim, H 1991: *Filosofi og dømmekraft*. Universitetsforlaget. Det blå bibliotek

Strand, S 1993: *Time in Transport: A Perverted Problem? Arguments for a Fresh Look at Time Utility Research and Its Application*. *Transportation Research Record* 1395: 10-14.

Strand, S 1998: *Finnmark og bortenfor, del 2: Usikkerheter i data og modeller*. Dr philos-avhandling, Universitetet i Bergen. 72s+.

- Strand, S 1999a: *Bottlenecks in Transportation Research: Understanding Behavior, the Alpha and Omega for the Applicability of Applied Research*. Transportation Research Record 1659: 46-50.
- Strand, S 1999b: *Forecasting the Future: Pitfalls in Controlling for Uncertainty*. Futures 31:333-50.
- Strand, S 2004. *Etterprøving av store (norske) samferdselsinvesteringer*. TØI-arbeidsdokument TR/1216, 55s.
- Strand, S 2006: *Å lære av suksess og fiasko. Etterprøving – læring – prognose. Noen forslag til hvordan evalueringen av samferdselstiltak kan nyttiggjøres*. TØI-arbeidsdokument PT/1852/2006
- Strand, S 2007: *Tid i transport. Noen synspunkter på faglig status*. TØI-arbeidsdokument OI/1941/2007
- Sverdrup, Sidsel (2002) *Evaluering. Faser, design og gjennomføring*. Fagbokforlaget.
- Taylor, Charles 1995. *Identitet, frihet och gemenskap*. Daidalos.
- Tennøy, A 2003: *Prediksjoner og usikkerhet i trafikkfaglige rapporter i KU*. Oslo: NIBR-rapport 2003:13, 98s.
- Tennøy, A 2004: *Transportanalysene hindrer bærekraftig byutvikling*. Samferdsel 9, 12-14
- Vedung, Evert 1998. *Utvärdering i politik och förvaltning*. Studentlitteratur 1998.
- Wachs, M 1990: *Ethics and Advocacy in Forecasting for Public Policy*. Business & Professional Ethics Journal 9:141-57.
- Weber, Max 1971 *Makt og byråkrati*. Gyldendal Norsk Forlag 1971.
- Aaheim, H. A 2003: *Hva er framtidens klima verdt?* Cicerone 1-2003

Vedlegg

Tabell 1. Kriterier for statistisk validitet slik disse formuleres hos Elvik (1999)

Undersøkelser	Kriterier for statistisk validitet
Enkelte	<ol style="list-style-type: none"> 1. Utvalgsmetode (valg av enheter) 2. Utvalgsstørrelse (antall enheter) 3. Målingers reliabilitet (uavhengig og avhengig variabel) 4. Systematiske feil i datagrunnlaget 5. Analyseteknikk (hvordan data analyseres)
Flere	<ol style="list-style-type: none"> 6. Sammenlignbarhet av definisjon av avhengige variabler 7. Publikasjonsskjevhet 8. Fordeling av resultater (modalitet, skjevhet, sterkt avvikende datapunkter) 9. Robust gjennomsnittresultat

Kilde: TØI-rapport 929/2007

Tabell 2 . Kriterier for intern validitet

Kriterier for intern validitet (kausaltitet)	Annen validitet
1. Statistisk sammenheng mellom årsaksvariabelen og virkningsvariabelen	Statistisk
2. Entydig årsaksretning (hva som er årsak og virkning er tydelig)	
3. Statistisk sammenheng må holde ved kontroll for andre mulige forklaringer	
4. Identifiserbar årsaksmekanisme som forklarer hvordan eller hvorfor årsaken skaper virkningen	Teoretisk
5. Sammenhengen mellom årsak og virkning er reproduisert under varierende betingelser i flere undersøkelser	Ekstern
6. Dose-responssammenheng mellom årsak og virkning, hvis både årsaks- og virkningsvariabelen har en stor nok variasjon	
7. Virkning av en årsak bør finnes kun i målgruppen og ikke i andre grupper, dersom det er mulig å identifisere en klar målgruppe for årsaksvariabelen	

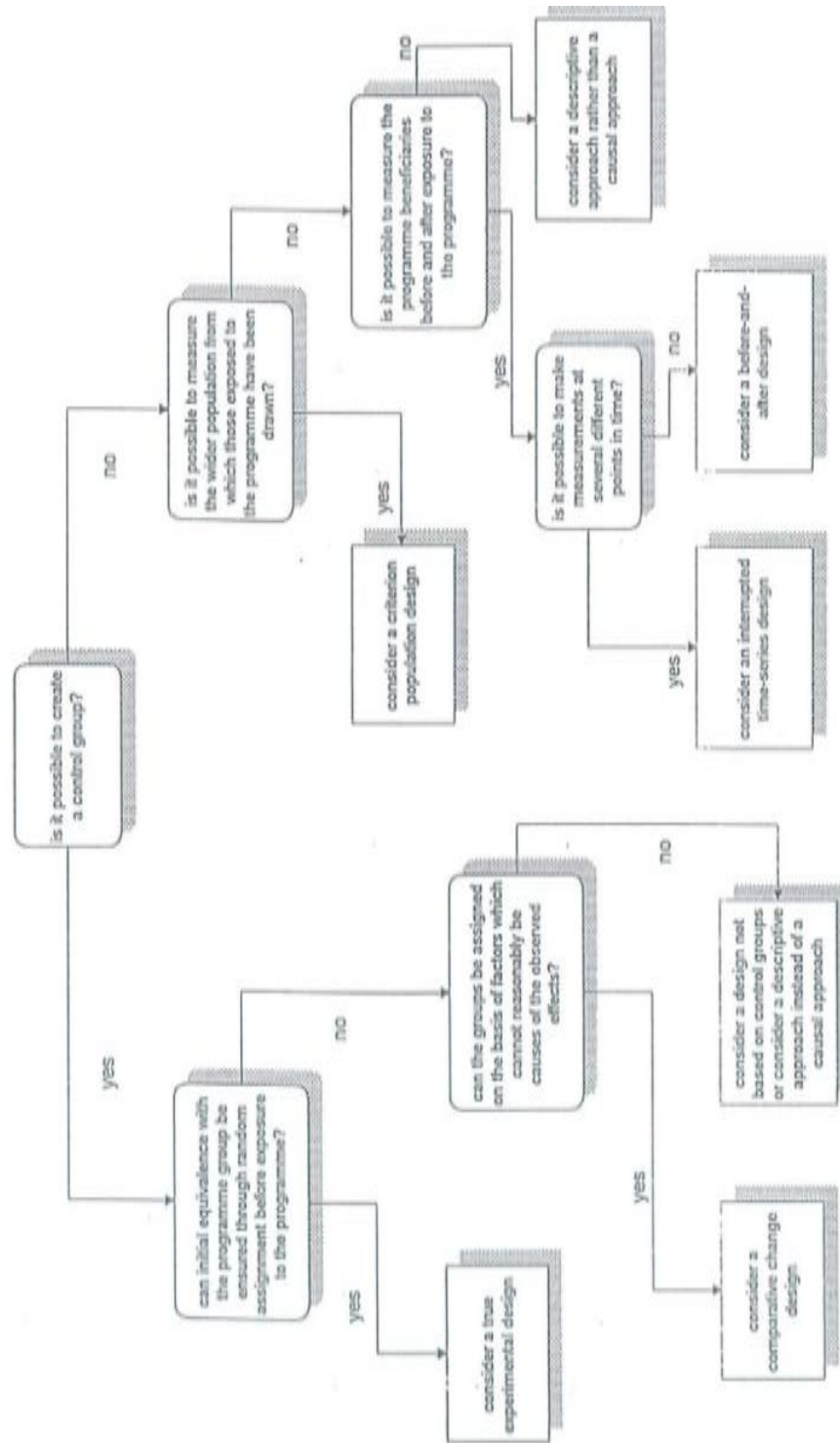
Kilde: TØI-rapport 929/2007

Tabell 3. Tiltak i tiltakspakkene (uavhengig variabel)

Byområde	Rutestruktur			Infrastrukturtiltak			Informasjon og markedsføring
	Økt frekvens	Kortere reisetid	Flere reiselasjoner	Holdplasser (flere eller opprusting)	Terminaler og knutepunkt	Framkommelighet	
Hundvågpakken i Stavanger	x			X	x		X
Tønsbergområdet	x			X	x		X
Drammensregionen	x			x		x	X
Nedre Glomma				X		x	X
Larviksområdet	x			X	x		x
Ålesundspakken	x			x		x	x
Grenland	x			X	x		x
Lillehammer og Gjøvik	x			X			x
"Buss 2000" i Tromsø				X	x		x
Trondheim	x	x	x		x		x
Kristiansand	x			X		x	x

Kilde: TØI-rapport 929/2007

Figure 4.2. Criteria for selecting an evaluation design (the causal approach)



Kilde: Nagarajan, N og Vanheukelen, M 1997

Besøks- og postadresse:

Transportøkonomisk institutt
Gaustadalléen 21
NO 0349 Oslo

Telefon: 22 57 38 00
Telefaks: 22 60 92 00
E-post: toi@toi.no

www.toi.no



**Transportøkonomisk institutt
Stiftelsen Norsk senter for samferdselsforskning**

- utfører forskning til nytte for samfunn og næringsliv
- har rundt 70 forskere med høy, flerfaglig samferdselskompetanse samarbeider med en rekke samfunnsinstitusjoner, forsknings- og undervisningssteder i Norge og i utlandet
- gjennomfører forsknings- og utredningsoppdrag av høy kvalitet innen områder som trafiksikkerhet, kollektivtransport, miljø, reisevaner, reiseliv, planlegging, beslutningsprosesser, transportøkonomi og næringslivets transporter
- driver aktiv forskningsformidling gjennom TØI-rapporter, Internett, tidsskriftet Samferdsel og andre nasjonale og internasjonale tidsskrifter
- deltar i CIENS, Forskningscenter for miljø og samfunn, i Forskningsparken nær Universitetet i Oslo