

ISBN 978-82-326-6156-5 (printet ver.)
ISBN 978-82-326-6178-7 (elektronisk ver.)
ISSN 1503-8181 (printet ver.)
ISSN 2703-8084 (elektronisk ver.)

 NTNU

Doktoravhandling hos NTNU, 2021:103

 NTNU

NTNU
Norges teknisk-
naturvitenskapelige universitet
Doktoravhandling for
Philosophiae Doctor
Fakultet for ingeniørvitenskap
Institutt for bygg- og miljøteknikk

Doktoravhandling hos NTNU, 2021:103

Lone-Eirin Lervåg

Evaluering av intelligente transportsystemer

Utvikling av evalueringsmetodikk basert på tiltakets programteori

 NTNU

Lone-Eirin Lervåg

Evaluering av intelligente transportsystemer

Utvikling av evalueringsmetodikk basert på tiltakets programteori

Doktoravhandling for Philosophiae Doctor

Trondheim, april 2021

Norges teknisknaturvitenskapelige universitet
Fakultet for ingeniørvitenskap
Institutt for bygg- og miljøteknikk



NTNU

Norges teknisknaturvitenskapelige universitet

Doktoravhandling for Philosophiae Doctor

Fakultet for ingeniørvitenskap
Institutt for bygg- og miljøteknikk

© Lone-Eirin Lervåg

ISBN 978-82-326-6156-5 (printet ver.)

ISBN 978-82-326-6178-7 (elektronisk ver.)

ISSN 1503-8181 (printet ver.)

ISSN 2703-8084 (elektronisk ver.)

Doktoravhandling hos NTNU, 2021:103



Printet av Skipnes Kommunikasjon AS

FORORD

Dette doktorgradsarbeidet er utført ved NTNU Institutt for bygg- og miljøteknikk, med finansiering fra Vegdirektoratets etatsprogram Smartere vegtrafikk med ITS (SmITS). Gjennom studiet har jeg vært ansatt som forsker ved SINTEF Mobilitet og samfunnsøkonomi.

Doktorgradsarbeidet har vært en spennende, men også svært arbeidskrevende prosess. I perioder har studiet ligget på vent, mens jeg har vært engasjert i andre forskningsprosjekter. Arbeidet er dermed strukket over flere år enn opprinnelig planlagt. Siden intelligente transportsystemer er et fagfelt i rask utvikling, er litteraturstudiet som er presentert i kapittel 2 i aller høyeste grad ferskvare.

Teknologistatus og relevante eksempler endres i høy takt. Kapittelet ble i hovedsak forfattet i 2016, og er senere revidert og oppdatert våren 2019. Selv om teknologien har fått stadig større utbredelse i løpet av disse årene, er hovedpoenget fortsatt gyldig: ITS er blitt en viktig del av transportsystemet, og vi trenger gode evalueringsmodeller for å frembringe et relevant og anvendbart kunnskapsgrunnlag. I nyere tid satses det i stadig økende grad på læring gjennom pilotering og utprøving av ny teknologi i transportsektoren. Dette gjør at problemstillingene i dette forskningsarbeidet oppleves minst like relevant i dag, som ved doktorgradsstudiets oppstart.

Dette prosjektet hadde ikke vært mulig uten støtte fra gode kollegaer i fagmiljøet. Takk til mine veiledere Eirin Olaussen Ryeng ved NTNU, Tomas Levin ved Statens vegvesen og Gunnar Deinboll Jenssen ved SINTEF. Takk til Ragnhild Wahl som tidlig viste meg tillit og ga meg ansvar som ung forsker. Takk til Per Johan Lillestøl som ga meg mulighet til å fordype meg i dette studiet innenfor SINTEFs trygge rammer. Takk til Anders Godal Holt, Eirik Skjetne, Erik Olsen og fagmiljøet i Vegdirektoratets SmITS-program for finansiering og tilrettelegging for innsamling av data. Takk til mine gode kollegaer Terje Moen, Trond Foss og Solveig Meland for samarbeidet i case-studiene. Takk til Terje Reitaas og Roar Norvik for tålmodighet og støtte i en lang periode hvor jeg har sjonglert doktorgradsarbeidet med andre forskningsprosjekt i SINTEF.

Takk til familie og venner som alltid heier på meg, og som gir meg gode grunner til å legge bort data og digitale skjermer når arbeidsdagen er over. Spesielt takk til Ståle, Mathea og Hermine for alle fine dager vi har sammen.

Trondheim, 1. desember 2020

Lone-Eirin Lervåg

SAMMENDRAG

EVALUERING AV INTELLIGENTE TRANSPORTSYSTEMER

UTVIKLING AV EVALUERINGSMETODIKK BASERT PÅ TILTAKETS PROGRAMTEORI

Intelligente transportsystemer (ITS) omfatter et mangfold av teknologiske løsninger og tjenester som forventes å spille en vesentlig rolle i utviklingen av et bærekraftig og fremtidsrettet transportsystem. Beslutningstakere etterspør relevant og anvendbar kunnskap som kan fremskynde innføringen av nyttig og lønnsom ITS, for å sikre at investeringer kommer samfunnet og befolkningen til gode. Kunnskap bygges blant annet gjennom evalueringer. Tilgang til gode evalueringsmodeller er derfor viktig i arbeidet med å utvikle fremtidens transportløsninger.

Det eksisterer flere vitenskapsteoretiske evalueringstilnæringer. Alle har sine styrker og svakheter, og evne til å fremskaffe ulike typer kunnskap. Transportfaget fremstår som svært metodestyrt, med den (kvasi-) eksperimentelle modellen som ideal for effektevalueringer, og nyttekostnadsmetodikken som grunnlag for samfunnsøkonomiske lønnsomhetsanalyser. ITS-tiltak har imidlertid en rekke egenskaper som gjør at de sjelden oppfyller de metodiske kravene som følger disse evalueringsmodellene:

- ITS er ikke et enhetlig tiltak eller virkemiddel. Det er derimot et mangfold av enkeltstående og sammensatte systemer, med komplekse årsakssammenhenger og et situasjonsavhengig effektbilde. Dette gjør det vanskelig å etablere en generell og felles evalueringsmetode for alle ITS-tiltak.
- ITS er ikke et stabilt tiltak. Teknologien er i stadig utvikling, noe som gjør at tiltaket fortløpende justeres og forbedres gjennom hele sin levetid. Dette gjør det vanskelig å sammenligne resultater over tid eller mellom ulike installasjoner.

Faglitteratur og effektkataloger viser derfor lite konsistente konklusjoner fra ulike evalueringsstudier, og ofte er både datagrunnlag og resultater beheftet med stor grad av usikkerhet. Kunnskapen blir lite anvendbar både for den som skal foreta beslutninger og for den som skal iverksette tiltak. I ytterste konsekvens begrenses innføringen av nyskapende løsninger, når evalueringsstudier ikke makter å dokumentere faktiske effekter og lønnsomhet.

Målsettingen med dette forskningsarbeidet har vært å bidra til bedre forståelse og kunnskap om et egnet metodegrunnlag for ITS-evalueringer. Dette er belyst gjennom drøfting av teorigrunnlag, tilpassing av evalueringsmetodikken og praktisk utprøving i to evalueringsstudier.

Litteraturstudiet av ITS som virkemiddel i transportsystemet har gitt innsikt i metodiske utfordringer knyttet til evaluering av ITS-tiltak. Arbeidet har ledet frem til fem konkrete anbefalinger for utvikling av evalueringsmetodikk for ITS-tiltak:

- 1) *Evalueringen bør ha et læringsperspektiv som bidrar til helhetlig kunnskapsbygging*
- 2) *Evalueringen bør kunne håndtere tiltak med ulik teknologisk modenhet*
- 3) *Evalueringen bør kunne identifisere de kontekstuelle faktorenes betydning*
- 4) *Evalueringen bør gi kunnskap om hvorfor et tiltak virker eller ikke*
- 5) *Evalueringen bør kunne identifisere og vurdere alle relevante konsekvenser, også kvalitative*

Evaluerings teorien har gitt innblikk i ulike evalueringsperspektiver, metodiske styrker og svakheter, samt tradisjoner for kunnskapsbygging innen ulike fagfelt. Med utgangspunkt i dette kunnskapsgrunnlaget, ble søkelyset rettet mot den *teoribaserte evalueringstilnærmingen* som synes å adressere flere typiske utfordringer med evaluering av ITS-tiltak: Den kan håndtere sammensatte og komplekse tiltak, den inkluderer kontekst i evalueringen, og den åpner for å kombinere flere metoder og analyseteknikker.

Dette har gitt grunnlag for å stille spørsmål om en dreining fra dagens metodestyrt tilnærming til en teoribasert evalueringsmetodikk kan styrke beslutningsgrunnlaget for implementering av ITS i transportsystemet. Problemstillingen er utforsket nærmere i to konkrete case-studier, hvor en teoribasert evalueringsprosess ble utviklet og anvendt i evaluering av ITS-tiltak.

Det første case-studiet omhandlet evaluering av ITS-tiltaket *NonStop klarering av tunge kjøretøy*. NonStop-tjenesten viste seg å ha begrenset funksjonalitet og vesentlig lavere teknologisk modenhet enn forutsatt ved prosjektets oppstart. Den teoribaserte fremgangsmåten var likevel praktisk gjennomførbar, selv med begrenset datagrunnlag og i en situasjon hvor premisene for en tradisjonell eksperimentell evaluering ikke var oppfylt. Metodevalget i forsøksdesignet ble styrt av relevante forskningsspørsmål og evalueringen evnet å frembringe relevant kunnskap om konseptet NonStop, til tross for at det konkrete tiltaket ikke fungerte tilfredsstillende. Situasjonsavhengige faktorer (kontekst) ble inkludert i evalueringen, og ga innsikt i hvilke eksterne faktorer som bidrar til å forsterke eller forhindre at man realiserer de effektene man ønsker å oppnå.

Det andre case-studiet omhandlet evaluering av ITS-tiltaket *Samkjøring*, med utgangspunkt i et pågående pilotprosjekt i Bergen. Samkjøringstiltaket var preget av høy kompleksitet med mange involverte aktører og aktiviteter, og en forventning om at omstendigheter rundt implementeringen

(kontekst) ville har stor betydning for realiserte effekter. Den teoribaserte evalueringsstrategien la til rette for en systematisk evaluering, med et forsøksdesign som kombinerte både kvasiexperimentelle og realistiske metoder. Den metodiske fremgangsmåten skilte seg fra tidligere forskning, ved at den i større grad la vekt på å identifisere rammebetingelser og motivasjonsfaktorer for å lykkes med samkjøringstiltaket. Evalueringen bidro dermed til å avdekke gunstige betingelser som muliggjør målrettet utforming og bruk av samkjøringstiltak i fremtiden.

Utprøving av den teoribaserte tilnærmingen i ITS-evalueringer har vist at den har kapasitet til å fremskaffe helhetlig forståelse og relevant kunnskap om virkemidlene i det moderne transportsystemet. Med utgangspunkt i erfaringene fra dette forskningsarbeidet er det foreslått en teoribasert evalueringsprosess for ITS-tiltak, som inkluderer følgende trinn:

Trinn 1: Vurdere tiltakets teknologiske modenhet

Trinn 2: Behovsanalyse

Trinn 3: Etablere tiltakets programteori (logikk)

Trinn 4: Definere evalueringskriterier og forskningsspørsmål

Trinn 5: Valg av evalueringsdesign

Trinn 6: Datainnsamling

Trinn 7: Analyser

Trinn 8: Presentasjon av resultater

Evalueringsmetodikken er metodenøytral og kan benyttes gjennom hele beslutningsprosessen fra tidlig konseptutvikling til tiltaket er iverksatt og fungerer i stabil drift. Hvor tiltaket befinner seg i dette utviklingsløpet påvirker hvilke evalueringskriterier som er relevante, hvilke forskningsspørsmål som er mulig å besvare, og hvilke metoder som er hensiktsmessig å benytte i evalueringsdesignet. Det teoribaserte evalueringsperspektivet skiller seg fra dagens metodestyrte tilnærming på to vesentlige områder: Det har en mekanistisk forståelse av kausale årsakssammenhenger og inkluderer kontekst som en relevant faktor i evalueringen. Dette gir stor forklaringskraft med hensyn til å besvare hva som bidrar til at et tiltak virker eller ikke virker – og i hvilke situasjoner man kan forvente å oppnå effekt.

Det teoribaserte metodegrunnlaget har sin styrke i evalueringer med tverrfaglig kunnskapsbygging som formål, og når tiltaket har høy grad av kompleksitet eller mange involverte aktører.

Teoribaserte evalueringsdesign kan også anvendes når det eksisterer lite empiri om

årsakssammenhenger og effekter på forhånd. Metodiske begrensninger er i hovedsak knyttet til spørsmål rundt evalueringer i et kontrollperspektiv (dokumentasjon av effektiviteten til et tiltak), fordi fremgangsmåten anses å ha svakere vitenskapelig evidens i kvantitative analyser og fordi samspillet med involverte aktører kan svekke evalueringens objektivitet. Det er imidlertid viktig å understreke at en overordnet teoribasert tilnærming ikke forkaster bruk av eksperimentell metode når betingelsene ligger til rette for dette, men åpner for å kombinere den med andre analyseteknikker i tillegg.

Erfaringene fra dette doktorgradsarbeidet gir grunn til å tro at en teoribasert tilnærming til ITS-evalueringer vil bidra til å skape et mer helhetlig kunnskapsgrunnlag og en bedre utnyttelse av evalueringsressurser innenfor fagfeltet.

SUMMARY

EVALUATION OF INTELLIGENT TRANSPORT SYSTEMS

DEVELOPMENT OF A THEORY-DRIVEN EVALUATION METHODOLOGY

Technological innovations hold a potential to change the way transport is performed, giving rise to the development of an efficient and sustainable future transport system. To obtain beneficial impacts for users and society, policy makers are in urgent need of knowledge to increase the understanding of how intelligent transport systems (ITS) should be developed and implemented. Building knowledge and evidence on benefits and impacts requires the presence of appropriate and valid evaluation models and methods.

Although there is a wide range of evaluation models available, research and evaluation studies within the transport sector commonly follow a rather strict method-driven tradition with the scientific experimental model as the gold standard for impact assessments and socio-economic analyses based on cost-benefit calculations. The prevailing methodological framework is, however, developed for and adapted to conventional road and traffic measures and it has yet proven inadequate for assessment of intelligent transport services for several reasons:

- ITS encompass a variety of technological solutions and services, with complex causal mechanisms and substantial contextual influence on results and outcome. This makes it rather challenging to develop a general and common evaluation methodology for ITS services.
- ITS is subject to accelerating technology development. Hence, measures and services undergo incremental improvements throughout its lifetime, making it rather difficult to compare outcome over time or between different implementations.

Method-driven evaluations have in many cases failed to provide relevant and applicable knowledge for policy makers, as the data quality and results are subjected to uncertainty. This is not necessarily due to weaknesses or shortcomings in the conventional methods themselves, but the fact that they are applied without fulfilling the prerequisites for use. Consequently, the introduction of innovative and beneficial ITS may be limited by the fact that evaluation studies are not able to document impacts and profitability.

The aim of this PhD study has been to accumulate knowledge and understanding of the methodological basis for ITS evaluations. This is elucidated by discussions of ITS and evaluation theory, adaptation of an evaluation methodology, and practical experience in two case studies.

A literature review of intelligent transport systems as an instrument to meet transport policies, has provided valuable insight into methodological challenges of ITS evaluations. Five specific recommendations for development of a framework for assessment of ITS are outlined:

- 1) *The evaluation should contribute to increased knowledge and improved practices*
- 2) *The evaluation should be able to deal with different levels of technological maturity of ITS*
- 3) *The evaluation should be able to identify the significance of contextual factors*
- 4) *The evaluation should provide knowledge to explain why the ITS service works or not*
- 5) *The evaluation should identify and address all relevant impacts, including the qualitative ones*

The evaluation theory has provided insight on evaluation perspectives, methodological strengths and weaknesses, as well as traditions for knowledge production within different scientific approaches. Consequently, the focus was set on *the theory-based evaluation approach* as it is suggested to address typical challenges related to ITS evaluations: It has the capacity to deal with highly innovative systems, the ability to include contextual influence in the assessment, and a methodological flexibility that allows for combination of multiple research designs and analyses.

The theoretical basis has led to further discussions on how moving from today's method-driven impact assessments to a theory-based evaluation approach are likely to improve our knowledge and understanding of how ITS services should be developed and implemented in the future transportation system. Practical experience is obtained through development of a theory-based evaluation process that is accordingly applied in two separate case studies on evaluation of ITS services.

The first case study involved the assessment of an automated pre-clearance system of heavy goods vehicles. The ITS service faced significant challenges in terms of limited functionality and a lower degree of technological readiness compared to what was originally intended. Still, the theory-based evaluation approach turned out to be feasible, and produced applicable knowledge for policy makers, despite of the drawbacks related to a limited data collection, and the fact that the terms of a traditional quasi-experimental before- and after study was not fully met. The theory-based methodology facilitated the development of an appropriate evaluation design in accordance with the evaluation purpose and research questions of interest. Contextual factors were included in the assessment and provided insights on external conditions that are likely to influence the future realization of the desired impact.

The second case study involved assessment of a dynamic ride-sharing service based on an ongoing pilot project in the city of Bergen. The service was characterized by high complexity in terms of the involvement of multiple actors and activities, combined with an expectation that the circumstances related to the initiative (context) would have a substantial influence on effects and outcome. The theory-based approach facilitated a systematic assessment, combining both quasi-experimental and realistic evaluation designs and methods. The theory-based evaluation methodology differed from previous research by emphasizing the external conditions, incentives and motivational factors relevant for a successful implementation of the dynamic ride-sharing service. Thus, the evaluation contributed to disclose critical requirements and favorable conditions that may enable targeted design and utilization of the ITS service in the future.

The theory-based evaluation approach is suggested to be beneficial in impact assessment of ITS, as it has proven to generate applicable knowledge as a basis for policy decisions on future development and deployment in both case studies. Based on the results and experiences achieved during the PhD study, a theory-based evaluation process for assessment of ITS is proposed:

Step 1: Technology readiness assessment

Step 2: Problem analysis and needs assessment

Step 3: Establish the program theory (logic) of the ITS service

Step 4: Define evaluation criteria and research questions

Step 5: Choice of evaluation design

Step 6: Data collection

Step 7: Analysis

Step 8: Presentation of results

This evaluation methodology is applicable throughout the whole innovation and decision-making process, ranging from early concept development phase until the ITS service is applied and ready for market deployment. It has a methodological flexibility, that allows for the combination of multiple evaluation designs and methods. The timing of the evaluation in terms of technology readiness is crucial for the development of relevant evaluation criteria and research questions of interest, as well as data collection possibilities and choice of methods. The main differences between the theory-based approach and conventional experimental designs are twofold: The theory-

based evaluation design has a mechanistic rather than counterfactual approach to determining causality, and the influence of contextual factors on results and outcome is included in the evaluation design. This provides a capacity to explain why the intervention provides a specific result, or otherwise why it fails to do so, making it possible to identify the circumstances and conditions that would enhance a successful implementation.

The theory-based evaluation approach is suggested to be beneficial when the purpose of the evaluation is to serve interdisciplinary knowledge development, and when the ITS service has a high complexity level or involvement of multiple stakeholders. The approach is also applicable when the empirical evidence on causal relationships and impacts is limited. The theory-based evaluation methods are however less likely to provide quantitative evidence with the same scientific power that is attended with the experimental model. Also, the objectivity may be questioned by the close cooperation between the evaluator and stakeholders during the evaluation process. Thus, the theory-based methods are considered to be more beneficial in evaluations with a learning perspective, and less suitable in producing documentation of effectiveness as a basis for control and accountability decisions. This may however be compensated by the fact that the theory-based approach allows for a methodological flexibility which enables experimental analyses to be included in the assessment when appropriate.

Lessons learned from this PhD study provides reason to believe that a theory-based approach to ITS evaluations will contribute to increased knowledge and improved leverage of evaluation resources within the ITS discipline.

1	INTRODUKSJON	1
1.1	BAKGRUNN	1
1.2	PROBLEMSTILLINGER	3
1.3	METODE	6
1.4	AVGRENSNING OG BEGREPSAVKLARING	7
1.5	STRUKTUR	8
2	INTELLIGENTE TRANSPORTSYSTEMER	11
2.1	DIGITALISERING AV HVERDAGEN	11
2.2	ITS SOM VIRKEMIDDEL I TRANSPORTSEKTOREN	17
2.2.1	TRANSPORTPOLITISKE RETNINGSLINJER OG STRATEGIER	17
2.2.2	TEKNOLOGI OG VIKTIGE BEGREPER	18
2.2.3	TJENESTER	24
2.2.4	FORSKNING OG UTVIKLING AV ITS	26
2.3	TEKNOLOGIEN INTRODUSERER NYE PROBLEMSTILLINGER OG ENDRET KUNNSKAPSBEHOV	30
3	EVALUERINGSTEORI	33
3.1	EVALUERING SOM BEGREP	33
3.2	EVALUERING SOM FAGFELT	36
3.3	EVALUERINGSPERSPEKTIVER	38
3.3.1	MÅLORIENTERT EVALUERING	39
3.3.2	BRUKERORIENTERT EVALUERING	40
3.3.3	PROSESSORIENTERT EVALUERING	41
3.3.4	MÅLFRI EVALUERING	42
3.3.5	DEVELOPMENTAL EVALUERING	43
3.3.6	TEORIBASERT EVALUERING	44
3.3.7	REALISTISK EVALUERING	45
3.3.8	SAMMENFATNING AV EVALUERINGSPERSPEKTIVER	46
3.4	EVALUERINGSDESIGN OG METODER	47
3.4.1	EKSPERIMENTELL METODE	47
3.4.2	IKKE-EKSPERIMENTELL METODE	49
3.4.3	VALIDITET	50
3.4.4	KAUSALITET	51
3.4.5	EVIDENSDEBATTEN	52
3.5	EVALUERING I TRANSPORTSEKTOREN	54
3.5.1	TRANSPORTFORSKNING SOM VITENSKAP	55
3.5.2	SAMFUNNSØKONOMISKE LØNNSOMHETSURDERINGER	57
3.5.3	EFFEKTEVALUERING	59

4	EVALUERING AV ITS	63
4.1	KUNNSKAPSUTVIKLING.....	63
4.2	METODISKE UTFORDRINGER I SAMFUNNSØKONOMISKE EVALUERINGER AV ITS.....	65
4.3	METODISKE UTFORDRINGER I EFFEKTEVALUERINGER AV ITS.....	67
4.4	ETABLERTE RAMMEVERK FOR EVALUERING AV ITS	72
4.4.1	<i>FESTA-METODIKKEN</i>	72
4.4.2	<i>EASYWAY GUIDELINES</i>	74
4.4.3	<i>RAMMEVERK FOR VURDERING AV TRAFIKKSIKKERHET</i>	74
4.4.4	<i>RAMMEVERK FOR EVALUERING AV BRUKERAKSEPT</i>	76
4.5	METODEDISKUSJON	77
4.6	BEHOV FOR VIDERE UTVIKLING AV METODEGRUNNLAGET FOR ITS-EVALUERINGER.....	80
5	KAN EN TEORIBASERT TILNÆRMING STYRKE ITS-EVALUERINGEN?	83
5.1	FRA METODESTYRT TIL TEORIBASERT EVALUERINGSSTRATEGI.....	83
5.2	PROGRAMTEORIEN - TILTAKETS TEORETISKE RAMMEVERK.....	85
5.3	TEORIBASERTE EVALUERINGSDESIGN	88
5.3.1	<i>THEORY OF CHANGE</i>	88
5.3.2	<i>REALISTISK EVALUERING</i>	89
5.4	ANVENDELSE AV TEORIBASERT EVALUERING I TRANSPORTFAGET	91
5.5	METODISKE VURDERINGER.....	93
5.6	STYRKER OG SVAKHETER I DEN TEORIBASERTE TILNÆRMINGEN	96
5.7	KUNNSKAPSPRODUKSJON I ET ANVENDT PERSPEKTIV	99
6	CASE A: NONSTOP KLARERING AV TUNGE KJØRETØY	103
6.1	BAKGRUNN.....	104
6.1.1	<i>KORT BESKRIVELSE AV ITS-TILTAKET NONSTOP KLARERING AV TUNGE KJØRETØY</i>	105
6.1.2	<i>KORT BESKRIVELSE AV PROSJEKTET NONSTOP</i>	106
6.2	EKSISTERENDE KUNNSKAP OM NONSTOP KLARERING AV TUNGE KJØRETØY	108
6.2.1	<i>POTENSIELLE EFFEKTER OG VIRKNINGER</i>	109
6.2.2	<i>HVILKE EVALUERINGSMETODER BENYTTES?</i>	115
6.3	OPPRINNELIG EVALUERINGSPLAN I NONSTOP-PROSJEKTET	119
6.4	ERFARINGER FRA PROSJEKTJENNOMFØRINGEN	122
6.5	UTVIKLING AV EN TEORIBASERT EVALUERINGSMETODIKK	124
6.5.1	<i>PROBLEMSTILLINGER</i>	125
6.5.2	<i>PROGRAMTEORI</i>	126
6.5.3	<i>EVALUERINGSKRITERIER</i>	128
6.5.4	<i>EVALUERINGSDESIGN</i>	129
6.5.5	<i>DATAINNSAMLING</i>	130
6.5.6	<i>ANALYSER</i>	135

6.5.7	RESULTATER.....	142
6.6	WHAT IF? POTENSIELLE RESULTATER FRA EN EKSPERIMENTELL EVALUERING	144
6.7	KONKLUSJONER FRA EVALUERINGEN	146
6.8	NY KUNNSKAP OM ITS-TILTAKET	147
6.9	NY KUNNSKAP OM EVALUERINGSMETODIKK	149
6.10	OPPSUMMERING AV CASE A: NONSTOP	153
7	CASE B: SAMKJØRING	155
7.1	BAKGRUNN.....	156
7.1.1	KORT BESKRIVELSE AV ITS-TILTAKET SAMKJØRING.....	157
7.1.2	KORT BESKRIVELSE AV PROSJEKTET EVALUERING AV SAMKJØRING	158
7.2	EKSISTERENDE KUNNSKAP OM SAMKJØRINGSTILTAK.....	158
7.2.1	POTENSIELLE VIRKNINGER, SUKSESSKRITERIER OG BARRIERER.....	161
7.2.2	HVILKE EVALUERINGSMETODER BENYTTES?	164
7.3	UTVIKLING AV EN TEORIBASERT EVALUERINGSMETODIKK	168
7.3.1	PROBLEMSTILLINGER	170
7.3.2	PROGRAMTEORI.....	171
7.3.3	EVALUERINGSKRITERIER.....	173
7.3.4	EVALUERINGSDESIGN.....	174
7.3.5	DATAINNSAMLING.....	176
7.3.6	ANALYSER	177
7.3.7	RESULTATER.....	183
7.4	WHAT IF? POTENSIELLE RESULTATER FRA EN EKSPERIMENTELL EVALUERING	186
7.5	KONKLUSJONER FRA EVALUERINGEN	186
7.6	NY KUNNSKAP OM ITS-TILTAKET	188
7.7	NY KUNNSKAP OM EVALUERINGSMETODIKK	191
7.8	OPPSUMMERING AV CASE B: SAMKJØRING.....	193
8	SAMMENFATNING OG DRØFTING AV RESULTATER.....	197
8.1	ERFARINGER OG RESULTATER FRA TEORIDELEN	197
8.2	ERFARINGER OG RESULTATER FRA CASE-STUDIENE	200
8.3	VURDERINGER AV VITENSKAPELIGHET	205
9	FORSLAG TIL RAMMEVERK FOR EVALUERING AV ITS-TILTAK.....	211
9.1	EN TEORIBASERT EVALUERINGSPROSESS	211
9.2	EKSEMPEL PÅ ANVENDELSE AV EVALUERINGSRAMMEVERKET	225
9.2.1	NORDICWAY.....	225
9.2.2	SMARTFEEDER	228
9.3	HVA KAN RAMMEVERKET TILFØRE FAGOMRÅDET?	230

10	BEMERKNINGER OG REFLEKSJONER	233
10.1	IMPLIKASJONER FOR FAGOMRÅDET	233
10.2	HAR FORSKNINGSSPØRSMÅLET BLITT BESVART?	235
10.3	METODISKE ERFARINGER OG BEGRENSNINGER	238
10.4	VIDERE FORSKNING OG UTVIKLING	240

OVERSIKT OVER FIGURER

<i>Figur 1: Evalueringsmetodikkens oppbygning</i>	<i>7</i>
<i>Figur 2: Prediksjon av overgang fra fossildrevne private kjøretøy til kollektive, elektriske og selvkjørende kjøretøy</i>	<i>15</i>
<i>Figur 3: Gartners hype-kurve for fremvoksende teknologier</i>	<i>16</i>
<i>Figur 4: Konvergens av autonome og samvirkende kjøretøy</i>	<i>22</i>
<i>Figur 5: Kjøreprosessens automatiseringsgrad i seks nivå</i>	<i>23</i>
<i>Figur 6: Evalueringens struktur</i>	<i>36</i>
<i>Figur 7: Det eksperimentelle evalueringsdesignet</i>	<i>48</i>
<i>Figur 8: Tradisjonelt evalueringsløp i transportsektoren</i>	<i>54</i>
<i>Figur 9: Sammenstilling av egenskaper ved klassiske effektstudier og en teoribasert evalueringsmetodikk</i>	<i>85</i>
<i>Figur 10: Skjematisk fremstilling av programteori; den underliggende logikken bak iverksettelsen av et tiltak</i>	<i>86</i>
<i>Figur 11: Eksempel på programteori med utgangspunkt i dette doktorgradsarbeidet</i>	<i>87</i>
<i>Figur 12: Generell tilnærming for effektevalueringer</i>	<i>92</i>
<i>Figur 13: Forenklet skisse av NonStop-systemet</i>	<i>105</i>
<i>Figur 14: Kontrollstasjonen sett ovenfra</i>	<i>123</i>
<i>Figur 15: Evalueringsprosessen i NonStop-prosjektet</i>	<i>125</i>
<i>Figur 16: NonStop-tiltakets programteori</i>	<i>126</i>
<i>Figur 17: Døgnfordeling av kjøretøy med indikasjoner på overlast eller andre forhold</i>	<i>139</i>
<i>Figur 18: Kategorisering av samkjøringsløsninger</i>	<i>159</i>
<i>Figur 19: Beslutningsmodell for vurdering av markedsgrunnlag for samkjøring</i>	<i>165</i>
<i>Figur 20: Evalueringsprosessen i Samkjøringsprosjektet</i>	<i>169</i>
<i>Figur 21: Samkjøringstiltakets programteori</i>	<i>171</i>
<i>Figur 22: Forenklet versjon av the reasoned action model</i>	<i>172</i>
<i>Figur 23: Evalueringens formål for samkjøringstiltaket</i>	<i>174</i>
<i>Figur 24: Rekrutteringsprosessen for samkjøringstiltaket fra deltakerne i Bergenspiloten</i>	<i>178</i>
<i>Figur 25: Vurdering av generelle trafikkreduserende tiltak</i>	<i>180</i>
<i>Figur 26: Motivasjonsfaktorer for deltakelse i samkjøringsprosjektet</i>	<i>181</i>
<i>Figur 27: Ressursinnsats og aktivitetsnivå for samkjøring i Bergenspiloten</i>	<i>182</i>
<i>Figur 28: Årsaker til å slutte med samkjøring</i>	<i>182</i>
<i>Figur 29: Oppdatert atferdsmodell med resultater fra evalueringsstudien</i>	<i>188</i>
<i>Figur 30: Kunnskapsproduksjon som produkt av resultatenes gyldighet og relevans</i>	<i>207</i>
<i>Figur 31: Forslag til teoribasert evalueringsprosess for ITS-tiltak</i>	<i>212</i>
<i>Figur 32: Technology Readiness Scale (TRL) for vurdering av teknologisk modenhet</i>	<i>213</i>
<i>Figur 33: Skjematisk fremstilling av programteori for ITS-tiltak</i>	<i>215</i>
<i>Figur 34: Generell programteori og ulike faser i ITS-tiltakets innføringsprosess</i>	<i>220</i>
<i>Figur 35: Sammenhengen mellom teknologisk modenhet og datainnsamlingsmetoder i NordicWay-prosjektet</i>	<i>222</i>
<i>Figur 36: Eksempel på programteori for den norske C-ITS-tjenesten for varsling av glatt vegbane i NordicWay</i>	<i>226</i>
<i>Figur 37: Programteori utviklet for den finske piloten i NordicWay-prosjektet</i>	<i>227</i>
<i>Figur 38: Innretning og omfang på de fem første pilotene med selvkjørende minibusser i Norge</i>	<i>229</i>
<i>Figur 39: SmartFeeder evalueringsmodell – programteorien for konseptet Smarte tilbringertjenester</i>	<i>229</i>

OVERSIKT OVER TABELLER

<i>Tabell 1: Et utvalg definisjoner av begrepet ITS.....</i>	<i>19</i>
<i>Tabell 2: Egenskaper ved modus 1- og modus 2-forskning.....</i>	<i>56</i>
<i>Tabell 3: Deltema som inngår i konsekvensanalysen iht. Statens vegvesens håndbok V712.</i>	<i>58</i>
<i>Tabell 4: Eksempler på vanlige forsøksdesign i transportevalueringer, rangert etter vitenskapelig styrke.</i>	<i>60</i>
<i>Tabell 5: Indikatorer for brukeraksept i henhold til Van Der Laan et al. (1997).</i>	<i>77</i>
<i>Tabell 6: Teoribasert evaluering som brobygger mellom effekt- og prosessevalueringer.</i>	<i>97</i>
<i>Tabell 7: Rollebeskrivelser i NonStop-prosjektet.....</i>	<i>107</i>
<i>Tabell 8: Opprinnelig evalueringsplan for NonStop-prosjektet.....</i>	<i>120</i>
<i>Tabell 9: Definerte indikatorer og hypoteser i NonStop-prosjektet</i>	<i>121</i>
<i>Tabell 10: Tiltakets underliggende logikk; forventet sammenheng mellom innsats og virkninger</i>	<i>127</i>
<i>Tabell 11: Evalueringskriterier i NonStop-prosjektet</i>	<i>128</i>
<i>Tabell 12: Evalueringsdesign i NonStop-prosjektet</i>	<i>130</i>
<i>Tabell 13: Fordeler og ulemper knyttet til datainnsamling for ulike bruksområder av NonStop-tiltaket.....</i>	<i>132</i>
<i>Tabell 14: Gjennomført datainnsamling i NonStop-prosjektet.....</i>	<i>134</i>
<i>Tabell 15: Trafikktall (tunge kjøretøy) på Sandmoen kontrollstasjon i demoperioden.....</i>	<i>137</i>
<i>Tabell 16: Tidsforbruk ved tungbilkontroll på Sandmoen kontrollstasjon i demoperioden</i>	<i>138</i>
<i>Tabell 17: Grunnlagsdata for nyttekostnadsanalysen.....</i>	<i>140</i>
<i>Tabell 18: Nyttekostnadsanalyse for NonStop klarering av tunge kjøretøy, effekter for transportbransjen</i>	<i>141</i>
<i>Tabell 19: Oversikt over hvilke deler av programteorien som ble adressert i evalueringsstudien</i>	<i>143</i>
<i>Tabell 20: Oppsummering av evalueringsresultater.....</i>	<i>144</i>
<i>Tabell 21: Gjennomsnittlig ventetid i perioder med og uten demo av NonStop-systemet</i>	<i>145</i>
<i>Tabell 22: Konklusjoner fra evalueringsarbeidet.....</i>	<i>147</i>
<i>Tabell 23: Evalueringsgrunnlag for ITS-tiltaket samkjøring.....</i>	<i>171</i>
<i>Tabell 24: Evalueringskriterier for samkjøringstiltaket.....</i>	<i>173</i>
<i>Tabell 25: Evalueringsdesign for samkjøringstiltaket.....</i>	<i>175</i>
<i>Tabell 26: Datainnsamling i evalueringsstudien av Samkjøringstiltaket.....</i>	<i>176</i>
<i>Tabell 27: Oppsummering av evalueringsresultater.....</i>	<i>184</i>
<i>Tabell 28: Suksesskriterier for innføring av tiltaket.....</i>	<i>185</i>
<i>Tabell 29: Evalueringskriterier og konklusjoner fra evalueringen av samkjøring.....</i>	<i>187</i>
<i>Tabell 30: ITS-tiltakenes ulike innretning i de to casestudiene.....</i>	<i>201</i>
<i>Tabell 31: Valg av evalueringsdesign og metode.....</i>	<i>221</i>

1 INTRODUKSJON

Samfunnet gjennomgår store og dyptgripende endringer som følge av mulighetene som ligger i anvendelsen av ny teknologi. Begreper som «den fjerde industrielle revolusjon» beskriver et gryende teknologisk paradigmeskifte, med potensiale til å endre de fleste samfunnssektorer.

Transportsystemet er en viktig del av denne utviklingen. Avanserte teknologiske løsninger som for få år siden syntes å være urealistiske framtidsscenarier, er i ferd med å realiseres. Selvkjørende og samvirkende kjøretøy representerer en fremtid som rykker stadig nærmere. Elektriske kjøretøy overtar for fossildrevet transport. Samtidig introduseres nye forretningsmodeller for organisering av transportløsninger, hvor en tiltagende delingsøkonomi flytter fokuset fra bileierskap til transportbehov og tjenester. Transportområdet er utvilsomt i ferd med å endre seg kraftig. Har vi da de verktøyene som trengs for å henge med på og styre utviklingen i riktig retning?

I dette forskningsarbeidet studerer jeg hvordan bruk av en teoribasert evalueringsmetodikk kan bidra til å styrke kunnskapen og beslutningsgrunnlaget for innføring av intelligente transportsystemer (ITS).

1.1 BAKGRUNN

Transportsektoren er preget av rask teknologisk utvikling på mange felt. Vi ser eksempler i form av stadig mer avanserte biler, hvor kjøretøyteknologi og førerstøttesystemer etter hvert er blitt en del av standardpakken. Vi benytter allerede mobile løsninger for reiseplanlegging og trafikaninformasjon, og det etableres stadig flere teknologibaserte infrastrukturtiltak i form av blant annet trafikkstyring og betalingsløsninger. Samtidig foregår det en storstilt utprøving og pilotering av fremtidens mobilitetstjenester. Selvkjørende kjøretøy har tilbakelagt et betydelig antall kilometer i storskala eksperimentering i USA, og de første selvkjørende bussene er allerede satt i drift som en del av det offentlige transporttilbudet her i Norge. Bilindustrien og tradisjonelle transportaktører utfordres av nye teknologiselskaper som tilbyr fleksible mobilitetstjenester i form av bildeling, taxiløsninger og elektriske utleiesykler. Bak utviklingen finnes en rekke sterke krefter som drar i samme retning; herunder EU-kommisjonen som gir betydelig støtte til forskning, utvikling og innføring av ITS, nasjonale transportmyndigheter, bilindustri og teknologiaktører, samt forbrukerne (trafikanterne) som med sin kjøpekraft bidrar med etterspørsel etter ny teknologi og løsninger som forbedrer deres transporthverdag.

Den overordnede transportpolitiske målsettingen om å utvikle et sikkert transportsystem som fremmer verdiskaping og er omstilt til lavutslippssamfunnet, stiller store krav til transportsektoren (Stortingsmelding 33, 2016-2017). Strategier og offentlige dokumenter viser at forventninger til ITS

som virkemiddel for å oppnå sikre, miljøvennlige og attraktive transportløsninger er sterkt forankret i det politiske miljøet både i Norge og internasjonalt (Directive 2010/40/EU, 2010; European Commission, 2016d; Samferdselsdepartementet, 2010; Statens vegvesen, 2012b; Stortingsmelding 26, 2012-2013; Stortingsmelding 33, 2016-2017). Norge har på flere måter særlige forutsetninger for å lykkes med utvikling og anvendelse av ITS, både fordi vi har et betydelig næringsliv basert på kunnskap og høyteknologi, og fordi vi har en høyt utdannet befolkning som er tidlig ute med å ta i bruk ny teknologi og tilgjengelige IKT-tjenester (European Commission, 2016b). Norge er også rangert helt i toppsjiktet blant land som er godt forberedt på innføring av selvkjørende teknologi i transportsystemet, kun passert av Nederland og Singapore (KPMG, 2019).

Et godt transporttilbud bidrar til livskvalitet for befolkningen og utgjør en viktig forutsetning for næringslivets konkurransekraft og verdiskaping. Samfunnet investerer derfor betydelige økonomiske ressurser i infrastruktur og mobilitetstjenester. Investeringer i transportsektoren har svært lang tidshorison, og beslutninger og teknologivalg som gjøres i dag legger føringer for hvordan transportsystemet vil se ut i mange tiår fremover. Det er derfor helt avgjørende å bygge kunnskap som kan fremskynde innføring av nyttig og lønnsom teknologi, samtidig som man sikrer at investeringer kommer samfunnet og befolkningen til gode. For transportmyndighetene representerer dette et vanskelig dilemma: På den ene siden skal de legge til rette for innovasjon og nytenking gjennom utvikling av en konkurransedyktig ITS-industri. Samtidig kan overdreven teknologioptimisme og forhastede beslutninger resultere i kostbare satsinger på teknologier som viser seg å være mislykket. I motsatt fall risikerer man også å bli sittende for lenge på gjerdet, i påvente av at teknologien skal bli moden. Konsekvensene blir da et lite aktivt innovasjonsmiljø og ITS-løsninger som i mindre grad er tilpasset lokale forhold og felleskapets behov.

Teknologiutviklingen legger grunnlaget for gjennomgripende endringer i transportsektoren, men konsekvensene er ikke entydig knyttet til teknologien i seg selv. Det handler minst like mye om hvordan teknologien tas i bruk og hvordan samfunnet tilpasser seg de nye mulighetene. Løsninger som ikke oppnår tilstrekkelig brukeraksept og tillit hos befolkningen, vil heller ikke utløse ønsket atferdsendring og effekt. For å ta steget videre mot et bærekraftig transportsystem, trenger vi kunnskap som sikrer at myndighetsbeslutninger og investering i infrastruktur bidrar til best mulig måloppnåelse. Denne kunnskapen frembringes blant annet gjennom pilotering og evaluering av lovende ITS-tiltak. Det er derfor behov for effektive metoder for innhenting av relevante data og utvikling av anvendbar kunnskap. På denne måten dannes et grunnlag for strategiske beslutninger som også inkluderer fremtidens intelligente transportsystemer.

1.2 PROBLEMSTILLINGER

Hittil har denne introduksjonen pekt på at nye teknologiske løsninger blir en stadig mer sentral del av transportområdet, og at det er behov for kunnskap om hvordan disse skal anvendes. Brukt på riktig måte er ITS et virkemiddel som kan bidra til bedre fremkommelighet, økt trafikksikkerhet og reduserte miljø- og klimautslipp. Samtidig må man være oppmerksom på nye utfordringer og utilsiktede virkninger. Transportmyndigheter og beslutningstakere må utarbeide gode strategier for å unngå å bli hengende på etterskudd av utviklingen, og samtidig ta ansvar for at løsningene som velges er til det beste for fellesskapet. Det naturlige spørsmålet blir da hvorvidt vi er i stand til å fremskaffe det kunnskapsgrunnlaget som trengs for å ta fornuftige steg mot fremtidens transportsystem? Her kommer problemstillinger knyttet til evalueringsmetodikk inn i bildet.

Et formål med evalueringer er å fremskaffe kunnskap som *øker forståelsen* av hvordan ITS som virkemiddel påvirker trafikanter, transportsystemet og samfunnet for øvrig. Dette er blant annet kunnskap om a) hvilken effekt tiltaket har i størrelse og omfang, b) hvilke virkninger disse effektene har innenfor ulike samfunnsområder, og c) hvordan trafikanter og samfunnet tilpasser seg tiltaket. Forsknings spørsmål som ligger til grunn for slike evalueringer inkluderer blant annet problemstillinger knyttet til i hvilken grad trafikanter vil akseptere og benytte ITS-tjenester, i hvilke situasjoner eller omstendigheter tiltaket forventes å gi effekt, samt hvordan tiltaket bør implementeres for å realisere sitt potensial. ITS er et forholdsvis nytt fagfelt, og derfor finnes det foreløpig lite historisk empiri som dekker bredden av effekter og årsakssammenhenger for alle ulike varianter av ITS-tiltak. Gjennom utprøving og evaluering av ITS fremskaffes erfaringer som gradvis bidrar til et felles akkumulert kunnskapsgrunnlag.

Et annet formål med evalueringer er å fremskaffe kunnskap som grunnlag for *kontroll og styring*. I et slikt perspektiv kan hensikten med evalueringen være å avvise eller legitimere tiltaket, eller gi grunnlag for sammenligning og valg mellom flere tiltak. Dette oppnås gjennom å dokumentere måloppnåelse og effektivitet i prosjektet. I praksis gjøres dette som regel i form av økonomiske analyser, hvor tiltakets virkninger vurderes i et nyttekostnadsperspektiv. Transportsektoren har begrenset økonomisk handlingsrom, og det er mange gode formål som konkurrerer om de samme investeringsmidlene. Beslutningstakere har behov for kunnskap som gjør det mulig å prioritere mellom ulike tiltak, uavhengig om de er basert på ITS eller mer tradisjonelle virkemidler.

Beslutninger i transportsektoren er forankret i en evidensorientert kunnskapsforståelse, som prioriterer empiri fremskaffet gjennom en naturvitenskapelig metodetradisjon. Konsekvensanalyser og samfunnsøkonomiske vurderinger er en viktig del av beslutningsgrunnlaget for offentlige

transporttiltak. De ulike transportetatene har utviklet metodehåndbøker for konsekvensanalyser, med egne modeller for nyttekostnadsberegninger forankret i økonomisk velferdsteori (Odeck & Welde, 2010; Statens vegvesen, 2012a). I disse modellene behandles ITS-tiltak på samme måte som andre infrastrukturtiltak, og til grunn for analysene legges effektanalyser, hovedsakelig basert på et eksperimentelt metodeoppsett med sammenligning av konsekvenser før og etter innføring av tiltaket.

Evaluering av ITS-tiltak har imidlertid vist seg å være krevende. ITS skiller seg fra tradisjonelle vegtiltak på flere måter, noe det gjøres rede for i kapittel 4.2 og 4.3 om metodiske utfordringer knyttet til ITS-evalueringer. Det tradisjonelle metodeverket er innrettet for å finne en veldokumentert og optimal løsning som resultat av grundig utforskning og testing over lengre tid, før innføring av et mest mulig standardisert vegtrafikktiltak. Dette fungerer svært godt når tiltaket er modent, og transportmyndighetene fungerer som en godkjenningsinstans som skal si ja eller nei til innføring av et spesifikt tiltak. Mange ITS-tiltak er derimot gjenstand for kontinuerlig teknologiutvikling som gjør at løsningen forandrer seg raskt og fortløpende gjennom sin levetid. I praksis betyr dette at ITS-tiltak ofte innføres gradvis, noe som gjør det vanskelig å definere klare før- og etterperioder i en evalueringsstudie. Videre er ITS en samlebetegnelse for et mangfold av ulike virkemidler som inkluderer alt fra enkle teknologier til komplekse sammensatte systemer, og gjerne med kompliserte årsakssammenhenger som kan være vanskelig å fange opp med tradisjonell evalueringsmetodikk. Resultatet er at faglitteratur og foreliggende synteserapporter (metaanalyser) viser lite konsistente konklusjoner mellom ulike evalueringsstudier, og ofte er både datagrunnlag og resultater beheftet med stor grad av usikkerhet, som vist i f.eks. Elvik (2011), Vaa, Penttinen, & Spyropoulou (2007), Høye, Elvik og Sørensen (2011), samt ulike ITS effektkataloger.

Motivasjonen for det foreliggende forskningsarbeidet ble sådd på en workshop i regi av Vegdirektoratets etatsprogram Smartere vegtrafikk med ITS (SmITS) i 2012, hvor landets sentrale forskningsmiljøer ble samlet for å drøfte erfaringer med evaluering av ITS-tiltak. Møtet konkluderte med en rekke metodiske problemstillinger, men få løsninger. Innspillene er oppsummert i Lervåg (2012): Et fellestrekk ved ITS-evalueringene er at de i stor grad synes å være styrt av en forhåndsdefinert metodikk, uten fleksibilitet for tilpassing av forskningsdesign til hver enkelt case. Det foregår lite drøfting rundt valg av metode, og det klassiske vitenskapelige eksperimentet med før- og etterundersøkelser er nesten enerådende på feltet. Ofte er dette designet allerede gitt av oppdragsgiver ved bestilling av evalueringoppdraget. Til og med når transportforskeren står fritt til å velge fremgangsmåte, kan det likevel synes som den eksperimentelle metoden anses som en gullstandard man strekker seg mot, også i tilfeller hvor de metodiske forutsetningene for

eksperimentet ikke oppfylles. I mange tilfeller skjer det derfor en praktisk tilpassing av datainnsamling og tillemping av analyser som til slutt kan gjøre det vanskelig å trekke sikre konklusjoner, til tross for at evalueringsrammeverket holder en høy vitenskapelig standard. Erfaringer fra norske ITS-evalueringer omfatter flere tilfeller hvor stor forskningsinnsats belønnes med relativt beskjeden grad av nyvunnet kunnskap, og hvor oppdragsgivers forventninger om entydig beslutningsgrunnlag ikke kan besvares (Gjerstad & Bayer, 2012; Høye, Sørensen, et al., 2011; Thorenfeldt, Bertelsen, & Øvstedal, 2011). Dette bidrar i liten grad til anvendbar kunnskap både for den som skal foreta beslutninger og den som skal iverksette tiltak. I ytterste konsekvens blir resultatet at ITS-løsninger med potensiale for å spare liv, redusere miljøulempen og effektivisere transportsystemet ikke blir tatt i bruk fordi evalueringsstudiene ikke makter å dokumentere effekter og lønnsomhet, som pekt på i Odeck og Welde (2010).

Med dette bakteppet er det grunn til å stille spørsmål om hvorvidt det finnes andre evalueringsstrategier som er bedre egnet til å fremskaffe kunnskapsgrunnlaget som trengs for å vurdere effekter og foreta fornuftige beslutninger om ITS-tiltak. Det er begrenset erfaring med utprøving av alternative metodiske tilnæringer i dag, og det er behov for å fylle dette kunnskapsgapet ved å videreutvikle det teoretiske og metodiske fundamentet for evaluering av ITS.

Målsettingen med dette forskningsarbeidet er å bidra til økt kunnskap om og forståelse for metodegrunnlaget for ITS-evalueringer. Dette har utspring i et ønske om at fremtidige evalueringsstudier skal gi mer anvendbare resultater, og dermed bidra til et bedre fundert beslutningsgrunnlag på veien mot fremtidens transportsystem. For å oppnå dette er det nødvendig å se nærmere på hva slags metodikk som legges til grunn for ITS-evalueringer i dag, og hvordan denne er forankret i generell evalueringsteori. I dette ligger det også en vurdering av hvorvidt dagens tilnærming er i stand til å oppfylle kunnskapsbehovet knyttet til innføring av ITS-tiltak. Kanskje finnes det andre metodiske innfallsvinkler som har potensial til å øke nytten av evalueringer innenfor ITS-faget? Viktige delmål i prosjektet blir dermed å frembringe kunnskap som styrker forskeren og oppdragsgiveren i metodiske valg i fremtidige evalueringer. Resultatet av denne forskningen kan åpne for en større grad av differensiering av evalueringsperspektiver, og derigjennom øke ITS-evalueringenes spillerom og anvendelsesområde.

Disse betraktningene har ført frem til forskningsspørsmålet som ligger til grunn for avhandlingsarbeidet:

Kan man med utgangspunkt i generell evalueringsteori utvikle eller tilpasse en evalueringsmetodikk som bidrar til å styrke kunnskapen og beslutningsgrunnlaget for implementering av ITS i transportsystemet?

Forskningen konkretiseres gjennom studier av følgende problemstillinger:

- a) *Hvilken metodikk legges til grunn for evalueringer av ITS i dag? Og hvordan er denne metodikken forankret i generell evalueringsteori?*
- b) *I hvilken grad er dagens evalueringstilnærming i stand til å oppfylle kunnskapsbehovet knyttet til implementering av ITS-tiltak?*
- c) *Finnes det andre innfallsvinkler som kan øke nytten av ITS-evalueringer? Hvordan kan dette eventuelt bidra til å styrke kunnskapen om og beslutningsgrunnlaget for implementering av ITS i transportsystemet?*

Hensikten med forskningsarbeidet er å utvide forståelsen av det metodiske mulighetsrommet i fremtidige ITS-evalueringer, gjennom drøfting av teorigrunnlag og praktisk utprøving i evalueringsstudier. På sikt vil resultatene kunne bidra til arbeidsmetoder som sikrer et bedre kunnskapsgrunnlag for beslutningstakere i transportsektoren.

1.3 METODE

Forskningsspørsmålet og problemstillingene er undersøkt og besvart gjennom en kombinasjon av *litteraturstudier* som har gitt kunnskap om tidligere forskning og teoretisk grunnlag for fagområdet, og *empiriske studier* som har medvirket til praktisk innsikt i problemområdet.

Litteraturstudiene er gjennomført med fokus på to hovedtema; a) intelligente transportsystemer som virkemiddel i transportsystemet og b) evalueringsteori. Hensikten med det første litteraturstudiet har vært å kartlegge det metodiske grunnlaget for evaluering av ITS-tiltak, samt avdekke hvorvidt særskilte egenskaper ved ITS som virkemiddel har betydning for valg av evalueringsmetodikk og kunnskapsproduksjon innenfor fagfeltet. Deretter har en gjennomgang av evalueringsteori gitt innblikk i det teoretiske grunnlaget for evalueringer generelt, samt hvilke tradisjoner og perspektiver som finnes innenfor ulike fagdisipliner. Til grunn for arbeidet lå et ønske om å teste ut en eller flere metodiske tilnærminger som kan gi positive bidrag til forskningen på ITS.

Litteratursøk har inkludert vitenskapelige databaser som Science Direct/Elsevier og SpringerLink, samt generelle litteratursøk i Oria (bibliotekets ressurser) og Google Scholar. Videre er det søkt i relevante offentlige publikasjonsdatabaser, eksempelvis fra EU, OECD og transportmyndigheter. Typiske søkeord har inkludert ulike kombinasjoner av ordene *ITS*, *Intelligent transport system*,

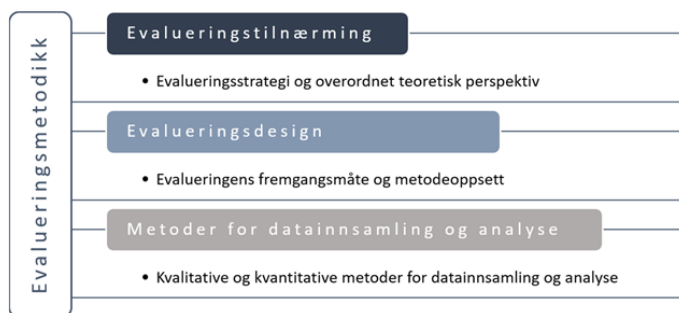
evaluation, methodology, theory, impact, assessment, analysis og method (engelsk) og *ITS, intelligente transportsystemer, evaluering, vurdering, teori, metode, effekt, nytte og virkning* (norsk). Siden ITS er et forholdsvis nytt og dynamisk fagfelt, var det forventet at en god del litteratur ville finnes som konferansepublikasjoner eller tilsvarende akademiske tekster på Google Scholar. Det er derfor gjort supplerende søk i relevante konferansedatabaser, som Association for European Transport: European Transport Conference Proceedings og ITS World Congress. Den generelle evalueringsteorien har en lengre tradisjon med faglige diskurser publisert i vitenskapelige tidsskrift, og hvor ulike teoretiske og metodiske perspektiver også er presentert i lærebøker.

Litteraturstudiene har gitt et teoretisk grunnlag for å utvikle eller tilpasse en evalueringsmetodikk som synes egnet til å fremskaffe relevant og anvendbar kunnskap om ITS. Empirisk utprøving av evalueringsmetodikken i to konkrete case-studier med ITS-tiltak har gitt praktisk innsikt i evalueringsarbeidet. Erfaringer fra case-studiene har således gitt grunnlag for en kvalitativ vurdering av hvorvidt evalueringsprosessen er formålstjenlig, i tillegg til å synliggjøre eventuelle begrensninger og behov for justeringer.

Metodiske erfaringer og begrensninger i doktorgradsarbeidet er drøftet nærmere i kapittel 9.3.

1.4 AVGRENSNING OG BEGREPSAVKLARING

Evaluering av intelligente transportsystemer favner et vidt spekter av aktiviteter som involverer ulike former for vurdering av teknologibaserte løsninger i transportsektoren. Denne studien er avgrenset til ITS-tiltak i vegtransportssystemet. Evalueringsbegrepets betydning varierer også vesentlig avhengig av kontekst og fagområde. Det finnes ingen entydig definisjon, og ulike varianter av begrepets bruk og meningsinnhold er nærmere diskutert i kapittel 3. Det er lagt til grunn en begrepsforståelse som innebærer at den konkrete *evalueringsmetodikken* er en operasjonell beskrivelse av hvordan den enkelte evalueringen skal gjennomføres, med utgangspunkt i en overordnet *evalueringstilnærming* (evalueringstrategi og vitenskapelig forankring) og med et bestemt *evalueringdesign* (fremgangsmåte) og *spesifikke metoder for datainnsamling og analyse*. Alle disse nivåene omtales og drøftes i avhandlingen, men hovedfokus er rettet inn mot den overordnede tilnærmingen for evaluering av ITS-tiltak.



Figur 1: Evalueringsmetodikkens oppbygning

I arbeidet med å utvikle en evalueringsmetodikk for ITS-tiltak, er det tatt utgangspunkt i fagområdets behov for relevant kunnskap om hvordan ITS kan bidra til å oppfylle transportpolitiske målsettinger om et effektivt, sikkert og miljøvennlig transportsystem i fremtiden. Utgangspunktet for dette arbeidet har vært norske transportmyndigheters behov for evaluering som gir kunnskap i en tidlig fase av innovasjonsløpet, for å redusere feilinvesteringer og fremskynde innføring av lovende teknologi. Dette innebærer at det er lagt vekt på evaluering som et verktøy for kunnskapsbygging og læring, og i mindre grad som instrument for bedømmelse og kontroll. Videre er det lagt til grunn et anvendt forskningsperspektiv, med vekt på å finne praktiske og anvendelige løsninger. Til sist er hovedfokus rettet mot *effektstudier av ITS-tiltak*, og avhandlingen går ikke i dybden på vurdering av metodikk for samfunnsøkonomiske lønnsomhetsvurderinger.

En sentral del av doktorgradsarbeidet omfatter utprøving av en *teoribasert tilnærming* til evaluering av ITS. Innenfor evalueringslitteraturen er begrepet *teoribasert evaluering* en etablert benevnelse på en evalueringsmodell som innebærer at evalueringen er strukturert rundt tiltakets programteori og med en mekanistisk kausalitetsforståelse¹. Ordet *teori* henspiller altså på tiltakets logikk og ikke på en overordnet teoretisk forankring. I denne avhandlingen brukes begrepet *teoribasert* utelukkende som fagbenevnelse på dette spesifikke evalueringsperspektivet, og det hentyder *ikke* at andre evalueringsmodeller som omtales mangler en teoretisk forankring.

1.5 STRUKTUR

Forskningsarbeidet inkluderer en teoretisk redegjørelse av intelligente transportsystemer som virkemiddel i transportsystemet, samt en gjennomgang av relevante evalueringsperspektiv og metoder. På bakgrunn av innledende litteraturstudier ble det tatt utgangspunkt i en teoribasert evalueringsmetodikk, som ble tilpasset og prøvd ut som fremgangsmåte i to konkrete case-studier. Erfaringer og resultater fra disse teoretiske og praktiske studiene har gitt grunnlag for utvikling av et rammeverk for teoribasert evaluering av ITS-tiltak.

Avhandlingen har følgende struktur:

Kapittel 2 presenterer hvordan intelligente transportsystemer har blitt et sentralt virkemiddel i transportsystemet. Det gjøres rede for begrepsbruk og teknologien som ligger til grunn for relevante ITS-tjenester.

¹ Den teoribaserte evalueringsmodellen er nærmere forklart i kapittel 5.

Kapittel 3 presenterer det teoretiske grunnlaget for evalueringer generelt, med tradisjoner og perspektiver innenfor ulike fagfelt, etterfulgt av en oversikt over det metodiske grunnlaget for evalueringer i transportsektoren.

Kapittel 4 gir et innblikk i hvordan ITS håndteres innenfor dagens metodiske rammeverk for evaluering av veg- og trafikktiltak. Til tross for at ITS allerede er en viktig del av transportsystemet, mangler det en felles etablert evalueringsmetodikk for ITS. Kapittelet konkluderer med fem metodiske anbefalinger for evaluering av ITS-tiltak.

Kapittel 5 gir en grundigere innføring i *det teoribaserte evalueringsperspektivet* med beskrivelse av fremgangsmåte, eksempler på teoribaserte metoder, og drøfting av den teoribaserte tilnærmingens styrker og svakheter.

Kapittel 6 presenterer en case-studie med evaluering av ITS-tiltaket *NonStop klarering av tunge kjøretøy*. Innledningsvis presenteres rammene for evalueringen, med beskrivelse av tiltaket og eksisterende kunnskap på fagfeltet. Videre gjøres det rede for evalueringsprosessen og erfaringer med denne, samt analyser, resultater og konklusjoner. Til slutt drøftes hvordan case-studiet har bidratt med ny kunnskap; både om effekter av det konkrete ITS-tiltaket og om evalueringsmetodikk for ITS-tiltak generelt.

Kapittel 7 presenterer en case-studie med evaluering av ITS-tiltaket *Samkjøring*, og følger samme struktur som forrige kapittel: Innledningsvis presenteres rammene for evalueringen, med beskrivelse av tiltaket og eksisterende kunnskap på fagfeltet. Videre gjøres det rede for evalueringsprosessen og erfaringer med denne, samt analyser, resultater og konklusjoner. Til slutt drøftes hvordan case-studiet har bidratt med ny kunnskap; både om effekter av det konkrete ITS-tiltaket og om evalueringsmetodikk for ITS-tiltak generelt.

Kapittel 8 oppsummerer og drøfter erfaringer og resultater fra foregående kapitler, basert på det teoretiske grunnlaget for ITS-tiltak og evaluering som fagdisiplin, samt den praktiske utprøvingen av en teoribasert evalueringsprosess i to konkrete case-studier. Arbeidet danner grunnlaget for anbefaling om å ta i bruk en teoribasert metodikk for evaluering av ITS-tiltak.

Kapittel 9 presenterer et forslag til rammeverk for evaluering av ITS-tiltak med utgangspunkt i en teoribasert tilnærming. Rammeverket gir en trinnvis veiledning for utvikling av en konkret evalueringsmetodikk for ITS-tiltak som inkluderer 1) vurdering av teknologisk modenhet, 2) behovsanalyse, 3) etablering av programteori, 4) utvikling av evalueringskriterier, 5) valg av evalueringsdesign, 6) datainnsamling, 7) analyser og 8) rapportering av resultater.

Kapittel 10 avrunder avhandlingsarbeidet med noen bemerkninger og refleksjoner rundt den foreslåtte evalueringsmetodikken og kunnskapen som er etablert.

Avhandlingen dokumenterer et omfattende forskningsarbeid med en overordnet teoretisk gjennomgang av intelligente transportsystemer og generell evalueringsteori, samt fordypning i metodegrunnlaget for ITS-evalueringer og det teoribaserte evalueringsperspektivet. Det er lagt til rette for at hvert kapittel skal kunne leses separat, med oppsummeringer og sammenfatninger underveis. Dette innebærer at hovedpoengene fra drøftingen av teorigrunnlaget er gjort rede for flere ganger: Kapittel 5 gir en sammenfatning av teorigrunnlaget i foregående kapitler, og utdyper begrunnelsen for valg av det teoribaserte evalueringsperspektivet i de etterfølgende case-studiene. Kapittel 8 oppsummerer og drøfter resultater og erfaringer både fra teoridelen og fra den praktiske utprøvingen av en teoribasert evalueringsmetodikk i to konkrete ITS-evalueringer. Hovedresultatet fra doktorgradsarbeidet er presentert i kapittel 9: En teoribasert metodikk for evaluering av ITS-tiltak.

2 INTELLIGENTE TRANSPORTSYSTEMER

I dette kapitlet gjøres det rede for et viktig premiss for dette doktorgradsarbeidet:

Transportsektoren føyer seg inn i en rekke av samfunnsområder som står overfor gjennomgripende endringer som følge av mulighetene som ligger i digitalisering, sensorteknologi og kunstig intelligens. Evaluering av innovative tiltak og tjenester er derfor helt avgjørende i arbeidet med å utvikle gode og bærekraftige transportløsninger for fremtiden.

2.1 DIGITALISERING AV HVERDAGEN

Samfunnet gjennomgår en sterk teknologisk utvikling. Digitale løsninger og sensorteknologi har på relativt kort tid blitt en integrert del av hverdagen for mange av oss. Vi sjekker rutetidene til bussen, gjennomfører elektroniske betalinger, leser nyheter og strømmer musikk fra internett på våre smarttelefoner. Bransjer som musikk, varehandel, fotografi og media er bare noen eksempler på industrier som på kort tid har erfart radikale strukturelle endringer som påvirker både forbrukernes atferd og grunnlaget for forretningsmodeller. Den teknologiske utviklingen berører flere og flere samfunnsområder i stadig økende tempo. Det hevdes at vi er på vei inn i et teknologisk paradigmeskifte med potensiale til å endre hvordan vi ser på verden og ikke minst hvordan vi forholder oss til den (Carlin, 2015; Finansdepartementet, 2016; Kjensli, 2016; Skryseth, 2016; World Economic Forum, 2015, 2016). Da verdens toppledere fra politikk, akademia og forretningsliv møttes til World Economic Forums årlige konferanse i Davos i 2016, var temaet for konferansen: *Mastering the forth industrial revolution*. Begrepet *den fjerde industrielle revolusjon* plasserer vår samtid inn i en rekke av historiske industrielle paradigmeskifter som har endret de sosiale, politiske og økonomiske strukturene i samfunnet (Schwab, 2016);

- *Den første industrielle revolusjon* på slutten av 1700-tallet var et resultat av tekniske nyvinninger knyttet til dampkraft og mekanisk produksjon, og medførte en overgang fra landbrukssamfunnet til industrisamfunnet. Innenfor transportområdet ga dette blant annet utfall i form av dampskip og senere damplokomotiv som kunne frakte mennesker og gods effektivt over lengre avstander.
- *Den andre industrielle revolusjon* tok form mot slutten av 1800-tallet med elektrifisering og masseproduksjon som utløsende teknologier. Forbrenningsmotoren ble utviklet og bruk av produksjonsmaskineri medførte en overgang fra små, lokale bedrifter til større fabrikklokaler og bruk av samlebånd i produksjonslinjen. Dette åpnet blant annet for storskalaproduksjon av biler, med Henry T. Fords Modell T som det mest kjente eksempelet fra denne tiden.

- *Den tredje industrielle revolusjon* tidfestes fra begynnelsen av 1970-årene og frem til i dag, med elektronikk og IT som grunnlag for automatisering av produksjonsprosesser. Informasjonsteknologien har medført enorme forandringer i folks hverdagsliv og hvordan vi kommuniserer med hverandre, med økt tilgang til PC, mobiltelefon, internett, forbrukerelektronikk og nye tjenester. I transportsektoren har ITS gjort sitt inntog med teknologibasert infrastruktur og stadig mer avanserte kjøretøy.
- *Den fjerde industrielle revolusjon* forventes å få sitt utløp som et resultat av digitalisering, datakraft, kunstig intelligens og sensorteknologi, hvor teknologien i større grad innrettes mot å foreta databaserte valg eller beslutningsstøtte. Alternative benevnelser for å beskrive den pågående teknologiske omveltningen inkluderer begreper som *Internet of Things*, *digitaliseringsbølgen* og innenfor transportområdet også *Mobilitet 4.0*. Mens de tidligere industrielle revolusjoner har blitt sett på som lineære, antas denne å følge en eksponentiell vekst, som gjør at teknologiske nyvinninger og trender utvikles med akselererende hastighet og større utbredelse enn man har opplevd tidligere i historien (Schwab, 2016; World Economic Forum, 2015).

Taktskiftet og hastigheten i endringene vi står overfor er krevende for myndigheter og premissgivere. Det er vanskelig å overskue virkninger og samfunnsmessige tilpasninger. Strategiske planer og beslutninger må ofte iverksettes parallelt med at endringene skjer – også når beslutningsgrunnlaget er usikkert. Ideologiske ståsted påvirker også oppfatningen av hvordan potensialet i teknologien skal realiseres til det beste for oss som samfunn og individer. Fremtidsforskere baserer ofte sine vurderinger på studier av megatrender som beskriver sannsynlige tendenser i et 10-15-årsperspektiv (ERTRAC, 2013; G. Larsen, 2006). Stiftelsen World Economic Forum (2015) har identifisert seks megatrender som forventes å gi store samfunnsmessige konsekvenser i forholdsvis nær fremtid:

1. *Vår digitale tilstedeværelse*: Den digitale tilgjengeligheten endrer våre relasjoner til hverandre og omgivelsene. Dette forsterkes ytterligere av at digitale komponenter integreres i klær og såkalte smartwear som sørger for kontinuerlig registrering av data om vår helse og atferd.
2. *Ubegrenset datakraft, kommunikasjon og lagringskapasitet*: Utviklingen av datamaskiner med stadig større datakraft på mindre fysiske enheter og til en rimelig pris, gjør at vi får nærmest ubegrenset tilgang til informasjon, regnekraft og lagringskapasitet blant annet gjennom våre mobiltelefoner.

3. *Internet of things (tingenes internett)*: Alt kobles på internett via små, billige og smarte sensorer. Dette gjør at fysiske komponenter som klær, husholdningsapparater, transportsystemet, byer, energiforsyning og produksjonsprosesser kan utveksle informasjon med sine omgivelser.
4. *Kunstig intelligens og big data*: Økt sensorbruk og internettoppkobling vil resultere i at det samles inn store mengder data som grunnlag for beslutningstaking og etterhvert også maskinlæring.
5. *Delingsøkonomi og distribuert tillit*: Internett og sosiale media har medført muligheter for å etablere og organisere effektive nettverk for deling av tjenester, samt systemer av verifiserte protokoller (blockchains) som holder oversikt over transaksjoner av verdier og sensitiv informasjon. Nettverkene verdi øker jo flere deltakere de oppnår, og utfordrer dermed tradisjonelle hierarkier og institusjoner ved å tilrettelegge for individuell organisering og helt nye forretningsmodeller.
6. *Digitalisering av materie*: Printing av fysisk materie (bl.a. 3D-printing) påvirker vareproduksjon og åpner muligheter for komplekse konstruksjoner innenfor blant annet helsesektoren.

Disse megatrendene påvirker et bredt spekter av samfunnsområder. Konsekvensene dette får for transportområdet og oss som trafikanter, avhenger av hvordan vi som samfunn velger å møte disse endringene. Myndighetene må ta stilling til hvordan ny teknologi skal implementeres og hvilke tjenester som skal støttes gjennom tilrettelegging av infrastruktur og regelverk. Eventuelle fremskritt forutsetter aksept hos brukerne ved at nytteverdien av teknologiske innovasjoner synliggjøres. På kort sikt kan man forvente å oppnå stegvise forbedringer av transporttjenestene vi allerede kjenner, mens det i et lengre perspektiv kan påvirke hele vår forståelse og anvendelse av transportsystemet. Dette gjør det vanskelig å forutsi langsiktige konsekvenser av de teknologiske nyvinningene, både for premissgiverne og hver enkelt av oss. I 2015 identifiserte World Economic Forum (2015) 21 sannsynlige teknologiskifter i samfunnet og predikerte såkalte *tipping points*, et kritisk tidspunkt hvor disse forventes å inntreffe. Nedenfor presenteres et utvalg på fire av disse milepælene som forventes å få avgjørende betydning innenfor transportsektoren;

- *Det forventes at 1 trillion² sensorer er tilkoblet internett innen år 2022.* Fysiske enheter vil være "smarte" og ha muligheten for kommunikasjon med sine omgivelser. Dette antas å få vesentlig

² Det legges til grunn at publikasjonen benytter den britiske betydningen av begrepet (ettall med 18 nuller bak) og ikke den amerikanske (ettall med 12 nuller bak). Uansett er det et enormt høyt tall som kan være vanskelig å fatte omfanget av innenfor den menneskelige evnen til mengdeforståelse.

betydning for trafikkikkerhet, effektivitet i varetransport (logistikk), overvåkning og kontroll (trafikkstyring) og utnyttelse av infrastruktur og kjøretøy.

- *Det forventes at flere bilreiser foregår med bildeling enn med private biler innen år 2025.* Delingsøkonomien vil få en enorm vekst basert på mulighetene som ligger i teknologiske plattformer for formidling og organisering av varer og tjenester. Dette vil blant annet påvirke behovet for å eie egen bil og flytte fokuset mot mobilitet som et tjenestetilbud.
- *Det forventes at vi får den første byen³ uten trafikklys innen år 2026.* Såkalte Smart Cities vil ha vegnettet koblet til internett med datastyrt håndtering av energi, logistikk og trafikk. Dette antas å ha positiv betydning for innbyggernes livskvalitet, med økt mobilitet og redusert luftforurensning, samt å gi bedre utnyttelse av transportressursene. De negative konsekvensene innbefatter blant annet en forverring av personvernet og økt sårbarhet for cyber-angrep.
- *Det forventes at selvkjørende biler utgjør 10 % av alle bilene på det amerikanske vegnettet innen år 2026.* De selvkjørende bilene vil opptre med større trafikkikkerhet og bedre effektivitet enn de menneskestyrte bilene og medføre en reduksjon i trafikkulykker, bilkøer og forurensning. På overordnet nivå antas det at eksisterende modeller for transport og logistikk endres, blant annet med elektrifisering av bilparken, redusert bilhold og økt mobilitet. På den negative siden vil mange tradisjonelle jobber innenfor transportsektoren gå tapt, og transportsystemet vil få økt sårbarhet for terror og cyberangrep.

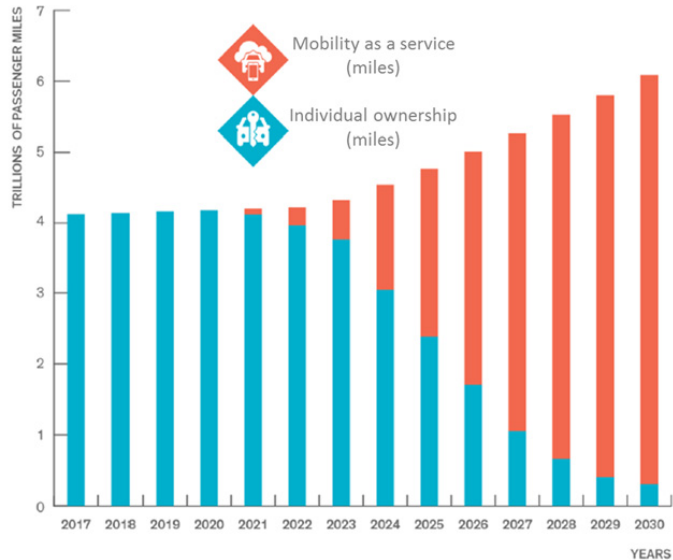
Fem år etterpå fremstår teknologiene fortsatt som høyst relevante, selv om tidsperspektivet kan synes noe urealistisk.

Det finnes en rekke ekspertanalyser med framtidsscenarier for transportsektoren (Arbib & Seba, 2017; Jordbakke, Salte, & Hammer, 2017; Little, 2018; McKinsey & Bloomberg, 2016). Selv om fremtidsbildene som tegnes opp varierer og tidsanslagene spriker, synes det å være generell enighet om at vi står overfor store og fundamentale endringer. Et eksempel er vist i Figur 2, hentet fra Arbib og Seba (2017) som analyserer hvordan teknologi knyttet til selvkjørende kjøretøy vil påvirke og trigge en rekke strukturelle forandringer i samfunnet. Studien konkluderer med at selvkjørende teknologi vil redusere prisen på transport betraktelig, noe som vil endre både eierskapsmodeller og trafikantenes reisemiddelvalg.

³ Større enn 50.000 innbyggere

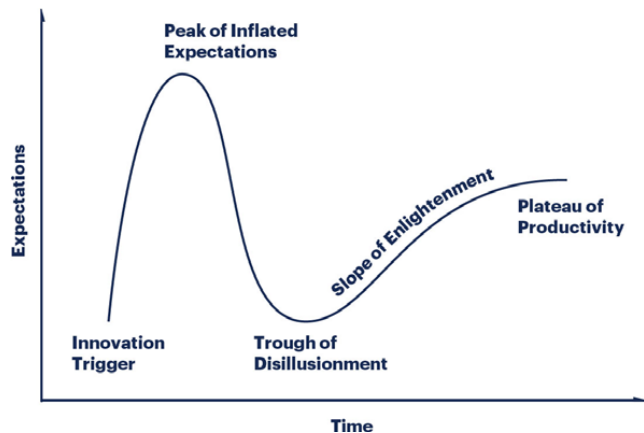
Figur 2: Prediksjon av overgangen fra tradisjonelle fossildrevne private kjøretøy til kollektive, elektriske og selvkjørende kjøretøy.

Passasjertransport (miles) per år. Eksempel basert på data fra US Department of Transportation (Arbib & Seba, 2017).



Analysen konkluderer med at fra det tidspunktet lovverket åpner for ordinær drift av selvkjørende kjøretøy på offentlig veg (hypotetisk satt til år 2021 i eksemplet), vil det i løpet av en tiårsperiode skje en fullstendig transformasjon hvor kollektive, elektriske og selvkjørende kjøretøy står for 95 % av transportarbeidet på vegen. Figuren kan neppe leses som en fasit på fremtiden, men den er en illustrasjon på at når teknologien og samfunnet er modent, kan endringene skje raskt og være omfattende. Det avgjørende momentet i dette tilfellet er at de kollektive mobilitetstjenestene (mobility as a service) blir vesentlig billigere enn å benytte en privateid fossildrevet bil, selv om denne allerede står ferdig nedbetalt i garasjen. I et slikt scenario, vil markedet for både brukte og nye fossilbiler forsvinne forholdsvis brått. Lignende simuleringer av overgangen fra privateide personbiler til kollektive mobilitetstjenester i byområder er dokumentert i International Transport Forum (2017b) og International Transport Forum (2017a) for byene Lisboa og Helsinki. Disse analysene indikerer at ved å erstatte dagens privatbilreiser med kollektiv bestillingstransport, kan man dekke reisebehovet med 3-4 % av bilparken. I tillegg kuttes CO₂-utslippene vesentlig, samtidig som det meste av parkeringsarealet frigjøres til andre formål. Slike fremtidsscenarier gir enorme implikasjoner for planlegging av fremtidens byer. I Jordbakke et al. (2017) presenteres også fire mulige scenarier for hvordan selvkjørende kjøretøy vil påvirke mobiliteten i Oslo og Akershus, avhengig av hvordan transporttjenestene reguleres og organiseres.

Gartners hype-kurve (Linden & Fenn, 2003) illustrerer nye teknologiers modenhet og fremvekst, som vist i Figur 3: Teknologisk gjennombrudd og tidlige prototyper trigger interesse fra media og publikum. Toppen av hype-kurven nås etter hvert som flere tidlige brukere tester ut og får erfaring med teknologien. Økt utprøving bidrar også til å avdekke begrensninger i funksjonalitet, og positiv publisitet



Figur 3: Gartners hype-kurve for fremvoksende teknologier.

erstattes av mediainnslag om feil og mangler ved teknologien. Interessen avtar, og det blir vanskeligere å få finansiering til videre utvikling. Samtidig må produktet forbedres for å oppnå tilstrekkelig aksept blant potensielle kunder, og 2. og 3. generasjons løsninger blir tilgjengelig. Markedet får etter hvert økt forståelse for teknologiens nytteverdi og forventningskurven er igjen stigende. Endelig nås et nivå hvor teknologien er tilstrekkelig moden for opptak i større brukergrupper. Forretningsmodeller blir klarere definert og markedsutbredelsen øker. I Gartners analyse fra 2015 var selvkjørende kjøretøy (autonomous vehicles) plassert på toppen av hype-kurven, med anslått opptak i de store brukergruppene i løpet av 5-10 år (Gartner, 2015). I 2019 er forventningsnivået betydelig korrigert: Full automatisering (SAE nivå 5) har rykket noe tilbake i modenhetsprosessen, mens selvkjøring på nivå 4 befinner seg på den nedadgående hype-kurven. Begge teknologiene anslås å bruke mer enn 10 år før de er tilgjengelig for massemarkedet (Gartner, 2019).

Det er vanskelig å predikere hvordan fremtiden vil se ut, og ingen kan forutsi nøyaktig hva som kommer og når det vil skje. Fremtidsbildene som er skissert uttrykker imidlertid ekspertenes forventninger om utvikling i de kommende tiårene, snarere enn utopier. Hvis prediksjonene holder mål, står vi ovenfor radikale endringer både i transportsystemet og samfunnet for øvrig. Hvorvidt vi som samfunn er i stand til å påvirke utviklingen i ønsket retning, i stedet for å passivt innordne oss i etterkant, avhenger av hvordan vi møter fremtiden. Teknologien står ikke alene, den er avhengig av brukeraksept, forretningsmodeller og et tilpasset regelverk. Et stikkord som går igjen i drøftinger av teknologiutviklingens samfunnsmessige konsekvenser, er det engelske begrepet *resilience*, som kan oversettes med *robusthet*, *elastisitet*, *motstandskraft*, *livskraft* og *utholdenhet*. Innenfor psykologien brukes begrepet om evnen til å håndtere stress og kriser. Fleksibilitet og evne til reorganisering vil være viktige krav til både myndigheter og infrastruktur i tiden fremover (Copenhagen Institute for

Future Studies, 2016; ERTRAC, 2013). Dette følges også opp av EU som i arbeidet med sitt 9. rammeprogram for perioden 2021-2027, understreker viktigheten av å møte en kompleks og hittil ukjent fremtid med fleksible strategier og en endringsorientert forsknings- og innovasjonspolitik (Weber et al., 2018).

Dette delkapittelet har pekt på hvordan samfunnet vårt er i ferd med å endres fundamentalt som følge av den pågående teknologiutviklingen og digitaliseringen av stadig flere bransjer og fagområder. Dette gir nye muligheter til å skape gode løsninger og økt livskvalitet for befolkningen, men bringer også med seg en del nye utfordringer. I neste delkapittel gjøres det rede for hvordan teknologiutviklingen påvirker transportsektoren, hvor intelligente transportsystemer og nye mobilitetstjenester er i ferd med å bli et stadig viktigere virkemiddel for å oppnå transportpolitiske målsettinger om et sikkert, effektivt og miljøvennlig transportsystem.

2.2 ITS SOM VIRKEMIDDEL I TRANSPORTSEKTOREN

2.2.1 TRANSPORTPOLITISKE RETNINGSLINJER OG STRATEGIER

Et velfungerende transportsystem er en forutsetning for et moderne samfunn. Næringslivet og økonomien er avhengig av en funksjonell infrastruktur, og mobilitetstilbudet er avgjørende for befolkningens bosetning, arbeidsmarked og bevegelsesfrihet. Dagens transportsystem er imidlertid ikke bærekraftig (Stortingsmelding 33, 2016-2017). Transportsektoren står for en betydelig og stigende andel av verdens klimagassutslipp, i tillegg til å forårsake lokal forurensning, ulykker, barrierer og støyproblematikk. Ifølge OECD (2016) genererer transportsektoren 23 % av det globale CO₂-utslippet fra drivstofforbrenning (tall fra 2013). I Norge er transport den største kilden til klimagassutslipp, og vegtransport står for over halvparten av utslippene (Miljødirektoratet, 2016). Hvert år går i størrelsesorden 150 menneskeliv tapt på norske veier, til tross for at vi er blant de mest trafikksikre landene i verden (OECD, 2015). På verdensbasis dør omtrent 1,35 millioner personer i trafikken hvert år (WHO, 2018). I tillegg til dødsfallene, regner man med at omtrent fire ganger så mange pådrar seg permanente invalidiserende skader, og åtte ganger så mange blir alvorlig skadd i trafikkulykker årlig. Ved siden av miljøulemper og trafikkulykker fører også overbelastning av vegnettet til kø og kapasitetsproblemer i områder med tett befolkning. Transportøkonomisk institutt har beregnet at de marginale eksterne kostnadene ved vegtrafikk i Norge utgjør omtrent 24 milliarder kroner i året (tall fra 2011), og da er ikke klimagassutslipp kalkulert inn i regnestykket (Thune-Larsen, Veisten, Rødseth, & Klæboe, 2014). Ulykkeskostnader står for det største bidraget (42 %) etterfulgt av kostnader knyttet til lokale utslipp, kø, drift (særlig vinterdrift), slitasje og støy. Det kan altså synes paradoksalt at vi på den ene siden legger til rette for

en voksende transportetterspørsel, samtidig som transportsektoren påfører oss store lokale og globale utfordringer.

Det finnes etter hvert betydelig dokumentasjon på at ITS forventes å spille en viktig rolle i omleggingen til et sikkert, effektivt og miljøvennlig transportsystem (Directive 2010/40/EU, 2010; European Commission, 2011, 2016c; KPMG, 2018; Samferdselsdepartementet, 2010; Statens vegvesen, 2012b; Stortingsmelding 26, 2012-2013; Stortingsmelding 33, 2016-2017). EUs transportpolitiske strategier er rettet inn mot å oppnå *økt konkurransekraft* gjennom et felles europeisk transportmarked og *bærekraftig mobilitet* gjennom utslippsreduksjoner, økt trafiksikkerhet og et effektivt og brukervennlig transportsystem. Det pekes på løsninger i form av innovative teknologier som optimerer kapasitet og utnyttelse av infrastrukturen, samt utvikling av tjenester som legger til rette for en mer bærekraftig forbrukeratferd (European Commission, 2011).

Norske transportplaner støtter i hovedsak opp om de europeiske retningslinjene. Økt bruk av ITS er en overordnet prioritering for å oppnå politiske målsettinger om et sikkert og effektivt transportsystem som er omstilt til lavutslippssamfunnet (Stortingsmelding 26, 2012-2013; Stortingsmelding 33, 2016-2017; Transportetatene, 2016). Innenfor alle transportformer satses det på utvikling og anvendelse av effektive styringssystemer, større grad av automatisering, samt sømløse og brukerrettede informasjons-, billetterings- og betalingsløsninger. Stortinget vedtok i 2015 en ny lov (ITS-loven) med formål å sikre at ITS skal fungere koordinert og sammenhengende innenfor vegtransport og med grensesnitt mot andre transportformer (Stortinget, 2015).

Transportetatene ønsker å fremskynde innføringen av ny teknologi, gjennom tilrettelegging for utprøving av ITS i pilotprosjekter. Nasjonal transportplan for perioden 2018-2029 fremmet en egen innovasjonssatsing (Pilot-T) på 900 mill. kroner til pilotering av ny teknologi og utprøving av nye forretningsmodeller i den kommende tiårsperioden (Stortingsmelding 33, 2016-2017). Formålet er å få teknologien raskere over i anvendelse og bidra til at norsk næringsliv tar del i verdiskapingen knyttet til omstillingen i transportsektoren.

Transportpolitiske føringer og forventninger til ITS som virkemiddel i transportsystemet har lagt til rette for vesentlig forskning på og utvikling av nye og innovative transportløsninger. Før dette presenteres nærmere i kapittel 2.2.3, er det behov for en forklaring av begrepsbruk og teknologier som benyttes innenfor fagfeltet.

2.2.2 TEKNOLOGI OG VIKTIGE BEGREPER

Betegnelsen *ITS* benyttes om løsninger og systemer som anvender informasjons- og kommunikasjonsteknologi (IKT) i transportsystemet. Det eksisterer flere definisjoner av begrepet,

hvor hovedbudskapet stort sett er likt, men anvendelsen og formålet med ITS vektlegges i ulik grad. I begrepets videste forstand inkluderes alle transportrelaterte løsninger som anvender IKT, mens en mer restriktiv tolkning innebærer en forutsetning om at formålet med anvendelsen er å bidra til oppfyllelse av transportpolitiske målsettinger. Et utvalg sentrale definisjoner er gjengitt i Tabell 1.

Tabell 1: Et utvalg definisjoner av begrepet ITS

Kilde	Definisjon på ITS
Directive 2010/40/EU (2010):	<i>Intelligent Transport Systems (ITS) means systems in which information and communication technologies are applied in the field of road transport, including infrastructure, vehicles and users, and in traffic management and mobility management, as well as for interfaces with other modes of transport.</i>
ERTICO (ITS Europa):	<i>ITS – Intelligent Transport Systems and Services – is the integration of information and communications technology with transport infrastructure, vehicles and users. By sharing vital information, ITS allow people to get more from transport networks, in greater safety and with less impact on the environment.</i>
ITS-loven (Stortinget, 2015)	<i>(ITS betyr) systemer for anvendelse av informasjons- og kommunikasjonsteknologi for vegtransport, herunder infrastruktur, kjøretøy og brukere, og innenfor trafikkstyring og mobilitetsstyring, samt grensesnitt mot andre transportformer.</i>
Statens vegvesens ITS-strategi for perioden 2010-2019 (Statens vegvesen, 2012b):	<i>ITS omfatter løsninger som benytter informasjons- og kommunikasjonsteknologi (IKT) i et trafikk- eller transportsystem, og har til hensikt å påvirke atferd og forbedre transportløsning og trafikkavvikling slik at man oppnår bedre fremkommelighet, trafiksikkerhet, miljø, tilgjengelighet og brukertilfredshet.</i>
Den felles nordiske ITS-ordboken (NVF, 2012):	<i>Anvendelse av informasjons- og kommunikasjonssystemer og –tjenester innen transport og tilhørende infrastruktur.</i>

Det er utarbeidet et omfattende begrepsapparat innenfor ITS-faget. Standardiseringsorganene CEN, ETSI og ISO tilbyr en rekke standarder for sentrale komponenter og anvendelser av ITS. I tillegg er det utviklet et nasjonalt rammeverk for multimodale ITS-tjenester (ARKTRANS), som blant annet beskriver relevante roller og ansvarsområder knyttet til økosystemet rundt en ITS-tjeneste. I dette avsnittet gjengis kun en oversikt over de mest sentrale elementene, mens mer inngående informasjon kan finnes i Natvig, Westerheim, Moseng og Vennesland (2009), Foss (2015), Foss (2017a), Standard Norge (2015) og Statens vegvesen (2013).

Selve ITS-løsningen betegnes gjerne *ITS-applikasjon*, som er det praktiske verktøyet som muliggjør funksjonalitet i form av en *ITS-tjeneste* som ytes til en trafikant eller *ITS-bruker*. Dette er en svært sentral distinksjon i forståelsen av ITS som innebærer at ITS må sees på som en muliggjørende teknologi og ikke som et mål i seg selv. Poenget er altså ikke å ta i bruk ITS for teknologiens skyld, men å rette fokus mot den tjenesten som tilbys trafikanter og andre ITS-brukere. I et anvendt perspektiv er hensikten med ITS-tiltaket å gi en nytteverdi for brukeren eller samfunnet for øvrig.

Denne forståelsen av ITS har muligens vært mindre fremtredende i den tidlige utviklingen av ITS-applikasjoner, som i stor grad har vært initiert og drevet frem av bilindustrien og teknologileverandører.

IKT-infrastrukturen i en ITS-applikasjon kategoriseres i fire undergrupper, *ITS sub-systemer*, avhengig av hvilket objekt ITS-applikasjonen er tilknyttet (Foss, 2015):

- *Personlig ITS sub-system* er mobile enheter eller utstyr som brukeren har med seg, eksempelvis i form av smarttelefon, nettbrett eller smartklokke. Det personlige ITS-systemet er helt sentralt for en del trafikantinformasjonstjenester og betalingsløsninger for kollektivtrafikk, der interaksjonen mellom tjenesten og brukeren for eksempel er knyttet opp mot bruk av applikasjoner (apper) på smarttelefon.
- *Kjøretøy ITS sub-system* er utstyr tilknyttet det aktuelle kjøretøyet. Moderne biler er utstyrt med en rekke sensorer og teknologier som registrerer og prosesserer data og informasjon til støtte for bilføreren. Slike førerstøttesystemer kan være rent informative, f.eks. varsling om glatt vegbane, eller de kan virke sammen med kjøretøyet elektroniske kontrollenheter og automatisere deler av kjøreprosessen, slik tilfellet er med blant annet elektronisk stabilitetskontroll (ESC) og adaptiv cruise control (ACC).
- *Vegkant ITS sub-system* er utstyr som monteres langs vegen, og som samler inn og prosesserer data ved hjelp av sensorer i infrastrukturen eller passerende kjøretøy. I noen tilfeller registreres data om trafikken (f.eks. trafikktegninger), mens i andre tilfeller er det data fra veginfrastrukturen (klimastasjoner e.l.) som benyttes til trafikkstyring ved bruk av skilt, lyssignal eller andre grensesnitt (f.eks. ved vegstenging eller tilfartskontroll).
- *Sentralt ITS sub-system* er systemer som ikke har egne sensorer eller grensesnitt mot brukerne, men som kommuniserer med trafikantene gjennom de andre tre sub-systemene. Slike systemer blir ofte fremstilt som en skyløsning. Et annet eksempel på et sentralsystem er IKT-systemet på Vegtrafikksentralen.

Når to eller flere av disse ITS sub-systemene samarbeider om å yte en ITS-tjeneste til brukerne, benevnes dette som *samvirkende* eller *kooperativ ITS*, gjerne forkortet *C-ITS*. I praksis betyr dette at det utveksles informasjon mellom kjøretøy, infrastruktur, sentralsystem og brukere, og man snakker om *bil-til-bil-kommunikasjon* (på engelsk forkortet til C2C eller V2V) og *bil-til-vegkant-kommunikasjon* (C2I/C2X). Standardiseringsorganene CEN, ETSI og ISO har kommet frem til en felles definisjon på samvirkende ITS:

A cooperative ITS is a subset of the overall ITS that communicates and shares information between ITS stations to give advice or facilitate actions with the objective of improving safety, sustainability, efficiency and comfort beyond the scope of stand-alone systems.

Mens benevnelsen *cooperative ITS* eller forkortelsen C-ITS er rådende i Europa, benyttes begrepet *connected vehicles* i stor grad i amerikansk litteratur. U.S. Department of Transportation gir følgende definisjon av *connected vehicles*; *Connected vehicles interact with each other (V2V), the roadside (V2I), and beyond (V2X) via wireless communications* (FHWA, 2014).

Kommunikasjonen mellom de ulike elementene i det samvirkende transportsystemet foregår i utgangspunktet over 5,9 GHz-frekvensen som er dedikert til datautveksling mellom sikkerhetsrelaterte ITS-applikasjoner (benevnes også *ITS-G5*). ITS-G5 muliggjør lokal trådløs datautveksling direkte mellom to eller flere enheter innenfor et geografisk begrenset område. For betalingsløsninger og bomstasjoner benyttes ofte *DSRC-teknologi*⁴ som blant annet inngår i AutoPass-brikken. Dersom ITS-applikasjonen skal ha større rekkevidde, må kommunikasjonen foregå via mobilnettet (4G/5G) eller internettforbindelse (WI-FI). Det pågår i dag pilotering av 5G-teknologi og nettverk som forventes å spille en vesentlig rolle i utviklingen av fremtidens smarte byer og intelligente transportløsninger⁵. Applikasjoner som kombinerer bruk av ITS G5 og mobilnettverk omtales gjerne som *hybrid C-ITS-applikasjoner*. Flere bilprodusenter jobber med ulike skyløsninger der informasjon fra kjøretøy eller infrastruktur lastes opp i en skytjeneste hvor data videreføres før det deles med andre kjøretøy/enheter som er en del av det samme kommunikasjonsnettverket.

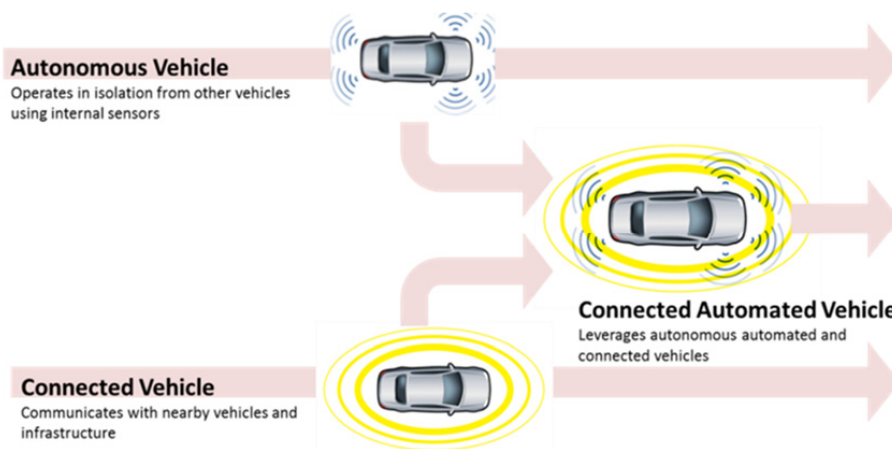
Kommunikasjon mellom kjøretøy og infrastruktur gir nye muligheter for datainnsamling og trafikkregistreringer. Der hvor man tidligere var avhengig av tellepunkter og fysiske installasjoner i vegbanen (Eulian registreringsteknikk), kan man ved å bruke bilen som sensor samle inn store mengder data om både kjøretøy, atferd og status på veg og omgivelser ved å følge individuelle kjøretøy (Lagrangian registreringsteknikk) (Herrera & Bayen, 2010; Work, Tossavainen, Jacobson, & Bayen, 2009).

Et sett med begreper som litt misvisende brukes om hverandre både i fagskrift og populærvitenskapelige fremstillinger, er *autonome*, *automatiserte*, *førerløse* eller *selvkjørende kjøretøy*. Begrepet *autonomt kjøretøy* beskriver i utgangspunktet kjøretøy eller biler som kan operere helautomatisk basert på sitt eget kontrollsystem, kunstig intelligens og sensorer som

⁴ Dedicated short range communications (DSRC)

⁵ Uttesting pågår både i Europa, Asia og USA. I Norge startet Telenor pilotering av 5G-teknologi i 2019.

registrerer data fra omgivelsene ved hjelp av blant annet integrerte radarer, lidarar (light detection and ranging), GPS-navigasjon og computersyn. Disse er altså ikke avhengig av input fra omgivelsene for å håndtere kjøreplassen, da all nødvendig data ligger lagret i kjøretøyet eller samles inn i sanntid av kjøretøyet egne sensorer. Det kan synes som om begrepet *autonom* har fått fotfeste med hensyn til å beskrive det som også egentlig er *automatiserte kjøretøy* (Wood, Chang, Healy, & Wood, 2012), hvor kjøreplassen riktignok er maskinstyrt, men hvor beslutningsprosessen er avhengig av input fra omgivelsene. Et kjøretøy som styres på autopilot med bruk av samvirkende ITS er eksempel på et automatisert kjøretøy som ikke er autonomt – fordi det baserer kjøreplassen på kommunikasjon og samspill med omgivelsene. U.S. Department of Transportation (2015) skisserer et mulig fremtidsbilde hvor utviklingen av samvirkende kjøretøy og autonome kjøretøy konvergerer, slik at de ulike teknologiene kompletterer hverandre. På sikt kan dette resultere i en robust løsning med selvkjørende kjøretøy som anvender samvirkende teknologi som supplerende støtte, som vist i Figur 4.



Figur 4: Konvergens av autonome og samvirkende kjøretøy, illustrasjon fra U.S. Department of Transportation (2015).

I media og i fremtidsbeskrivelser av vårt transportsystem benyttes ofte begrepet *førerløse biler*. Dette er i prinsippet en samlebetegnelse for alle tilfeller hvor kjøreplassen er automatisert og kan utføres uten menneskets innblanding, uavhengig av om kjøretøyet er autonomt eller ikke. Begrepet benyttes imidlertid også når konvensjonelle kjøretøy er i bevegelse uten fører til stede (f.eks. når et parkert og forlatt kjøretøy uheldigvis ruller av gårde). Norsk faglitteratur anbefaler isteden å bruke begrepet *automatisert* eller *selvkjørende* som brukes i den norske lovteksten (Foss, 2017a; Stortinget, 2017a).

I realiteten er automatisering av kjøretøyet eller kjøreoppgaven en *gradvis* prosess som allerede er påbegynt. SAE International⁶ presenterte i 2014 en stegvis modell for automatisering, som senere har fått bred anvendelse innenfor fagområdet (ERTRAC, 2015; SAE International, 2014, 2018). Modellen består av seks trinn, fra *ingen automatisering* til *full automatisering*; hvor de tre første stegene beskriver en tilstand hvor *føreren overvåker kjøreomgivelsene* og de tre siste stegene beskriver en tilstand hvor *teknologien overvåker kjøreomgivelsene*. Kjøretøyets automatiseringsprosess er illustrert i Figur 5, oversatt til norsk fra IEEE Spektrum (2020).

SAE - Automatiseringsnivå

Full Automatisering					
0	1	2	3	4	5
INGEN AUTOMATISERING	FØRERSTØTTE (ADAS)	SEMI – AUTOMATISK	BETINGET AUTOMATISERING	HØY AUTOMATISERING	FULL AUTOMATISERING
Null Autonomi Føreren utfører alle kjøreoppgaver	Kjøretøyet kontrolleres av fører, men deler av kjøreoppgaven er automatisert	Kjøretøyet kontrolleres av sammensatte ADAS funksjoner, men føreren må følge med på trafikkbildet til enhver tid	Fører er nødvendig, men føreren må ikke følge med på trafikkbildet hele tiden Føreren må likevel være forberedt på å ta over kontroll på kort varsel	Kjøretøyet kan utføre alle kjøreoppgaver under visse betingelser En person kan ta over kontroll om ønskelig	Kjøretøyet kan utføre alle oppgaver under alle forhold En person kan ta over kontroll om ønskelig

Figur 5: Kjøreprosessens automatiseringsgrad i seks nivå. Oversatt til norsk fra IEEE Spektrum (2020).

I henhold til SAEs automatiseringsprosess (SAE International, 2014, 2018), befinner de mest avanserte kommersielle kjøretøymodellene seg i hovedsak på nivå 2 med delvis automatisering. Moderne biler har førerstøttesystemer for både styring og fartskontroll basert på informasjon fra omgivelsene, men føreren må delta aktivt i kjøreprosessen, både ved å utføre øvrige oppgaver og inneha den overordnede kontrollen. Nivå 3 er omstridt med hensyn til trafiksikkerhet, da dette nivået innebærer at teknologien har overtatt kjøreprosessen, samtidig som det forventes at føreren fortsatt er årvåken og klar til å overta kontrollen dersom det kreves. Forskning på menneskets evne til å opprettholde oppmerksomheten og foreta gode valg og handlinger i slike situasjoner, tyder imidlertid på at mennesker har begrenset kapasitet på dette feltet (Kyriakidis et al., 2017).

Tesla ga sine kunder tilgang til automatiserte funksjoner (nivå 2) i 2015, og har senere annonsert at de sikter mot full selvkjørende funksjonalitet, under forutsetning av at lovgrunnlaget åpner for

⁶ Verdensomfattende forening for ingeniører og tekniske eksperter innenfor romfart og kjøretøyindustrien.

nødvendige godkjenninger. Tesla-bilene selges i dag med hardware som i prinsippet skal kunne opereres selvkjørende kun ved hjelp av oppdatert software⁷. Googlebilen (Waymo) er et eksempel på høy grad av automatisering (nivå 4), og flere store bilprodusenter har utviklet prototyper med selvkjørende funksjoner i visse situasjoner (for eksempel parkering). De selvkjørende minibussene fra EasyMile og Navya som er testet her til lands oppgis også fra produsentene å kunne operere på nivå 4 (blant annet med restriksjoner på kjørestrekning).

Et kooperativt utviklingsløp innebærer både at stadig flere operasjoner kan utføres automatisk, og at man etter hvert kan få dedikerte strekninger eller områder som tilbyr høy grad av samvirkende tjenester. Det er imidlertid viktig å ta høyde for at en fullstendig fornying av bilparken og vegnettet vil skje over tid, og mens denne utviklingen pågår skal både transportsystemet og trafikantene håndtere en tilværelse med blandet trafikk. Dette innebærer både at de enkelte kjøretøyene befinner seg i ulike faser av automatiseringsprosessen og at vegnettet og infrastrukturen har ulik teknologisk standard.

2.2.3 TJENESTER

Det er utviklet et stort antall ITS-tjenester som både har ulike formål og som er rettet mot ulike målgrupper. I henhold til ISO TC 204/SC⁸ kan tjenestene kategoriseres i tolv ulike tjenesteområder (oversatt og tilpasset norske forhold i Foss (2017b)):

Trafikantinformasjon: Dette tjenesteområdet omfatter levering av statiske og dynamiske data om vegnettet og trafikken til brukerne av ITS-tjenestene.

Trafikkstyring: Dette tjenesteområdet dekker styring av alle typer kjøretøy og alle typer trafikanter i veg- og gatenettverk. Området inkluderer både automatisk overvåking, kontroll og styring og beslutningsprosesser som adresserer sanntids hendelser og andre forstyrrelser i transportsystemene. Det adresserer også styring av etterspørsel slik at det opprettholdes en optimal avvikling og mobilitet i transportsystemene.

Kjøretøyrelaterte ITS-tjenester: Dette tjenesteområdet omfatter levering av ITS tjenester i kjøretøyet. Området fokuserer på en forbedring av sikkerhet og effektivitet i styring av kjøretøyet både gjennom advarsler og anbefalinger til brukeren eller som input til styringen/kontrollen av kjøretøyet. Området omfatter både bruk av intern (i kjøretøyet) og ekstern informasjon.

⁷ Etter flere ulykker og hendelser med automatiserte kjøretøy på SAE nivå 2 har Tesla begrenset funksjonaliteten i sin autopilot, slik at den overholder EUs reviderte forskrift om førerstøttesystemer (Regulativ 79 FN/ECE).

⁸ ISO Technical committee 204 Intelligent transport systems

Kollektivtransport: Dette tjenesteområdet dekker styring av kollektivtrafikkmidler slik at kollektivselskapene kan levere effektive og pålitelige kollektivtransporttjenester. Dette inkluderer også innsamling av informasjon om avviklingen av kollektivtrafikken og informasjon til reisende om status. Området dekker også styring av etterspørselsbasert kollektivtrafikk.

Godstransport: Dette tjenesteområdet dekker transport av gods og styring av kjøretøyflåter med næringskjøretøy. Området inkluderer også administrative tjenester som skal medvirke til sikker og effektiv transport av gods.

Betalings tjenester: Dette tjenesteområdet dekker aktiviteter som gjør det mulig å gjennomføre elektronisk betaling for transporttjenester gjennom bruk av elektroniske verdier og non-stop systemer.

Overvåking av vær og miljø: Dette tjenesteområdet dekker aktiviteter som overvåker været og miljøet og varsler om forhold som kan ha virkninger for både transportbrukere og operatører av transportinfrastruktur.

Nødtjenester: Dette tjenesteområdet dekker aktiviteter som både sender og mottar meldinger om nødsituasjoner, dirigerer utrykningskjøretøy i tilknytning til nødmeldinger og som bidrar til å finne igjen stjalne kjøretøy.

Drift av infrastruktur: Dette tjenesteområdet dekker aktiviteter knyttet til vintervedlikehold som brøyting og salting.

Katastrofehåndtering: Dette tjenesteområdet dekker aktiviteter som håndterer ressurser fra ulike fagdisipliner i forhold til naturkatastrofer, store ulykker, terrorhandlinger o.l.

Personikkerhet: Dette tjenesteområdet dekker aktiviteter som skal sørge for personikkerheten i transportsystemer. Området dekker primært sikkerheten til kollektivtrafikanter, fotgjengere og spesielt sårbare trafikanter.

Håndtering av terrortrusler: Dette tjenesteområdet er ikke inkludert i den norske veilederen.

Hver av disse tjenesteområdene inneholder en rekke ITS-tjenestegrupper, som totalt omfatter 138 konkrete ITS-tjenester.

Hensikten med denne oversikten er å vise at det er stor variasjon i ITS-tjenestene som utvikles, både når det gjelder teknologi, nødvendig infrastruktur og målgrupper for tjenesten. Når ITS løftes frem som et lovende virkemiddel for å oppnå transportpolitiske mål er begrepet i realiteten en

samlebetegnelse for et mangfold av teknologibaserte tiltak, som har ulik grad av modenhet og et svært sammensatt effektbilde.

2.2.4 FORSKNING OG UTVIKLING AV ITS

Verden over legger myndigheter til rette for utvikling og implementering av fremtidens transportteknologier, gjennom strategisk satsing på forskning og uttesting, samt tilpasning av lovgivning og rammebetingelser. Allerede i 1991 ble det opprettet interesseorganisasjoner for ITS både i USA (ITS America) og Europa (ERTICO).

I USA har langsiktige, nasjonale programmer bidratt til å posisjonere industri og offentlige transportaktører for storskala implementering av automatiserte kjøretøy (National Highway Traffic Safety Administration, 2013). Mange stater var tidlig ute med å revidere lovverket for å tillate selvkjørende kjøretøy på offentlig veg, noe som har gitt USA et forsprang i utprøving denne teknologien. I Japan ble en landsomfattende implementering av samvirkende systemer påbegynt allerede i 2011, med de første tjenester tilgjengelige for trafikk i virkelige omgivelser fra 2015 (European Commission, 2016d). Storbritannia har også lang tradisjon for samarbeid mellom myndigheter, industri og akademiske miljø for å realisere samfunnsgevinstene som ligger i å oppnå intelligent mobilitet. Landet var blant annet tidlig ute med å demonstrere selvkjørende passasjertransport i tilknytning til terminal 5 på Heathrow flyplass (Bly & Lowson, 2007), og myndighetene var blant de første i Europa til å foreta juridiske betraktninger og utarbeide testprosedyrer med sikte på testing av selvkjørende transport (Department for Transport, 2015; ERTRAC, 2015).

EU-kommisjonen har gjennom mange år gitt betydelig økonomisk støtte til forskning, utredning og innføring av ITS. De første store ITS-prosjektene, PROMETHEUS⁹ (1987-1995) og DRIVE¹⁰ (1989-1992) var knyttet opp mot kjøretøyteknologi og trafikksikkerhet (Nowacki, 2012). Fokuset har med tiden blitt utvidet fra de tekniske aspektene ved ITS til større vektlegging av anvendelser og effekter i piloter og demonstrasjonsprosjekter (European Commission, 2016d), som f.eks. CVIS (Kompfner, 2010), SAFESPOT (Andreone et al., 2010), FOTsis (Pou et al., 2015) og CityMobil (Stam & Alessandrini, 2012). I 2008 ble det utviklet en europeisk handlingsplan for innføring av ITS (ITS Action Plan) etterfulgt av et direktiv (Directive 2010/40/EU, 2010) som forplikter medlemslandene til oppfølging av politikken som del av EØS-avtalen. Handlingsplanen trekker opp

⁹ Program for a European Traffic System with Higher Efficiency and Unprecedented Safety (PROMETHEUS). Initiativ fra den europeiske bilindustrien for å øke konkurransekraften mot japansk bilproduksjon.

¹⁰ Dedicated Road Infrastructure for Vehicles Safety in Europe (DRIVE). EU-initiert forskningsprogram med fokus på samfunnets og trafikantens utfordringer og interesser knyttet til ITS.

prioriteringer og satsingsområder, og målet med direktivet er å akselerere og koordinere utnyttelsen av ITS på vegtrafikkområdet og i grenseflaten mot andre transportformer. Ifølge en oversikt utarbeidet av EU-kommisjonen i 2011 (referert i Stortingsmelding 26 (2012-2013)) er Norge blant de landene i Europa som har vært mest aktive når det gjelder å innføre ITS-tjenestene som er spesielt prioritert i ITS-direktivet.

De siste årene har fokuset i stadig større grad rettet seg mot automatiserte og samvirkende systemer (ERTRAC, 2015; FHWA, 2014; Transportetatene, 2016). I 2013 ble det inngått en offisiell samarbeidsavtale om å etablere en sammenhengende korridor for uttesting og anvendelse av samvirkende ITS på det europeiske vegnettet mellom Rotterdam, Frankfurt og Wien (Amsterdam Group, 2013). I etterkant er det etablert en tilsvarende teststrekning i Frankrike (Direction generale des Infrastructures des Transports et de la Mer, 2014), og også i Norden samarbeides det om utvikling av den kooperative korridoren *NordicWay* mellom Helsinki, Stockholm, Göteborg, Oslo og København (Statens vegvesen, 2015). I tillegg samarbeider Norge og Finland om uttesting av samvirkende teknologi langs E8 mellom Skibotn, Kilpisjärvi og Kolari gjennom Aurora Borealis-prosjektet (Statens vegvesen, 2017). I mars 2016 ble Amsterdamerklæringen *Cooperation in the field of connected and automated driving* signert av alle EUs medlemsstater og Norge. Erklæringen fastsetter felles mål for å jobbe mot et europeisk rammeverk for automatiserte kjøretøy og tilrettelegge for utprøving av automatiserte kjøretøy i eget land og på tvers av landegrenser (Stortinget, 2017b).

I Europa har forskning på og utvikling av ITS i stor grad blitt drevet frem av land med en aktiv bilindustri, som Frankrike, Tyskland, Storbritannia, Spania og Sverige. I tillegg har særlig Nederland vært sentral i innovasjon og tilrettelegging for ITS i vegsystemet (ERTRAC, 2015). Nederland har vært et foregangsland med store ambisjoner og lang tids erfaring med testing av avanserte førerstøttesystemer som ISA (intelligent fartstilpasning), LDW (elektronisk feltskiftevarsling) og ACC (adaptiv cruise control). Nederlandske aktører har deltatt i store europeiske forskningsprosjekter på samvirkende systemer (f.eks. CVIS og SAFESPOT), samt inngått strategiske allianser for storskala anvendelse av samvirkende ITS i Europa (Amsterdam Group, 2013). Samarbeid mellom offentlige og private aktører har tilrettelagt for introduksjon av selvkjørende kjøretøy¹¹, blant annet gjennom tilpassing av lovverket som tidlig tillot feltforsøk med automatiserte kjøretøy på offentlig veg i Nederland. Den offensive nederlandske tilnærmingen til digitalisering av transportsystemet blir gjerne beskrevet med varianter av uttrykket *learning by*

¹¹ Dutch Automated Vehicle Initiativ (DAVI) er et offentlig-privat samarbeid mellom myndigheter, forskningsmiljø og interesseorganisasjoner for bil som utvikler automatiske kjøretøy for forskning og testing på offentlige veier i Nederland, med fokus på trafiksikkerhet og såkalte human factors.

doing and making it happen (ERTRAC, 2015; Government of the Netherlands, 2016), som peker på at fremskritt og utvikling bygger på aktiv uttesting og læring gjennom prøving og feiling. Også Østerrike har vært en pådriver for ITS-basert mobilitet gjennom nasjonale støtteprogram og handlingsplaner (ERTRAC, 2015; Paulweber, Rom, & Affenzeller, 2016). Fokuset har spesielt vært rettet mot teknologiløsninger og sensorer i infrastrukturen, med samvirkende ITS og kommunikasjon mellom kjøretøy og vegkant (COOPERS¹², Amsterdam Group (2013); ERTRAC (2015)).

I Frankrike, Tyskland og delvis også Spania foreligger en rekke nasjonale initiativ for å støtte opp om egen bilindustri og utviklingen av fremtidens kjøretøy. Disse landene har lang tradisjon for satsing på ITS og førerstøttesystemer, og har med åpenbare økonomiske interesser fra bilindustrien vært sentrale i en rekke store europeiske teknologiprojekter (COOPERS, CVIS, SAFESPOT). Dette videreføres mot utvikling og demonstrasjon av samvirkende systemer og selvkjørende kjøretøy, med støtte fra både myndigheter og forskningsinstitusjoner (ERTRAC, 2015). Frankrike har vært førende i utviklingen av selvkjørende, elektriske minibusser, med kommersielt tilgjengelige produkter fra EasyMile og Navya fra 2016. Siden den gang er det etablert en rekke demonstratorer og piloter med slike busser rundt omkring i verden.

I Sverige er ITS-arbeidet i stor grad motivert av myndighetenes nullvisjon og bilprodusenten Volvo Cars' fokus på trafikksikkerhet. I det som omtales som verdens første store pilotprosjekt for automatisert kjøring, *Drive Me: Selfdriving cars for sustainable mobility*, samarbeider Volvo Cars, Trafikverket, Transportstyrelsen, Lindholmen Science Park og Gøteborg by om en demonstrator med selvkjørende biler i hverdagstrafikk på vegnettet rundt Gøteborg (Volvo Cars, 2016). Ved oppstart i 2014 hadde prosjektet ambisjoner om å sette 100 selvkjørende biler på vegen i løpet av 2017. Hensikten var å se på hvilke fordeler samfunnet oppnår med automatisert transport, samt å sette Volvo og Sverige i ledelsen for utvikling av fremtidens transportsystem. Ved utgangen av 2017 var imidlertid prosjektet vesentlig nedskalert og utviklingen av fullt ut selvkjørende funksjonalitet var utsatt til 2021 (Volvo Cars, 2017)¹³.

Finland har stått i front i utviklingen av mobility-as-a-service-tankegangen (MaaS), som sikter mot optimalisering av det totale transportarbeidet, gjennom plattformer som tilbyr mobilitet som tjeneste for befolkningen. Hensikten er å samle alle transporttilbydere for ulike transportformer i en felles

¹² COOPERS (Cooperative systems for intelligent road safety) er et EU-finansiert forskningsprosjekt innenfor 6. rammeprogram i perioden 2006-2010.

¹³ Per desember 2017 var det kun inngått testavtale med to familier, og kjøretøyene som ble benyttet hadde kun delvis selvkjørende funksjonalitet.

tjeneste som omfatter både reiseinformasjon, bestillingstjenester og betalingsmodeller. Løsningen representerer en innovativ forretningsmodell basert på delingsmobilitet istedenfor eierskap til transportmidler.

I Norge har vegmyndighetene inntatt en sentral rolle i satsing på samvirkende ITS som forventes å kunne gi betydelige sikkerhets- og effektivitetsgevinster (Statens vegvesen, 2018). Eksempelvis gjennomføres det avansert teknologisk utprøving av nye ITS-applikasjoner på det nordiske vegnett i regi av NordicWay-prosjektene (Statens vegvesen, 2015) og Borealis-prosjektet på E8 (Statens vegvesen, 2017). Offentlige og private aktører har i fellesskap stimulert til forskning og utvikling av automatiserte transporttjenester. Allerede i 2009 ble selvkjørende minibuss demonstrert i Trondheim i regi av EU-prosjektet CityMobil (Stam, Alessandrini, & Marco, 2010). I 2016 gikk slike busser fra å være forskningsbaserte prototyper til å bli kommersielle produkter klare for utprøving på virkelig veg. Dette avstedkom en rekke demonstratorer omkring i landet, hvor kjøretøy og teknologi ble vist frem for publikum og beslutningstakere i transportsektoren på initiativ fra private teknologiaktører¹⁴. I januar 2018 ble det innført et nytt lovgrunnlag for utprøving av selvkjørende kjøretøy på offentlig veg i Norge (Stortinget, 2017a). Dette åpnet for pilotering av selvkjørende kjøretøy på definerte vegstrekninger, etter risikovurdering og nødvendig godkjenning fra Vegdirektoratet. I perioden 2018-2020 ble det realisert fem selvkjøringspiloter i blandet trafikk på offentlig veg her i landet; Forus, Fornebu, Gjøvik, Kongsberg og Ruterpiloten i Oslo, og i løpet av 2019 ble selvkjørende minibusser satt inn som en del av det ordinære kollektivtilbudet både på Kongsberg (Brakar) og i Oslo (Ruter). Forskningsprosjektet SmartFeeder har fulgt aktivitetene rundt den første utprøvingen av selvkjørende kjøretøy i Norge (Lervåg, Foss, & Wahl, 2018). Hensikten med prosjektet er å bruke de tidlige piloterfaringene til å bygge kunnskap om hvordan selvkjørende kjøretøy bør innføres i fremtidens transportsystem på en måte som styrker de kollektive transportløsningene.

Til sammen har alle aktivitetene på fagfeltet resultert i at Norge nå er rangert som det tredje beste landet i verden med hensyn til å være forberedt på innføringen av selvkjørende kjøretøy i transportsystemet, etter Nederland og Singapore (KPMG, 2019). Norge scorer spesielt høyt på faktoren *teknologi og innovasjon*. Norge er sammen med Nederland og resten av Skandinavia blant de mest digitaliserte landene i verden, målt i henhold til kriterier som dekningsgrad, digital kompetanse og offentlige tjenester (European Commission, 2016b, 2016e). Dette gir oss særskilte

¹⁴ I Norge ble det i løpet av de neste par årene gjennomført 23 demonstratorer med over 20.000 passasjerer. Tall samlet inn av forfatteren fra aktørene Acando, Forus Prt og Applied Autonomy per november 2018.

fortrinn med hensyn til muligheten for å ta i bruk og realisere nyttegevinster av nye teknologiske tjenester.

2.3 TEKNOLOGIEN INTRODUSERER NYE PROBLEMSTILLINGER OG ENDRET KUNNSKAPSBEHOV

De transportpolitiske strategier og forskningsaktiviteter som er presentert i dette kapitlet, viser at det er en sterk politisk vilje og høye forventninger til ITS som virkemiddel innenfor transportsektoren. Skal man oppnå økonomisk vekst og velferd for befolkningen også i fremtiden, er man helt avhengig av å finne transportløsninger som sikrer smart og bærekraftig mobilitet. Dette forutsetter fleksibilitet og vilje til å tenke nytt med hensyn til økonomiske, juridiske og organisatoriske strukturer innenfor transportområdet. ITS har potensiale til å løse mange av dagens utfordringer knyttet til sikkerhet, miljø og vegkapasitet - men introduserer samtidig noen nye problemstillinger blant annet knyttet til datasikkerhet, personvern, juridiske ansvarsforhold, kompatibilitet over landegrenser, forretningsmodeller og organisatoriske forhold (ERTRAC, 2013, 2015; European Commission, 2011, 2016a; Stortingsmelding 26, 2012-2013; Transportetatene, 2016).

Den omfattende teknologiutviklingen og digitaliseringen innenfor stadig flere samfunnsområder, har skapt debatt om i hvilken grad teknologien skal få spillerom til å forme vårt samfunn og levesett. Forsøk på å påvirke eller kontrollere teknologiutviklingen drøftes blant annet med referanse til et dilemma fremsatt av Collingridge (Genus & Stirling, 2018) i boken *The Social Control of Technology*: Det er vanskelig å forutsi konsekvensene av en teknologi før den er moden, langt utviklet og mye brukt (kunnskapsproblem) – men når teknologien først er godt etablert er det vanskelig å kontrollere eller endre teknologien (styrkeproblem). Downes (2009) drøfter også denne problemstillingen når han viser til at sosiale, økonomiske og juridiske systemer endres trinnvis, mens teknologien endres eksponentielt. Dette medfører at myndighetenes respons og regulering (lovgivning) nesten alltid vil ligge på etterskudd av teknologisk innovasjon. En følge av Collingridge-dilemmaet er "føre-var-prinsippet" som innebærer at nye innovasjoner ikke skal omfavnes eller iverksettes før tjenestetilbyderen kan bevise med vitenskapelige metoder at den ikke vil være til skade for helse, miljø eller samfunn. Kelly (2009) kritiserer imidlertid denne tilnærmingen til teknologi, og påpeker at det hever terskelen dramatisk for implementering av nyskapende løsninger.

Den store variasjonen i ITS-tjenester, både med hensyn til teknologi, infrastruktur, målgrupper og tilsiktende virkninger, gjør at det er vanskelig å etablere et felles beslutningsgrunnlag for ITS-tiltak. Verdien i bruk av ITS er ikke begrenset til trafikale virkninger av tiltak og tjenester i

transportsystemet, men er i like stor grad knyttet til innsamling av data og hvordan disse kan anvendes. En moderne bil er fylt med datateknologi og sensorer som gjør det mulig å registrere data om de fleste egenskaper ved kjøretøyet og kjøreatferden. I tillegg registreres informasjon om trafikken fra en stadig økende bruk av teknologi i infrastrukturen. Ved å hente ut og kombinere informasjonen fra de ulike datakildene, får man tilgang til kunnskap om trafikken og trafikksystemet som endrer premissene for transportplanlegging og kan effektivisere og forbedre trafikkstyring på et overordnet nivå. ITS består av komplekse løsninger, involverer flere aktører og gir andre virkninger enn tradisjonelle veg- og trafikktiltak. Dette reiser en rekke problemstillinger knyttet til vurdering av tiltakenes nytteverdi og kompliserer beslutningsprosessen for investeringer og prioritering mellom tiltak i transportsektoren.

Med høy innovasjonstakt, blir samspillet mellom industrielle kommersielle aktører, forskning og offentlige myndigheter viktig for å øke utviklingstempoet og realisere potensialet innenfor transportsektoren. Böhm, Flechl, Aigner og Visser (2016) peker på at det i dag er et betydelig gap i kommunikasjonen mellom beslutningstakere med begrenset teknologiforståelse på den ene siden - og teknologifokuserte ITS-aktører med manglende evne til å sette seg inn i kundens reelle behov på den andre siden. Når man i tillegg mangler historiske erfaringer om tiltakenes effekter og lønnsomhet, blir dette en vesentlig barriere for investeringsbeslutninger knyttet til konkrete ITS-tiltak. En endringsorientert transportpolitikk forutsetter at man fortløpende foretar velinformerte beslutninger som håndterer en uforutsigbar fremtid. Spørsmålet er da om vi har passende verktøy for å tilegne oss et nødvendig kunnskapsgrunnlag. Kunnskap bygges blant annet gjennom evalueringer. I hvilken grad er tradisjonell evalueringsmetodikk tilstrekkelig i en tid hvor endringer skjer raskt og på mange fronter samtidig? Dette temaet drøftes nærmere i de to neste kapitlene.

3 EVALUERINGSTEORI

Dette kapittelet gir et innblikk i temaet evaluering som fagdisiplin, med vekt på rådende teorier, perspektiver og metoder innenfor fagfeltet. Ambisjonen er ikke å gi en utfyllende gjennomgang av alle retninger og diskurser innenfor feltet. Hensikten er derimot å introdusere sentrale perspektiver som gir et faglig grunnlag for den videre diskusjonen av evalueringer innenfor transportsektoren.

3.1 EVALUERING SOM BEGREP

Begrepet *evaluering* inngår i dagligtalen som et synonym for ordet *vurdering*, *kvalitativ bedømmelse* eller *verdifastsetting*, og meningsinnholdet kan variere i ulike sammenhenger. Selve ordet beskriver et bredt spekter forskjelligartede aktiviteter, fra menigmanns intuitive vurdering av tilgjengelige alternativer i en valgsituasjon, til en mer systematisk og forskningsbasert gjennomgang av ulike aspekter og konsekvenser av et tiltak eller prosjekt. Det finnes mangfoldige definisjoner av begrepet evaluering, blant annet avhengig av hvilket fagområde som er i fokus. I det følgende gis det noen eksempler fra internasjonal evalueringslitteratur, fra offentlig anvendelse og fra norsk lærebokmateriell i evaluering.

En av de ledende forfatterne innenfor evalueringslitteratur (Scriven, 1991) definerer evaluering som: «...*judging the worth or merit of something or the product of the process*». Han skiller distinkt mellom *samfunnsfaglig forskning* som kun forholder seg til observerte, målte eller kalkulerede data og som i sin natur er objektiv til de resultatene som avdekkes – i forhold til *evaluering* som tar resultatene fra forskningen videre til en sammenligning og vurdering opp mot et sett forhåndsdefinerte kriterier eller standarder. I et slikt perspektiv har evaluering en tilleggsdimensjon til det som tradisjonelt anses som samfunnsfaglig forskning.

OECDs definisjon fra evaluering av bistands- og utviklingsprogrammer er en mye brukt referanse på forklaring av evalueringsbegrepet (1991, s. 5):

An evaluation is an assessment, as systematic and objective as possible, of an on-going or completed project, programme or policy, its design, implementation and results. The aim is to determine the relevance and fulfilment of objectives, developmental efficiency, effectiveness, impact and sustainability. An evaluation should provide information that is credible and useful, enabling the incorporation of lessons learned into the decision-making process of both recipients and donors.

Sentrale poenger i denne definisjonen fremholder at evalueringen er *systematisk* og *objektiv*, at den kan foregå *underveis eller etter* en implementeringsprosess, og at hensikten med evalueringen er å

bestemme *relevans, måloppnåelse, virkningsgrad, effektivitet, påvirkning/effekt* og *bærekraftighet*. Målet med evalueringen er å frembringe *troverdig og nyttig informasjon* som basis for *fremtidige beslutningsprosesser*. Denne tolkningen av begrepet har et målrettet og anvendt perspektiv, i det formålet med evalueringen er å vurdere i hvilken grad et tiltak eller prosjekt har oppfylt (forhåndsdefinerte) krav, og hvor hensikten er at resultatene skal støtte eller legge grunnlag for fremtidige beslutninger om tiltaket som vurderes.

I Finansdepartementets evalueringsveileder (2005, s. 8) defineres evaluering som:

En systematisk datainnsamling, analyse og vurdering av en planlagt, pågående eller avsluttet aktivitet, en virksomhet, et virkemiddel eller en sektor. Evalueringer kan gjennomføres før et tiltak iverksettes (ex ante), underveis i gjennomføringen, eller etter at tiltaket er avsluttet (ex post). Evalueringen kan utføres av interne eller eksterne fagmiljøer.

Denne definisjonen favner bredt og tar høyde for at evalueringen kan foregå på ulike tidspunkt i implementeringsprosessen, også som del av en planleggingsprosess *før* tiltaket innføres. Denne tolkningen tar ikke stilling til valg av metode eller formålet med evalueringen. Det stilles heller ingen eksplisitte krav til gjennomføring eller objektivitet, utover at evalueringen skal være *systematisk*.

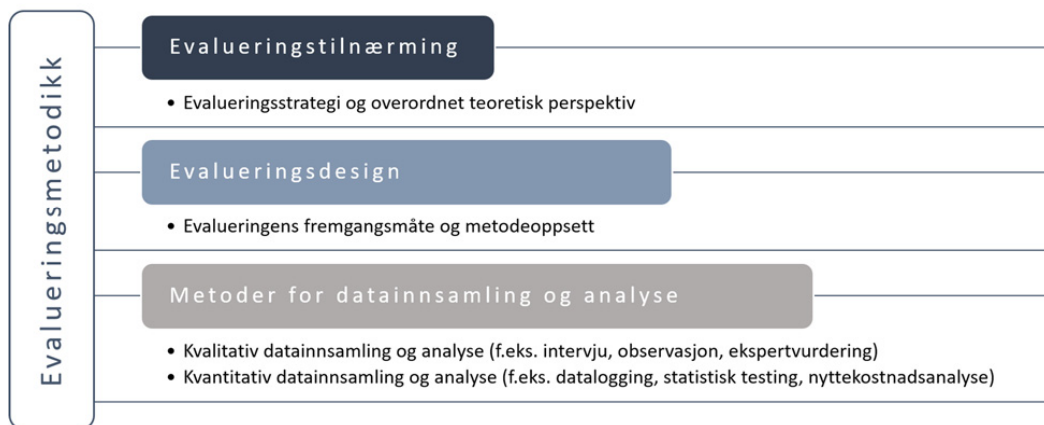
Evaluering som forskningsbasert virksomhet kjennetegnes ifølge Sverdrup (2002) av en *metodisk og teoretisk forankring*, en *empirisk orientering* og vektlegging av *systematikk og grundighet* i analyser. I dette ligger en anvendelse av aksepterte forskningsbaserte prinsipper som skal sikre gyldighet og pålitelighet i resultatene som fremkommer. Det endelige siktemålet for forskningsbasert evaluering er *akkumulering av kunnskap*. Denne beskrivelsen vektlegger bruk av objektive metoder og teorier som grunnlag for forskningen, og nøyaktighet og etterprøvbarhet i gjennomføringen. Dette skal sikre sanne og allmenngyldige resultater som bidrar med ny kunnskap innenfor det aktuelle fagfeltet. I denne tolkningen legges det ingen føringer med hensyn til evalueringsperspektiv (valg av teoretisk forankring), og det er heller ingen fokus på hvordan evalueringsresultatene eller kunnskapen som oppnås skal anvendes. Halvorsen (2013) skiller mellom forskningsbasert evaluering versus andre evalueringer med utgangspunkt i fire kriterier; *generaliserbarhet, formidling, vitenskapelig innstilling* og *vitenskapelig metode*. Betingelsen som må oppfylles for at evalueringen skal kategoriseres som *forskningsbasert* innebærer en intensjon om generaliserbarhet og bruk av teori, formidling til forskningsfellesskapet, søking av tilleggsinformasjon utover oppdragsgivers spørsmål, samt vurdering av metodevalg med vekt på åpenhet omkring svakheter. Forfatteren peker på at det finnes en rekke evalueringer som er

forskningsbaserte, og mange som ikke er det - uten at skillet mellom dem nødvendigvis fremstår som klart.

Det foregår en faglig diskusjon omkring grensegangen mellom *evaluering* og *forskning*, og i denne problemstillingen inngår også bruk av begrepene *forskningsbasert evaluering* og *evalueringsforskning*, som altså beskriver evalueringer som benytter seg av vitenskapelige metoder og prinsipper (Halvorsen, 2013; Sverdrup, 2002). Det er ulike oppfatninger om forholdet mellom evaluering og forskning, og diskusjonen begrunnes blant annet i den teoretiske forankringen av evaluering og synet på objektivitet som en grunnverdi innenfor forskningsverdenen. På den ene siden kan det hevdes at visse typer evaluering er en ren teknisk øvelse uten forankring i teorier, hvor utredninger foretas med en metodisk fremgangsmåte (diskutert i Sverdrup (2002) og Halvorsen (2013)). Dette perspektivet sidestiller evaluering med utredninger eller revisjoner, og i denne tolkningen kan det ligge en oppfatning av at evaluering ikke anses som fullverdig forskning. Det finnes også argumenter for at evaluering ikke kan skilles fra øvrig samfunnsfaglig forskning. Halvorsen (2013) viser til Campbells (uspesifisert årstall) tolkning av evaluering som anvendelse av den vitenskapelige metode (eksperimentet), og Guba og Lincolns (1989) konstruktivistiske tilnærming og metodemangfold som eksempler på at evaluering er en form for anvendt samfunnsfaglig forskning. Scriven (1991) og Shadish (1998) går enda lengre i sine vurderinger og argumenterer for at evalueringer er vesentlig mer krevende enn samfunnsforskning, fordi det forutsetter at forskeren/evaluatorene er i stand til å vurdere resultatene i en større kontekst. Der samfunnsforskeren baserer sine resultater på ren hypotesetesting, må evalueringsforskeren vurdere resultatene i relasjon til relevante tekniske, juridiske og vitenskapelige verdier (Scriven, 1991). Weiss (1972), referert i Balla, Lodge og Page (2015), går så langt som å hevde at: «*Evaluation research calls for a higher level of skills than research that is under the researcher's complete control*». I dette ligger det også en dimensjon av at evaluering foregår i en praktisk verden hvor evalueringsforskeren må inneha både kunnskap om tiltaket som evalueres, samt en helhetlig evalueringsfaglig kompetanse som gir fleksibilitet i metodevalg (også diskutert i Baklien (1993)). Et av fenomenene som gjør diskusjonen noe uklar og som muligens gjør det vanskelig å oppnå konsensus, er at begrepet evaluering rommer så vidt forskjellige aktiviteter og at hver aktør vil legge sin egen forståelse av begrepet til grunn for diskusjonen.

Dette delkapittelet har vist at begrepet *evaluering* ikke har en entydig definisjon og meningsinnholdet varierer i ulike sammenhenger. Felles for definisjonene som presenteres er at begrepet benyttes på et overordnet nivå, og at innholdet først konkretiseres når evalueringsformål og valg av evalueringsperspektiv og metode blir beskrevet. Figur 6 illustrerer hvordan en spesifikk

evalueringsmetodikk er bygd opp med utgangspunkt i en overordnet tilnærming (evalueringsstrategi og teoretisk perspektiv), med et bestemt forskningsdesign (fremgangsmåte) og konkrete metoder for datainnsamling og analyse. *Metodebegrepet* er i utgangspunktet knyttet til beskrivelsen av evalueringsdesign, datainnsamling og analyser, selv om det i dagligtalen også forekommer i benevnelse av den overordnede teoretiske tilnærmingen.



Figur 6: Evalueringens struktur

I neste delkapittel presenteres hovedlinjene i den historiske utviklingen til evalueringsforskning som fagfelt. Videre gjøres det rede for noen sentrale perspektiver og overordnede tilnærminger innenfor fagfeltet i kapittel 3.3, etterfulgt av en strukturell gjennomgang av evalueringsmetoder og praksis i kapittel 3.4.

3.2 EVALUERING SOM FAGFELT

Evalueringsforskning anses å være et relativt ungt fagfelt, med utspring i anvendt samfunnsvitenskapelig forskning i USA fra begynnelsen av det forrige århundret.

Evalueringslitteraturen peker riktignok på enkeltksemples med studier eller praksis av evalueringsmessig karakter fra tidligere år, skjønt disse eksemplene varierer avhengig av fagfelt og ståsted hos den enkelte forfatter. Det synes imidlertid å være enighet om at evaluering som systematisk og mer eller mindre faglig forankret praksis, oppsto i utdannings- og helsesektoren i USA tidlig på 1900-tallet (Pawson & Tilley, 1997; Rossi et al., 2007; Torvatn, 1993). Fra 1920-tallet finnes det etter hvert evalueringsstudier av arbeidsliv og produktivitet, med Hawthorn-studiene¹⁵ som det mest kjente eksempelet.

¹⁵ Hawthorne-studiene undersøkte hvordan ulike faktorer påvirket fabrikkarbeidernes produktivitet, men oppdaget at arbeidsinnsatsen økte også når faktorene ble holdt uendret, fordi arbeiderne jobbet hardere når de ble observert. Dette fenomenet er senere omtalt som "Hawthorne-effekten"; som beskrivelse av at forskningssituasjonen i seg selv påvirker resultatene av studien.

I Rossi et al. (2007) diskuterer forfatterne hvordan evaluering som fagfelt har utviklet seg som resultat av de politiske endringene gjennom århundret. Utover 1930-tallet iverksatte de amerikanske myndighetene store sosialpolitiske programmer som resultat av økonomiske nedgangstider, stor arbeidsledighet og sosiale problemer i det som i ettertid er omtalt som den store depresjonen i verdenshistorien. Dette bidro til forskning på effekter av sosiale programmer innenfor blant annet helse, ungdomskriminalitet og fengselsvesen. I løpet av andre verdenskrig skiftet fokus mot militær virksomhet, med studier av soldatenes moral og effekter av propagandavirksomhet.

Tiden etter andre verdenskrig var preget av vekst og velstand, og evalueringsforskningen opplevde en markant oppblomstring. Store offentlige satsinger på byutvikling, boligpolitikk, utdanning, jobbtrening og forebyggende helsearbeid ble fulgt av evalueringsforskning med fokus på resultater og effekter av tiltakene. På slutten av 1950-tallet var evaluering blitt vanlig praksis både i USA og Europa, og evalueringskomponenter ble i økende grad inkludert i planleggingen av offentlige sosiale programmer. Med stadig økende omfang og erfaring med evalueringsstudier, økte også bevisstheten omkring evalueringsmetodikk, datainnsamling og statistiske analyser.

Utover 1960-tallet ble de første artiklene og lærebøkene om evalueringsforskning og metode utgitt, og på 1970-tallet ble evalueringsforskning etablert som en egen profesjon innenfor samfunnsfaget, med opprettelse av egne universitetskurs i evalueringsmetodikk i USA. I 1976 ble også det første vitenskapelige tidsskriftet for evaluering utgitt; *Evaluation Review* fra Sage forlag. Med økt kunnskap og bevissthet omkring evalueringsforskning som fagfelt, fulgte også metodediskusjoner og kritikk av kvaliteten på tidligere gjennomførte evalueringer av sosiale programmer (Weiss, 1972). Store offentlige utgifter til velferd og sosiale programmer utløste krav om og behov for innsikt i og kunnskap om effekter og effektivitet, og det ble etter hvert større fokus på kostnadene i de offentlige prosjektene. Med Reagen-administrasjonen i USA og Margareth Tatchers privatisering og reformering av offentlig virksomhet i England på 1980-tallet (New Public Management) ble de statlige bevilgningene til sosiale programmer redusert, og evalueringer ble i større grad rettet mot vurdering av lønnsomhet og effektivitet i programmer og organisasjoner.

I Norge oppsto evaluering som fagfelt relativt sent, og selv om det kan være vanskelig å skille evaluering fra generell anvendt samfunnsforskning, regner man at evalueringsforskningen tok til først på 1980- og 1990-tallet (Askim, Døving, & Johnsen, 2013; Baklien, 1993; Dahler-Larsen, 2013; Sverdrup, 2002; Tornes, 2013; Torvatn, 1993). I følge Dahler-Larsen (2013) er dette i store trekk sammenfallende med evalueringsforskningens utvikling i Danmark, mens Sverige kan vise til en lengre evalueringstradisjon, blant annet begrunnet i en omfattende utveksling med amerikansk samfunnsvitenskap. Evaluering som fagfelt har imidlertid hatt en sterk utvikling i Norge, med

tiltagende etterspørsel etter evalueringer innenfor et bredt spekter av fag- og anvendelsesområder. Det finnes flere eksempler på store offentlige evalueringsprosjekter, og ifølge statistikk fra Direktoratet for økonomistyring¹⁶ er det i perioden 2000-2016 registrert 2019 bestilte og eksternt gjennomførte evalueringer fra departementer og statlige virksomheter. Det norske evalueringsfeltet har i stor grad adoptert og anerkjent metodikk og tilnærming fra det amerikanske og europeiske forskningsfeltet (Torvatn, 1993), og til tross for en sterk innflytelse fra New Public Management-bølgen på 1980-tallet, anvender evalueringsprosjektene i dag et bredt utvalg evalueringsmodeller. Den danske evalueringsforskeren Dahler-Larsen (2013, s. 52) påpeker at Norge i særskilt grad har tilrettelagt for at en stor grad av evaluering- og utredningsarbeidet utføres i forskningsregi, og at evalueringsarbeidet derfor har oppnådd «... en soliditet, respekt og et teori- og metodegrunnlag som mange lande, der i høyere grad hylder «evaluering» som et særskilt fænomen, må misunde Norge».

Utover 2000-tallet har evalueringsfeltet generelt vært preget av større grad av brukermedvirkning, involvering av interessegrupper og samarbeid mellom ulike aktører (Fitzpatrick, Sanders, & Worthen, 2012). I dette ligger også en dreining fra målfokus til prosessfokus i evalueringene. Rossi, et al. (2007) peker på at evalueringsforskningen i tidlige år ble formet av samfunnsforskerne (evaluatorer), mens det i senere år har vært en større påvirkning fra de som anvender forskningsresultatene (beslutningstakere, planleggere av sosiale program og samfunnet forøvrig). Fagfeltet kan således synes å ha gjennomgått en transformasjon fra et hovedsakelig akademisk utgangspunkt, til et mer politisk og anvendt perspektiv. Samtidig har evalueringsmodellene utviklet seg fra hypotetisk-deduktive mål- og effektorienterte studier av dokumenterbare virkninger, til induktive prosessorienterte studier som i større grad har som formål å påvirke og skape endring underveis i prosessen (Saunders, Evans, & Joshi, 2005; Sverdrup, 2002).

Den teoretiske utviklingen av fagfeltet diskuteres nærmere i neste delkapittel.

3.3 EVALUERINGSPERSPEKTIVER

Evalueringstilnærmingen representerer ulike modeller for å organisere evalueringsarbeidet, blant annet avhengig av hvem som forestår evalueringene, type virkemiddel eller tiltak som evalueres, spørsmål og evalueringskriterier som vektlegges, metoder som benyttes, samt hva som er formålet med evalueringen. Dette omtales gjerne som *evalueringsperspektiver*, *evalueringsparadigmer*, *evalueringsteorier* eller *evalueringmodeller*. Disse begrepene beskriver en overordnet metodisk og/eller teoretisk strategi (approach) som kan variere avhengig av det vitenskapsteoretiske

¹⁶ Statistikk fra www.evalueringsportalen.no. Det tas forbehold om at evalueringsportalen ikke har 100 % dekningsgrad, så det reelle antallet gjennomførte evalueringer er trolig høyere.

utgangspunktet. Dette må ikke forveksles med den konkrete evalueringsmetodikken som er en operasjonell beskrivelse av hvordan den spesifikke evalueringen skal gjennomføres (som vist i Figur 6). Selv om ulike evalueringstradisjoner er typisk forbundet med enkelte metoder, vil de fleste evalueringsperspektivene kunne involvere og kombinere et bredt spekter av metoder.

Michael Scriven, Professor i filosofi og skoleforskning, innførte mot slutten av 1960-tallet begrepene *formativ evaluering* og *summativ evaluering* som skiller mellom henholdsvis evalueringer som foregår samtidig med det aktuelle prosjektet og som har til hensikt å forbedre prosjektet eller tiltaket underveis i prosessen (formativ), og evalueringer som gjennomføres i etterkant av et tiltak med hensikt å avdekke effekter og måloppnåelse (summativ) (Scriven, 1991). Mens en formativ evaluering kan gjennomføres for et relativt umodent tiltak som fortsatt er under utvikling, vil en summativ evaluering forutsette at tiltaket er veletablert og at eventuelle atferdsendringer og virkninger av tiltaket har tredd i kraft. Ut ifra den historiske gjennomgangen som er presentert i dette kapitlet, kunne de tidligste evalueringene kategoriseres som summative, mens man senere har fått en større dreining mot evalueringer av formativ karakter.

Opp igjennom årene er disse to perspektivene utvidet og supplert med nye teorier, som hver for seg er lansert med litt ulike begrep og innretning avhengig av den enkelte forfatters ståsted og fokus (Alkins, 2013; Funnel & Rogers, 2011; Halvorsen, 2013; Hansen, 1991; Johansson, Denvall, & Vedung, 2015; Patton, 2006, 2012; Pawson & Tilley, 1997; Rossi et al., 2007; Shadish, 1998; Sverdrup, 2002). Dette har medført at evalueringfeltet kan fremstå noe uoversiktlig, med et betydelig antall tilnæringsmåter, som i en viss grad konkurrerer med hverandre og i en viss grad utfyller hverandre. Det finnes ingen felles akseptert entydig evalueringsteori, og de fleste evalueringer av en viss størrelse vil kombinere perspektiver fra ulike teoretiske ståsted. I de følgende avsnittene er det valgt å gruppere noen sentrale evalueringsperspektiver i et utvalg hovedkategorier.

3.3.1 MÅLORIENTERT EVALUERING

Målorientert evaluering fokuserer på måling av resultater og effektivitet av et tiltak eller prosjekt i relasjon til forhåndsdefinerte målsettinger. Denne tilnærmingen kalles også *beslutningsorientert* eller *resultatorientert* evaluering, fordi evalueringen brukes som et styringsmiddel for fremtidige beslutninger. På engelsk benyttes blant annet begrepene *goal-oriented evaluation* og *objective-oriented evaluation*. Utviklingen av målorientert evalueringsteori tilskrives utdanningsforskeren Ralph W. Tyler, som fra begynnelsen av 1930-tallet ledet evalueringer av store skoleprogrammer i USA (Hogan, 2007). Den klassiske tilnærmingen innenfor målorienterte modeller innebærer at

tiltakets eller programmets hensikter defineres ved igangsettelsen, slik at man i etterkant av tiltaket kan måle i hvilken grad målsettingene er realisert. Denne tilnærmingen har en logisk og vitenskapelig tilnærming, som er anvendbar for evaluatorene, fordi kriteriene eller standarden som resultatene skal måles mot er forhåndsdefinert. Kritikken mot denne modellen er at evalueringen kan bli for snever, fordi ikke alle virkninger av tiltaket studeres (kun de som er definert i prosjektets målsettinger inkluderes), og fordi enkelte effekter kan ha egenskaper som gjør dem mindre målbare (Jabeen, 2016). Dette kan bidra til en skjevfordeling (bias) av forskningens resultater og dermed også i kunnskapsutviklingen. Et eksempel på en typisk målorientert evaluering, er den internasjonale PISA¹⁷-undersøkelsen i skoleverket, hvor elevenes kunnskapsnivå i de enkelte skolefag måles som en indikator for måloppnåelse og suksess i undervisningssystemet.

For å veie opp mot begrensningene i den opprinnelige målorienterte evalueringsmodellen, er det etter hvert videreutviklet ulike former for *systemorientert evaluering* som i større grad fokuserer på tiltaket i et systemperspektiv (Walton, 2014). Dette betegnes også med det engelske begrepet *complexity evaluation* eller med det spesifikke uttrykket *CIPP-evaluering* (Stufflebeam, 1971), etter en modell basert på kontekst (C), innsats (I), prosess (P) og produkt (P). Denne tilnærmingen innebærer at analyser av innsats/virkemidler, prosess og resultat vurderes både opp mot forhåndsdefinerte målsettinger og i lys av en kontekst i form av rammebetingelser som styrker eller hindrer måloppnåelse (Hansen, 1991). Bruk av systemevalueringer representerer en utvidelse av det målorienterte perspektivet, og gir rom for vurdering av utilsiktede effekter; altså positive eller negative bieffekter som ikke var planlagt ved utforming av tiltaket. Det er viktig å merke seg at systemperspektivet også kan benyttes innenfor andre evalueringsretninger enn den målorienterte.

Den viktigste innvendingen mot den målorienterte modellen, er at den er beskrivende og ikke forklarende (Tornes, 2013), i den forstand at den kun avdekker måloppnåelse og ikke fanger opp årsaker til at målene oppfylles i mer eller mindre grad. Dette blir også omtalt som *black box*-problematikk, hvor evalueringen vektlegger det som puttes inn i programmet og hva som kommer ut; men vier lite oppmerksomhet til prosessene som foregår imellom og som faktisk forårsaker programmets virkninger. Dette gir i liten grad innsikt i tiltakets forbedringspotensial.

3.3.2 BRUKERORIENTERT EVALUERING

Den brukerorienterte evalueringen retter fokus mot brukernes meninger og interesser isteden for forhåndsdefinerte mål eller kriterier. Brukere kan i denne sammenheng omfatte både myndigheter (beslutningstakere), industri (produktutviklere), enkeltindivider (som berøres av tiltaket), forskere,

¹⁷ Programme for International Student Assessment (PISA)

andre interessegrupper og samfunnet for øvrig, og evalueringen har til hensikt å verdsette utbyttet for de ulike aktørene. Evalueringer innenfor denne tilnærmingen benevnes også *stakeholder-evaluering* eller *interessent-evalueringer*. Nytteaspektet og sosial kontekst står sterkt i denne tilnærmingen, og evalueringskriteriene som benyttes er således de enkelte interessenters behov og verdier, og hvordan disse påvirkes av tiltaket.

Pattons (2012) *utilization-focused evaluation* er et eksempel på en evalueringstilnærming med sterkt brukerfokus, hvor nytteaspektet og brukernes perspektiver vektlegges. Det brukerorienterte perspektivet er også sentralt i Guba og Lincolns (2001) *konstruktivistiske modeller* som også benevnes *forth generation evaluation*, hvor blant annet subjektive opplevelser av virkningene av et tiltak legges til grunn, i motsetning til bruk av objektive kriterier. Innenfor den brukerorienterte retningen benyttes gjerne kvalitative og dialogbaserte metoder. Typisk anvendelse av den brukerorienterte tilnærmingen er for eksempel i evalueringer av virksomheters organisasjonsstruktur, hvor søkelyset rettes mot sentrale elementer som kultur, kommunikasjon og menneskene i organisasjonen. Slike evalueringer kan benyttes for å dokumentere i hvilken grad virksomheten er i stand til å oppfylle en gitt strategi, og dermed danne et kunnskapsgrunnlag for justering eller endring av styringssett eller organisasjon.

En styrke ved å involvere interessegrupper i evalueringen, er at selve prosessen kan bidra til å forsterke tiltakets effekter gjennom læring og bevisstgjøring hos involverte aktører. Når brukernes læringsprosess og eierskapsfølelse blir hovedmålet med evalueringen, omtales den gjerne som en *empowerment-evaluering*. Dette er blant annet en mye brukt innfallsvinkel i folkehelseprosjekter. Utfordringer med den brukerorienterte modellen er gjerne knyttet til håndteringen av uklare og motstridende interesser hos involverte aktører, samt at evalueringsarbeidet fort får et meget stort omfang hvis alle interessegrupper skal inkluderes (Pawson & Tilley, 1997).

3.3.3 PROSESSORIENTERT EVALUERING

Den prosessorienterte evalueringen legger vekt på selve implementeringen av et tiltak eller program, for å skape forståelse for og innsikt i hvilke strukturer og forhold knyttet til iverksettelsen av programmet som påvirker de resultatene som oppnås. Mens den målorienterte evalueringen fokuserer på *hvorvidt* og *i hvilken grad* måloppfyllelse skjer, avdekker prosessevalueringen *hvordan* virkningene oppstår, og dermed *hvorfor* et tiltak virker eller ikke virker (Saunders et al., 2005). Dette er blant annet nyttig når man ønsker å bruke evalueringsresultatene i planleggingen av nye tiltak (f.eks. bruke tiltaket flere steder). I en prosessevaluering evalueres hele årsakvirkningskjeden, fra programmets innsats til resultater og virkninger. Evalueringen foregår som regel samtidig med

implementeringen av tiltaket, slik at det kan gis innspill om justeringer og forbedringer etter hvert som ulike forhold avdekkes (formativ), men en prosessevaluering kan også gjennomføres i etterkant av tiltaket (summativ). Aksjonsforskning og følgeforskning (tilnærminger hvor forskeren er involvert i eller tett på området det forskes på) er eksempler på type forskning som typisk innehar prosessperspektivet (Sverdrup, 2002). Prosessperspektivet utelukker heller ikke andre evalueringsteoretiske perspektiver, og en prosessevaluering kan godt utføres komplementært til andre evalueringer (f.eks. den målorienterte). Eksempler på anvendelse av prosessevalueringer omfatter typisk flerårige evalueringer av offentlige reformer, som NAV¹⁸-reformen (2006-2014) og Samhandlingsreformen¹⁹ (2012-2015).

3.3.4 MÅLFRI EVALUERING

I det målfrie evalueringsspektivet fokuseres det på vurdering av de totale virkningene av et tiltak eller program, uten at disse knyttes opp mot et forhåndsdefinert mål eller verdigrunnlag. Hensikten med denne tilnærmingen er å sikre at alle effekter og konsekvenser av et tiltak avdekkes, enten disse er tilsiktet ved iverksettelsen av tiltaket eller ikke. Idealet er at evalueringen gjennomføres av en ekstern aktør som ikke har kjennskap til hensikten med tiltaket, og som derfor kan vurdere hva tiltaket faktisk har oppnådd, uten å være forutinntatt (Youker, 2013). Søkelyset rettes da mot innsats, effekter og nytte, og det anvendes gjerne et før-/etterperspektiv hvor man dokumenterer hva som har skjedd og i hvilken grad dette kan knyttes opp mot tiltaket (Hansen, 1991). Scriven (1991) argumenterer for at det målfrie evalueringsspektivet er teorinøytralt og således kan benyttes som del av eller supplement til andre evalueringsteorier, så lenge disse ikke er basert på en målorientert tilnærming. En variant av den målfrie tilnærmingen kan være å relatere virkninger til interessentenes behov istedenfor programmets mål.

Rene målfrie evalueringer forekommer sjelden, og et søk blant 2462 rapporter i Evalueringsportalen²⁰, ga ingen treff på evalueringer som eksplisitt har benyttet en målfri evalueringstilnærming. Likevel har perspektivene innenfor denne teoretiske retningen hatt betydning i form av å øke bevisstheten omkring bieffektene av et program eller tiltak, på en slik måte at evalueringer i dag ofte søker å avdekke eller omfatte også utilsiktede effekter.

¹⁸ NAV-reformens målsetting var å oppnå bedre arbeidsretting, bedre brukerretting og mer effektiv forvaltning.

¹⁹ Samhandlingsreformens målsetting var å oppnå helhetlige og koordinerte helse- og omsorgstjenester.

²⁰ www.evalueringsportalen.no

3.3.5 DEVELOPMENTAL EVALUERING

Developmental²¹ evaluering er en ganske nylig utviklet tilnærming for håndtering av komplekse evalueringssituasjoner, hvor hensikten er å bidra til innovasjon og læring (Gamble, 2006; Patton, McKegg, & Wehipeihana, 2015; Pawson, 2013). Den developmentale evalueringsmodellen er svært brukerfokuset og prosessorientert, men ikke avgrenset til spesielle metodevalg eller evalueringsdesign. Den legger vekt på at evalueringen skal foregå parallelt med utviklingen av tiltaket eller innovasjonen, med kontinuerlig tilbakemelding i sanntid til involverte aktører. Tilnærmingen er lite resultatorientert og har et internt fokus, for eksempel rettet mot etterlevelse av organisasjoners verdier, strategier og visjon. Evalueringer innenfor dette perspektivet tar utgangspunkt i en organisasjon eller beslutningstaker som ennå ikke vet hvilken intervensjon og type implementering som bør velges, enten fordi a) problemet ikke har en kjent løsning, b) det er flere mulige retninger fremover, og man vet ikke hvilken man skal velge, og/eller c) det er behov for innovasjon, utprøving og eksperimentering.

I motsetning til formativ prosessevaluering (som har til hensikt å forbedre et eksisterende tiltak) og summativ målorientering (som har til hensikt å dokumentere effektene av et tiltak), er formålet med den developmentale evalueringen å bestemme retning for videre satsing. Dette er spesielt relevant i komplekse, usikre eller dynamiske situasjoner, når man står overfor innovasjoner (nye løsninger med ukjent effekt) og når målet med tiltaket er potensielt store endringer. Typisk anvendelse synes å være innenfor sosialt entreprenørskap for bekjempelse av komplekse utfordringer som sult, vold, fattigdom og kriminalitet. Tilnærmingen er preget av fleksibilitet og pragmatisme i metodevalg, systemtenkning og langsiktighet, stor grad av involvering av både interessegrupper og evaluator, med fokus på innovasjoner, systemendring og læring av evalueringen mens den pågår (Gamble, 2006; Patton, 2006; Patton et al., 2015). Kompleksiteten og mangfoldet i evalueringstilnærmingen stiller særskilte krav til evaluatoren, både med hensyn til helhetlig metodekompetanse, analytiske ferdigheter og forståelse for innovasjonen som skal evalueres.

I Gamble (2006) presenteres et eksempel på vellykket bruk av developmental evaluering, hvor det canadiske skole- og nærmiljøprogrammet *Green Street* ble omstrukturert og videreført på bakgrunn av evalueringen som blant annet avdekket et stort forbedringspotensial i programmets

²¹ Begrepet er forholdsvis nytt, og det er ikke registrert noen formell norsk oversettelse i forbindelse med dette doktorgradsarbeidet. På nettstedet www.evalueringportal.no som drives av Direktoratet for økonomistyring og Det danske evalueringsselskaps nettsted (www.danskeevalueringsselskab.dk) benyttes det engelske begrepet *developmental evaluering*, med forkortelsen DE. En direkte oversettelse til *utviklingsorientert evaluering* kan skape misforståelse da dette begrepet også benyttes i forbindelse med formative evalueringer generelt.

beslutningsprosesser. Developmental evaluering inngår som et eksempel blant flere kompleksitetsorienterte tilnærminger.

3.3.6 TEORIBASERT EVALUERING

Den teoribaserte evalueringen²² har et systemperspektiv som innebærer at tiltaket settes inn i en kontekst med rammebetingelser og eksterne faktorer som påvirker hvorvidt man lykkes med iverksettelsen. Tilnærmingen bygger på en grunnleggende teoretisk forståelse av tiltaket (eller programmet) som skal evalueres, etterfulgt av en empirisk undersøkelse av hvordan tiltaket fungerer i praksis. Grunntanken innenfor den teoribaserte fremgangsmåten er at evalueringens utforming og metodebruk må styres av et konseptuelt rammeverk kalt programteori (Chen, 2012). Programteorien til et tiltak, prosjekt, program eller annen intervensjon, er en eksplisitt modell av sammenhengen mellom tiltaket og dets forventede effekter i form av i) *hvordan tiltaket forventes å medføre endringer*, ii) *hvorfor og i hvilken kontekst endringer eller resultater forventes å inntreffe*, samt iii) *hvilke konsekvenser resultatene vil medføre på kort og lang sikt*. Programteorien representerer således tiltakets underliggende logikk, i form av prosjekteiers forventninger til hvordan og hvorfor planlagte virkemidler skal føre til et ønsket resultat. På denne måten skiller det seg fra målorienterte black-box-evalueringer ved at de ikke bare skal gi svar på *om* et tiltak virker eller ikke, men også *hvorfor* og *hvordan* det virker.

Programteorien fremstilles gjerne i form av et diagram eller logisk rammeverk (logic frame), som kan være både lineært, hierarkisk eller matrisebasert, og som tilpasses det aktuelle prosjektet. Programteorien kan enten brukes som et fast referansepunkt gjennom hele evalueringen, eller revideres underveis i prosessen (dynamisk). En fordel med bruk av programteori som evalueringsgrunnlag, er at det er egnet som kommunikasjonsverktøy for avklaring av forventninger mot oppdragsgiver eller interessenter underveis i prosessen (Kaplan & Garrett, 2005). Videre fungerer programteorien også som et rammeverk for utvikling og kvalitetssikring av den operasjonelle evalueringsplanen.

Teoribaserte evalueringer er ansett å ha sin styrke i evnen til å håndtere komplekse tiltak med uklare årsakssammenhenger, og er egnet til å fremskaffe informasjon som er formålstjenlig for forbedring av eksisterende eller fremtidige tiltak (Blamey & Mackenzie, 2007; Hills & Junge, 2010). Den teoribaserte tilnærmingen trekkes også frem som spesielt relevant for beslutningstakere (Riché, 2012; van der Knaap, 2004), fordi den både vurderer tiltakets (potensielle) virkning, samt

²² *Theory-driven evaluation* eller *theory-based evaluation* på engelsk

fremskaffer kunnskap om hvordan tiltaket bør implementeres og hvilke rammebetingelser som må være oppfylt for å lykkes.

Til tross for at begrepet *programteori* er relativt nytt (Karlsen & Jentoft, 2013), er det ifølge Funnel og Rogers (2011) blitt vanlig i de fleste evalueringstilnæringer. Basert på tilgjengelig litteratur kan det imidlertid synes som om bruk av programteori som eksplisitt metode stort sett er knyttet til evaluering av sosiale programmer innenfor eksempelvis helse, utdanning og kriminalomsorg, med noen unntak. Et norsk eksempel på bruk av teoribasert evaluering er dokumentert i Vareide, Kobro og Storm (2013) som har etablert og anvendt en programteori for attraktivitet i et stedsutviklingsperspektiv, på oppdrag fra Kommunal- og moderniseringsdepartementet.

3.3.7 REALISTISK EVALUERING

Realistisk evaluering ble introdusert som en alternativ og realismeorientert evalueringsmodell av de engelske kriminologiforskerne Tilley and Pawson på slutten av 1990-tallet (Funnel & Rogers, 2011; Pawson, 2002a, 2002b, 2013; Pawson & Tilley, 1997). Tilnærmingen er sterkt teoriorientert, og omtales også som en spesifikk underkategori av det teoribaserte evalueringsperspektivet (Blamey & Mackenzie, 2007). Den realistiske modellen baserer evalueringsgrunnlaget på utvikling av programteorier som uttrykkes ved konfigurasjoner av begrepene kontekst (context), mekanisme (mechanism) og virkning (outcome). Det bærende prinsippet i dette perspektivet, er at evalueringen skal gi svar på spørsmålene; *hva virker, for hvem, i hvilken kontekst og på hvilken måte?* I dette ligger en erkjennelse av at virkningene av et tiltak eller program oppstår i spesifikke (fordelaktige) situasjoner og implementasjoner, og for spesifikke deltakere eller målgrupper. Konteksten for tiltaket blir derfor en viktig del av programteorien (eller *CMO-konfigurasjonen* som det benevnes innenfor den realistiske evalueringsmodellen).

Det som skiller den realistisk orienterte tilnærmingen fra øvrig teoribasert tilnærming, er at programteorien (CMO) i større grad spesifiserer de underliggende mekanismene som utløser programmets virkninger, samt de situasjonsspesifikke forholdene som påvirker hvorvidt disse mekanismene vil inntreffe. Det realistiske evalueringsperspektivet er metodenøytralt (forutsetter ikke bruk av spesifikke metoder), men vektlegger vitenskapelighet i metodevalget. Viktige grunnprinsipper ved realistiske evalueringer er at forskningen skal forankres i kjent kunnskap og at den har et anvendt perspektiv, i den forstand at den skal frembringe kunnskap for beslutningstakere til utvikling av policy og praksis.

Pawson og Tilley viser eksempler på realistiske evalueringsprosjekter i sine publikasjoner (Pawson, 2013; Pawson & Tilley, 1997). Foruten disse, er det blant annet publisert en dansk medisinsk studie,

hvor den realistiske tilnærmingen anvendes og diskuteres i forhold til klassisk eksperimentell metode (Berger & Vrangbæk, 2011). Forfatterne konkluderer med at den realistiske evalueringen har klare styrker med hensyn til vurdering av komplekse og sosiale teknologier og med hensyn til å skape forståelse for hvordan kontekst innvirker på muligheten for å oppnå effekter. På den andre siden, kan den realistiske tilnærmingen bli ressurskrevende, og resultatene er ikke nødvendigvis egnet til å trekke generelle konklusjoner om hvilke teknologianvendelser som har størst effekt i en absolutt betraktning.

3.3.8 SAMMENFATNING AV EVALUERINGSPEPERSPEKTIVER

De foregående delkapitlene har vist hvordan evalueringer kan kategoriseres på ulike måter, blant annet ut fra formålet med evalueringen, hva som evalueres, tidspunktet for evalueringen og vitenskapsteoretisk ståsted. Evalueringperspektivene er dels konkurrerende, dels utfyllende, dels overlappende og dels komplementære. En evalueringsmetodikk kan ha et metodisk eller teoretisk utgangspunkt. I *metodestyrt evalueringer* vil en forhåndsbestemt forskningsmetode gi føringer for fremgangsmåte og analyser. Med en *teoribasert evalueringstrategi* bestemmes evalueringdesignet og metodebruk av formålet med evalueringen og forskningsspørsmålene som ønskes besvart.

Det skilles gjerne mellom evalueringer som har et *kontrollformål* og evalueringer som har et *læringsformål* (Finansdepartementet, 2005). Evalueringer med kontrollformål har til hensikt å fremskaffe kunnskap som grunnlag for sanksjoner og styring, gjennom å avdekke måloppnåelse eller effektivitet i en virksomhet eller prosjekt. I en slik evaluering er det viktig å opprettholde avstand mellom (ekstern) evaluator og oppdragsgiver, for å sikre objektivitet i resultatene. Evalueringer med læringsformål vektlegger i større grad involvering og kommunikasjon med involverte aktører, blant annet fordi hensikten med evalueringen er å fremskaffe kunnskap som grunnlag for selvinnsikt og forståelse.

Pawson (2013) konkluderer med at evalueringfeltet har gjennomgått en rask utvikling fra en tidlig summativ målorientert tradisjon til en situasjon hvor verktøykassa og metodebruken hele tiden utvides for å møte behovet fra stadig mer komplekse problemstillinger. I de nyere lærebøker innenfor feltet anbefales det gjerne å kombinere flere evalueringperspektiver, med mangfoldige problemstillinger og metodetriangulering i datainnsamling og analyser. Pawson og Tilley (1997), Pawson (2013) og Walton (2014) kritiserer imidlertid en slik ustrukturert pluralistisk tilnærming med argumentasjon om at a) man aldri vil få tilstrekkelig med ressurser til å forske både i bredden og dybden på alt, b) en for omfattende tilnærming gjør det vanskelig å holde fokus på det som er vesentlig og c) en overfylt verktøykassa gjør evalueringen uhåndterbar.

Hittil er det presentert en rekke evalueringsperspektiver på overordnet nivå. I neste delkapittel rettes søkelyset mot innholdet i verktøykassa; de konkrete metodene som benyttes i evalueringsarbeidet.

3.4 EVALUERINGSDESIGN OG METODER

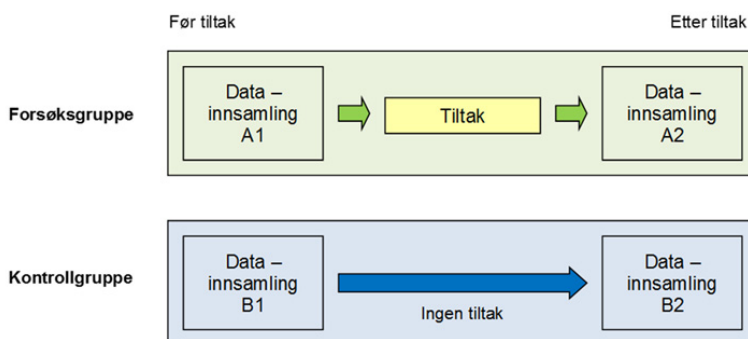
Mens evalueringstilnærmingen beskriver den overordnede strategiske evalueringsprosessen, handler metodevalget om den operative gjennomføringen med datainnsamling og analyser. Det er vanlig å skille mellom *kvantitative* og *kvalitative* evalueringsmetoder. De kvantitative metodene baserer analysene på et stort datamateriale (mange enheter) om et begrenset antall variabler som kan telles eller rangeres ved hjelp av for eksempel spørreskjema, registerdata eller logging av hendelser. Kvalitative metoder søker i større grad dybdekunnskap om flere variabler fra et fåtall enheter, gjennom intervjuer, observasjonsstudier eller dokumentanalyser. De kvantitative undersøkelsene gir typisk svar på hvor mange, hvor ofte eller hvor sterkt en effekt inntreffer, mens de kvalitative undersøkelsene kan forklare hvorfor og hvordan effekten inntreffer. De fleste større evalueringer vil involvere både kvalitative og kvantitative metoder.

Videre kan man skille mellom *induktive* evalueringsstudier som gjerne benytter kvalitative metoder for å utforske et fagfelt hvor det finnes lite kunnskap fra før, og *deduktive* studier som benytter hypotesetesting for å bekrefte eller avkrefte antakelser innenfor et fagfelt med godt etablert empiri. I det første tilfellet benyttes praktiske erfaringer til å fremskaffe kunnskap som kan bidra til å etablere ny teori – mens i det andre tilfellet testes etablert teori mot virkeligheten.

Effektanalyser av tiltak deles i hovedsak inn i to metodiske kategorier, som gjenspeiles i evalueringens forskningsdesign; *eksperimentell metode* og *ikke-eksperimentell metode*. Disse metodiske hovedretningene er nærmere beskrevet i påfølgende kapitler 3.4.1 og 3.4.2.

3.4.1 EKSPERIMENTELL METODE

Det eksperimentelle evalueringsdesignet innebærer en kvantitativ sammenligning av effekter før og etter innføring av et tiltak, og hensikten er å påvise at den observerte effekten er sann og allmenngyldig. I praksis betyr dette at man søker å finne en generell effekt som er overførbart til alle andre implementeringer av tiltaket. Metoden har sitt utspring fra naturvitenskapelig medisinsk forskning, hvor randomiserte kontrollerte eksperimenter (RCT) regnes som gullstandarden for kliniske laboratorieforsøk. Se illustrasjon av det eksperimentelle forsøksdesignet i Figur 7.



Figur 7: Det eksperimentelle evalueringsdesignet

Evalueringsdesignet innebærer at alle relevante data samles inn og sammenlignes i forsøksgruppen (A) og kontrollgruppen (B) både før og etter innføring av tiltaket. Den eksperimentelle fremgangsmåten legger til grunn en rekke forutsetninger som må være på plass for at evalueringen skal gi gyldige resultater (Campbell & Stanley, 1966; Shadish, Cook, & Campbell, 2002):

- *Randomiserte utvalg*: Deltakerne i eksperimentet skal være tilfeldig fordelt i en forsøksgruppe og kontrollgruppe, slik at man oppnår at forsøkspersonene i gjennomsnitt er lik i de to gruppene som sammenlignes. Hensikten med dette er å utelukke at eventuelle observerte forskjeller kan tilskrives egenskaper ved forsøkspersonene. Ofte vil praktiske eller forskningsetiske hensyn medføre at dette kravet ikke oppfylles i effektevalueringer innenfor transportområdet (i likhet med en del andre praktiske fagfelt hvor evalueringen foregår utenfor laboratoriet). Når randomiseringsprinsippet ikke er ivaretatt, benevnes det metodiske grunnlaget som kvasi-eksperimentelt. I det kvasi-eksperimentelle forskningsdesignet blir kontrollgruppen isteden konstruert med ulike matching-teknikker (fordeling av kjønn, alder og andre relevante egenskaper) for å etablere mest mulig like sammenligningsgrupper. Det ideelle forsøksoppsettet innebærer at fordelingen i forsøks- eller kontrollgruppe er ukjent både for forsøkspersonene og den som gjennomfører evalueringen (dobbelblindet studie).
- *Etablering av baseline med kontrafaktiske analyser*: Beskriver tilstanden som ville inntruffet dersom tiltaket ikke ble iverksatt. Innenfor transportområdet gjøres dette gjerne ved å benytte et sammenligningsalternativ (baseline) hvor ellers like trafikanter eller vegstrekninger sammenlignes over tid med forsøksgruppen som blir eksponert for et tiltak. Hensikten med dette er å utelukke at eventuelle effekter kan tilskrives en generell trend eller eksterne faktorer. Når kontrafaktiske analyser utelates fra evalueringsdesignet av praktiske eller ressursmessige hensyn, medfører dette utfordringer for resultatenes validitet (gyldighet).

- *Datainnsamling i før-situasjon:* Metodeoppsettet forutsetter innsamling av all relevant data i forkant av at tiltaket iverksettes, både i forsøksgruppen, kontrollgruppen og for sammenligningsalternativet. Muligheten for dette avhenger i stor grad av tidspunktet for evalueringen, da det kan være vanskelig å innhente tilstrekkelig informasjon dersom evalueringen ikke er planlagt når tiltaket innføres.
- *Entydig forhold mellom årsak og virkning:* Sammenhengen mellom tiltaket som innføres og effekten som utløses skal være entydig og allmenngyldig. Dette innebærer også at tiltaket opptrer stabilt og uforanderlig over tid, slik at datainnsamlingen kan foregå over en tilstrekkelig lang periode.
- *Kontroll over eksterne faktorer:* Eksterne faktorer som kan påvirke effektene av tiltaket må holdes konstant, slik at eventuelle effekter som observeres er "rene" effekter av tiltaket, uavhengig av kontekst og situasjonsbestemte faktorer.

Den (kvasi-) eksperimentelle metoden benytter statistisk hypotesetesting og falsifikasjon som grunnlag for å verifisere om et tiltak medfører en signifikant effekt. Når forutsetningene for modellen er oppfylt, er dette en vitenskapelig sterk metode for å dokumentere om et tiltak virker eller ikke. Denne fremgangsmåten danner også grunnlaget for den evidensorienterte forskningen som har til hensikt å akkumulere objektiv kunnskap om effektene av ulike tiltak, gjennom å samle resultatene fra gjentatte repliserbare studier i meta-analyser eller såkalte effektkataloger. Den eksperimentelle metoden omfatter imidlertid ikke forklaringsmodeller knyttet til kausale årsakskjeder (hvorfor et tiltak virker eller ikke) eller eksterne faktorer (i hvilke situasjoner et tiltak kan forventes å ha effekt).

3.4.2 IKKE-EKSPERIMENTELL METODE

Det ikke-eksperimentelle evalueringdesignet innebærer observasjonsstudier av et tiltak, uten sammenligning mot en kontrollgruppe. Dette metodeoppsettet benyttes når hensikten med evalueringen er å øke forståelsen for og kunnskapen om hvordan tiltaket fungerer i samspill med målgruppen og sine omgivelser. Fremgangsmåten stiller mindre formalistiske krav til tilgjengelig datagrunnlag, og benyttes også når betingelsene eller ressursene for å gjennomføre eksperimentelle studier ikke er til stede, eksempelvis hvis det ikke er praktisk eller etisk mulig å etablere en passende kontrollgruppe. Ikke-eksperimentelle metoder inkluderer blant annet observasjonsstudier, case-studier, retrospektive studier, tidsseriestudier, studier av ulike målgrupper og dokumentanalyser.

Hensynet til validitet forutsetter at det etableres kunnskap om og forståelse av hvordan tiltaket forventes å forårsake effekter i målgruppen (årsakssammenheng), for eksempel gjennom etablering av en velfundert programteori. Metodeoppsettet innebærer så at det gjøres systematiske vurderinger av hvorvidt resultatene fra studien er konsistent og samsvarer med den forhåndsdefinerte teorien. Denne fremgangsmåten benevnes gjerne som et teoribasert evalueringsdesign.

3.4.3 VALIDITET

Gode evalueringsstudier må utformes slik at det er grunn til å tro at resultatene viser de reelle virkningene av tiltaket som studeres. Dette benevnes som *validitet* eller *gyldighet*, og er et mål på studiekvaliteten; i hvilken grad forskningen gjenspeiler virkeligheten (at man måler det man har til hensikt å undersøke). Generelt skilles det mellom *indre validitet* som beskriver i hvilken grad resultatene er gyldige for det utvalget og fenomenet som er undersøkt, og *ytre validitet* som beskriver i hvilken grad resultatene kan overføres til andre utvalg og situasjoner. Elvik og Vaa (2004) opererer med fire hovedkriterier for studiekvalitet:

- *Statistisk validitet* betegner graden av tallmessig nøyaktighet, feilfrihet og representativitet. Dette punktet er knyttet til datainnsamling og statistiske analysemetoder.
- *Teoretisk validitet* betegner muligheten til å gi en teoretisk forklaring av resultatene fra en undersøkelse. Dette punktet er knyttet til begrepsbruk, hypoteser og virkningsmekanismer i et eksplisitt formulert teorigrunnlag.
- *Intern validitet* betegner muligheten for å hevde at det er årsakssammenheng mellom tiltak og effekt. Dette punktet er knyttet til studiens kausalitet (se kapittel 2.4.4).
- *Ekstern validitet* betegner i hvilken grad resultatene kan overføres til andre utvalg og situasjoner. Dette punktet er knyttet til muligheten for å generalisere resultatene av undersøkelsen.

Osland, Leiren, Hoff, Tennøy og Strand (2007) har brukt denne inndelingen som kvalitetskriterier i en gjennomgang av et utvalg evalueringsstudier innenfor transportsektoren. Rapporten konkluderer med at tre av kriteriene (teoretisk, intern og ekstern validitet) er hensiktsmessig for få frem styrker og svakheter ved ulike evalueringer og evalueringstradisjoner, men at det derimot er gode argumenter for å fravike kravene til statistisk validitet. Muligheten til å oppnå statistisk validitet avhenger av størrelsen på undersøkelsen; jo flere enheter (N) som inngår i utvalget, desto mindre blir den statistiske usikkerheten. Dette er hensiktsmessig for kvantitative analyser, men utelukker

bruk av kvalitative metoder. Kravet om statistisk validitet vil dermed innebære å forkaste kunnskap som kan være vesentlig, men som ikke kan fremskaffes gjennom statistiske analyser.

3.4.4 KAUSALITET

I evalueringsforskningen er det særlig interessant å identifisere årsakssammenhenger, fordi dette gir kunnskap om hvilke konsekvenser som vil følge av alternative tiltak. Bestemmelse av sammenhengen mellom en årsak og virkning benevnes med begrepet *kausaltet*. Dette er ikke ensbetydende med samvariasjon eller korrelasjon som beskriver sammenhengen mellom to variabler (og kan skyldes bakenforliggende faktorer). Det råder uenighet innenfor vitenskapsfilosofien om hvordan kausalitetsbegrepet skal defineres og praktiseres, og den vitenskapelige diskursen er gjort rede for i Tufte (2013), Finseraas og Kotsadam (2013) og Bhaskar (2008). Sentrale tolkninger av kausalitetsbegrepet inkluderer blant annet;

- *Regularitetsprinsippet*: Kausalitet forstås som en lovmessighet som opptrer automatisk i form av konstante observerbare sammenhenger, dersom tiltaket er til stede. Dette er den klassiske naturvitenskapelige tolkningen av kausalitetsbegrepet.
- *Det kontrafaktiske prinsippet*: Kausalitet forstås som en kontrafaktisk avhengighet som innebærer at en effekt ikke ville inntruffet uten at tiltaket var til stede.
- *Mekanismeprinsippet*: Kausalitet ses på som en generativ kraft som øker sannsynligheten for at en effekt vil inntreffe når tiltaket er til stede. En slik forståelse av kausalitetsbegrepet innebærer at fokus rettes mot å trekke konklusjoner basert på teoretisk validitet for spesifikke tiltak og målgrupper (lokal kausalitet) heller enn å avdekke universelle lovmessigheter.

Beviskravet for kausalitet er tett knyttet opp mot valg av evalueringsdesign. I det eksperimentelle designet ivaretas kravet av en metodisk stringens, hvor kausaliteten både forstås som en regularitet og bestemmes gjennom kontrafaktiske analyser. Ikke-eksperimentelle evalueringsdesign kan benytte ulike strategier for bestemmelse av kausalitet, inkludert det kontrafaktiske prinsippet som beskrevet i Finseraas og Kotsadam (2013), og det mekanistiske prinsippet som beskrevet i Pawson og Tilley (1997) og Bhaskar (2008). De forskjellige kausalitetsanalysene avdekker ulike former for kausalitet, og har alle sine styrker og svakheter som det er gjort nærmere rede for i Tufte (2013), Finseraas og Kotsadam (2013) og Pawson og Tilley (1997).

3.4.5 EVIDENSDEBATTEN

I dagens samfunn stilles det stadig større krav til og forventning om at politikktutforming og faglige avgjørelser skal være forankret i en kunnskapsbasert praksis (Alvsvåg, 2009; Davies, Nutley, & Smith, 2000; Johansson et al., 2015; Sanderson, 2002; Urban, Hargraves, & Trochim, 2014).

Begrepet *kunnskapsbasert praksis* er en oversettelse av det engelske *evidence based practice*, som opprinnelig ble introdusert i medisinsk forskning på 1990-tallet (Guyatt et al., 1992), og som senere har spredd seg til stadig flere samfunnsområder. Kjernen i evidenstankegangen er å finne ut *hva som virker* og anbefale dette til utforming av politikk og lokale beslutninger. I diskusjonen om hva som virker, blir uttrykkene *evidensbasert*, *forskningsbasert* og *kunnskapsbasert* brukt om hverandre (Skulberg, 2008), og begrepsforståelsen varierer både innenfor og mellom ulike fagfelt (Johansson et al., 2015).

Sackett, Rosenberg, Gray, Haynes og Richardson (1996) definerte opprinnelig evidensbasert praksis (i medisinfaget) som en kombinasjon av erfaringsbasert ekspertise og beste tilgjengelige evidens fra systematisk forskning, i en fortolkning som også inkluderer pasientens ønsker og behov. Ifølge Johansson et al. (2015) og Alvsvåg (2009) er imidlertid evidensbegrepet omstridt, og forståelsen av kunnskap/evidens deler seg i to atskilte retninger:

- A) Evidensbegrepet er avgrenset og ensbetydende med resultater som er fremkommet fra bruk av randomiserte kontrollerte forsøk (den eksperimentelle metoden, også omtalt som RCT). I denne sammenhengen tolkes evidens som den eneste sanne og akseptable kunnskap.
- B) Evidensbegrepet er tredelt og omfatter følgende komponenter i tråd med Sackett et al. (1996);
 - forskningsbasert kunnskap (randomiserte kontrollerte forsøk)
 - erfaringsbasert kunnskap (ekspertise)
 - brukervurderinger

Evidensdebatten er egentlig en metoddebatt, hvor eksperimentell metodikk settes opp mot ikke-eksperimentell metodikk. I følge Urban et al. (2014) har det vokst frem en kultur i løpet av de siste par tiårene som favoriserer RCT-basert evidens (A), noe som gjør at beslutningstakere føler seg presset til å prioritere eksperimentelle evalueringer (dokumentere tiltakets effektivitet), til og med når tiltaket er umodent og i en prematur fase av innføringsløpet. Internasjonale evidensorganisasjoner som Cochrane og Campbell Collaboration følger også en smal fortolkning av evidensbegrepet (A) i sitt arbeid med å etablere systematiske oversikter over effektene av tiltak, innenfor henholdsvis medisin og sosial velferd og utdanning. I denne sammenhengen handler evidens om hvorvidt tiltakene virker, og ulike evalueringsdesign graderes i henhold til sin evne til å produsere sikker kausal evidens om effekten av tiltak. Johansson et al. (2015) kommenterer at

tilhengere av evidensbevegelsen undervurderer ekspertkunnskap (professional wisdom) og brukererfaringer, til fordel for det vitenskapelige eksperimentet med randomiserte kontrollerte forsøk. De gjengir videre et evidenshierarki, hvor forsknings- og evalueringsmetoder rangeres i prioritert rekkefølge – med randomiserte kontrollerte forsøk ubestridt på topp (Johansson et al., 2015):

1. Randomiserte kontrollerte forsøk (RCT)
2. Kvasieksperimentelle studier (matchet kontrollgruppe)
3. Sammenligning før- og etter tiltak i en gruppe
4. Tverrsnittstudier med randomiserte utvalg
5. Prosessevaluering, formative studier og aksjonsforskning
6. Kvalitative case-studier og etnografisk forskning
7. Deskriptive retningslinjer og eksempler på god praksis
8. Ekspertvurderinger
9. Brukervurderinger

Dette skiller seg altså vesentlig fra den bredere forståelsen av evidensbegrepet (B), hvor både praktiske erfaringer (ekspertise) og brukervurderinger inkluderes i kunnskapsgrunnlaget. Flere er imidlertid kritisk til den RCT-orienterte retningen som innebærer at ikke all kunnskap er like gyldig, fordi det restriktive evidenskravet gjør at mange relevante studier blir utelatt (Alvsvåg, 2009; Johansson et al., 2015; Sackett et al., 1996; Skulberg, 2008; Urban et al., 2014). Sackett et al. (1996) understreker at evidens fra RCT-studier skal være til støtte for (informere) ekspertvurderingene, men at det aldri kan erstatte ekspertisen, da selv eksellent forskningsbasert evidens kan være ubrukelig (inapplicable) eller uhensiktsmessig (inappropriate) i den konkrete situasjonen hvor tiltaket vurderes brukt. Dette støttes også av Urban et al. (2014) som påpeker at selv om et tiltak er effektivt i kontrollerte omgivelser, er det ingen automatikk i at det vil fungere i den virkelige verden. Skulberg (2008) hevder også at evidensfokuset undervurderer sosiokulturelle faktorer og mangler blick for betydningen av kontekst – altså forhold som er knyttet til implementering og anvendelse av tiltaket. Han advarer videre mot å kopiere den naturvitenskapelige metoden (RCT) til samfunnsmessige problemstillinger som ikke lar seg rendyrke i laboratorier, fordi kravet om høy statistisk signifikans vil underkjenne relevant og viktig kunnskap. Higen etter å gjennomføre eksperimentelle evalueringer på umodne tiltak, kan medføre at potensielt lovende tiltak ser lite effektive ut, og dermed blir forkastet før de egentlig har fått en reell mulighet til å oppnå effekter (Urban et al., 2014). Til og med innenfor medisinsk forskning møter den evidensfokuserende tolkningen (A) kritikk, da flere fagområder (eksempelvis rusmiddelforskning og sykepleie) handler om komplekse sosiale årsakssammenhenger som ikke lar seg studere i

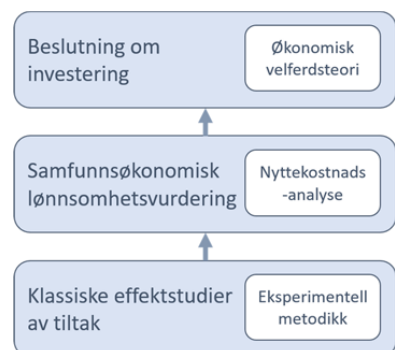
kontrollerte laboratorier (Alvsvåg, 2009; Ravndal, 2007). Alvsvåg (2009) advarer også mot at ensidig RCT-orientert evidensforståelse gir et innsnevret kunnskapsgrunnlag, men påpeker samtidig at RCT-basert evidens ikke er problematisk i seg selv, så lenge den metodiske fremgangsmåten begrenses til områder den har gyldighet på.

Uavhengig av hvor strengt man tolker evidensbegrepet, er ikke evidens en absolutt størrelse – men noe som foreligger i større eller mindre grad. Det er derfor utarbeidet flere standarder for klassifisering av evidensnivå til bruk i kunnskapsoppsummeringer (se flere eksempler innenfor helsesektoren presentert i Martinussen et al. (2016)). De ulike variantene benytter ulike kategorier og legger forskjellig beviskrav til grunn for inndelingen, men felles for alle synes å være at de tar utgangspunkt i hvor godt tiltakets virkning er dokumentert (virker tiltaket?). Det er i dette arbeidet ikke funnet tilsvarende eksempler på klassifiseringer av kunnskapsstatus som inkluderer andre forskningsspørsmål, for eksempel knyttet til aspekter ved gjennomføringen av tiltaket (implementeringsfaktorer, suksesskriterier eller barrierer).

3.5 EVALUERING I TRANSPORTSEKTOREN

Innenfor transportsektoren foregår det et betydelig evalueringsarbeid i regi av offentlige myndigheter, både i form av evalueringer som gjennomføres av fagfolk i offentlige etater (som f.eks. Statens vegvesen og kommuner) og ved bestilling av eksterne evalueringsoppdrag fra forskningsinstitutt og konsulentfirma. Tradisjonelt har evaluering innenfor transportområdet inngått som siste ledd i en synoptisk (rasjonell) planprosess, med hensikt å vurdere i hvilken grad valgte løsninger har bidratt til måloppnåelse (Statens vegvesen, 2014). Etter hvert har evalueringsarbeidet i større grad handlet om å fremskaffe beslutningsgrunnlag for offentlige investeringer og prioritering mellom tiltak i transportsektoren.

Beslutningsprosessen for offentlige investeringer i transportsystemet legger avgjørende føringer for det metodiske rammeverket innenfor transportfaget, som vist i Figur 8. Myndigheter skal sikre at bruk av offentlige midler gir størst mulig nytte for fellesskapet. Dette ivaretas gjennom *samfunnsøkonomiske lønnsomhetsvurderinger* – noe som fører til at nyttekostnadsorienterte evalueringer blir prioritert. De fleste evalueringsstudier er dermed rettet inn mot å fastsette nyttekostnadsverdien for et spesifikt transporttiltak, og/eller *effektstudier* som avdekker



Figur 8: Tradisjonelt evalueringsløp i transportsektoren

virkingen av et tiltak som input til fremtidige nyttekostnadsanalyser. Evalueringsfokus og metodeutvikling har dermed over tid blitt styrt mot indikatorer som inngår i nyttekostnadsanalysen.

I påfølgende avsnitt drøftes det vitenskapelige grunnlaget for transportforskning, deretter gjøres det nærmere rede for teori- og metodegrunnlaget for samfunnsøkonomiske lønnsomhetsvurderinger og effektevalueringer i transportsektoren.

3.5.1 TRANSPORTFORSKNING SOM VITENSKAP

Transportforskning er en forholdsvis ny vitenskap, som har vokst frem som et resultat av privatbilismens utbredelse fra midten av forrige århundre. I takt med at transport og mobilitet har spilt en stadig viktigere rolle i samfunnet, har man også opplevd de negative effektene av transport – i form av trafikkulykker, rushtrafikk og kødannelse, arealforbruk og barrierevirkninger, samt miljøproblemer i form av lokal luftforurensning og globale CO₂-utslipp. For å bøte på dette, har man forsøkt å komme frem til tiltak og virkemidler som motvirker de negative konsekvensene. Dette har avstedkommet en rekke effektevalueringer av tiltak og analyser med sikte på å avdekke hvilke faktorer som påvirker folks reisevalg og reiseatferd. Fagfeltet har således etablert seg som et resultat av samfunnets behov for å løse utfordringer knyttet til transportsektoren, med gradvis utvikling av ulike teoretiske rammeverk og modeller for å forklare trafikantatferd.

Transportforskningen har imidlertid ikke lyktes å oppnå konsensus om et enhetlig teoretisk grunnlag for fagområdet (Carsten, 2007; Elvik, 2004; Engström & Hollnagel, 2007; Ranney, 1994). Dette gjør at transportfaget kan fremstå som teoretisk umodent sammenlignet med en rekke andre tradisjonelle fagdisipliner, og kvalitetsbedømming av tiltak og prosjekter er i større grad knyttet til implementering av effektive løsninger enn til tradisjonell vitenskapelig publisering med fagfellebedømming.

Synet på hva som oppfattes som vitenskap og hvordan vitenskap oppstår er i stadig endring. Shapin (2008) beskriver hvordan vitenskapen i løpet av de siste hundre årene har beveget seg fra filosofiens søken etter sannhet, til en mer pragmatisk tilnærming som kommer til syne i form av ingeniørenes vektlegging av anvendbarhet. Dette kan illustreres med følgende utsagn fra Shapin (2008, s. 443):

Our culture used to insist on massive differences between science and technology and between the role of the scientist and that of the engineer. It's a distinction that now makes less and less sense: we're all engineers now, and the authority of science is increasingly based not on what scientists know but on what they can make happen.

Nowotny, Scott og Gibbons (2003) diskuterer hvordan kunnskapsdannelse og forskningsprosesser har gjennomgått en radikal transformasjon, som innebærer at det tradisjonelle, disiplinbaserte, akademiske regimet (modus 1) er erstattet av en mer tverrfaglig, problemløsningsorientert kunnskapsproduksjon (modus 2). Tabell 2 gir en oversikt over typiske karakteristikk av modus 1- og modus 2-forskning, i henhold til Hessels og van Lente (2008).

Tabell 2: Egenskaper ved modus 1- og modus 2-forskning (Hessels & van Lente, 2008)

Modus 1: Tradisjonelt akademisk	Modus 2: Tverrfaglig problemløsning
Akademisk kontekst	Anvendelseskontekst
Disiplinær	Transdisiplinær (flerfaglig)
Homogenitet	Heterogenitet
Autonomi/selvstendighet	Refleksivitet, sosial ansvarlighet
Tradisjonell kvalitetskontroll (fagfellebedømming)	Nye former for kvalitetskontroll

Transportforskningen kjennetegnes av en rekke egenskaper som knyttes til modus 2-teoriene med vekt på tverrfaglighet, problemløsning og anvendbarhet. Dette synes å være en generell trend for flere nye vitenskaper, hvor styring av forskningsprioriteringer, kommersialisering av forskning, samt økt bruk av vitenskapelige tellekantsystem bidrar til en generell endring av kunnskapsproduksjonen (Nowotny et al., 2003).

Fagfeltet fremstår som flerfaglig, med behov for stor grad av samarbeid mellom ulike fagprofesjoner for å fremskaffe kunnskap om komplekse problemstillinger i et anvendt perspektiv. Likevel synes fagfeltet i liten grad å utnytte mangfoldet i tilgjengelige evalueringsmodeller. Evalueringer innenfor transportsektoren er i stor grad styrt av beslutningsprosessen for offentlige investeringer, og dermed rettet mot vurdering av lønnsomhet og effektivitet. Metodegrunnlaget er basert på den økonomiske nyttekostnadsmodellen som rettesnor for samfunnsøkonomiske lønnsomhetsanalyser og den eksperimentelle modellen med hypotesetesting som ideal for effektevalueringer. Fagfeltet preges således av en positivistisk forskningstradisjon, som innebærer at forskningsmetodikken er en ufravikelig betingelse for at kunnskapen som fremskaffes oppfattes som eksplisitt og garantert (Hjørland, 1996). Transportforskning foregår imidlertid sjelden i kontrollerte laboratorier, og effektevalueringer er derfor preget av ulike kvasi-eksperimentelle studier (med eller uten kontrollgrupper og sammenligning), hvor den indre validiteten svekkes jo lengre bort man beveger seg fra det ideelle randomiserte kontrollerte designet.

3.5.2 SAMFUNNSØKONOMISKE LØNNSOMHETSVURDERINGER

Investeringer i offentlige prosjekter innenfor transportsektoren er tradisjonelt forankret i økonomisk velferdsteori. Dette innebærer at allokering av ressurser styres mot tiltak som gir størst mulig velferd (sosial nytte) per kostnad for samfunnet som helhet. Det økonomiske teorigrunnlaget forutsetter et prinsipp om rasjonell forbrukeratferd som legger til grunn at hver enkelt konsument (trafikanter) har full informasjon om sine valgmuligheter og vil oppføre seg rasjonelt med hensyn til optimering av personlig nytte.

Samfunnsøkonomiske lønnsomhetsvurderinger av veg- og transporttiltak innebærer en systematisk vurdering av alle relevante fordeler og ulemper tiltaket vil føre til for samfunnet. Evalueringen er i stor grad basert på nyttekostnadsanalyser som sammenligner alle nytteeffektene av et tiltak med kostnadene til etablering og drift av tiltaket, i en *ex ante* evaluering i en tidlig fase av prosjektet. Her undersøker man ikke hvilken nytte tiltaket har gitt, men benytter isteden fastlagt empiri fra tidligere undersøkelser. Metodikken forutsetter dermed at man har kjennskap både til hvilke nytteeffekter som kan påregnes, samt kostnadene som er forbundet med tiltaket. Dette innebærer også en pris- eller verdisseting av alle variablene som inngår i analysen, som grunnlag for en endelig monetær sammenstilling av forholdet mellom nytte og kostnad (ofte en nyttekostnads-brøk). Analysene innbefatter også en sammenligning av det aktuelle tiltaket med alternative løsninger, hvor nullalternativet (do nothing) som regel representerer baseline for sammenligningen. Håndteringen av kvalitative variabler som ikke inngår i selve nyttekostnadsanalysen kan variere noe mellom ulike modeller, men fokuset tillegges i stor grad den monetære sammenstillingen som gjerne suppleres med en oppstilling av øvrige variabler.

Den norske modellen er nedfelt i Statens vegvesens *Håndbok V712 Konsekvensanalyser* (2012a) som omfatter både prissatte og ikke-prissatte konsekvenser. Bruksområdet for evalueringer på dette nivået omfatter beslutningsgrunnlag for Nasjonal transportplan, konseptvalgutredninger, kommunedelplaner og andre overordnede planverk, hvor hensikten er å vurdere verdien av et prosjekt eller tiltak før det gjennomføres. Temaene som inngår i konsekvensanalysen er gjengitt i Tabell 3.

Tabell 3: Deltema som inngår i konsekvensanalysen iht. Statens vegvesens håndbok V712.

Aktører	Tema	Form	Deltema
Trafikant og transportbrukere	Nytte for trafikant og transportbruker	Prissatt	Distanseavhengige kjørekostnader, andre reiseutlegg, tidsbruk, ulempekostnader i ferjesamband og ved vegstengning, helsevirkninger av økt gang- og sykkeltrafikk, trygghet for gående og syklende.
Operatører	Operatørnytte	Prissatt	Operatørselskapenes (kollektivselskap, bompengeselskap, ferjeselskap, parkeringsselskap) kostnader, brukerinntekter og overføringer.
Det offentlige	Budsjettvirkning	Prissatt	Investering, drift og vedlikehold, tilskudd til kollektivtrafikk, skatteinntekter.
Miljø og samfunn	Trafikkulykker	Prissatt	Personskadeulykker og materiell- skadeulykker
	Restverdi	Prissatt	Framtidig nytte av tiltaket etter beregningsperioden
	Skattekostnad	Prissatt	Effektivitetstap knyttet til skattefinansiering, 20 % av offentlige utgifter
	Støy og luftforurensning	Prissatt	Støyplage innendørs. Lokal, regional og global luftforurensning.
	Landskapsbilde	Ikke prissatt	Endringer i ubebygde strøk, spredtbygde strøk, by og tettbygde strøk.
	Nærmiljø og friluftsliv	Ikke prissatt	Endrede kvaliteter i boligområder, identitetsskapende elementer, uteområder, friluftsområder, service, veg- og stinett for gående og syklende.
	Naturmangfold	Ikke prissatt	Inngrep i enkeltlokaliteter/forekomster og inngrep på landskapsnivå. Her under landskapsøkologiske sammenhenger, viktige naturtypeområder, verneområder etter naturmangfoldloven, viltområder, funksjonsområder i ferskvann, naturhistoriske områder og påvirkning av økologisk tilstand i vann.
	Kulturmiljø	Ikke prissatt	Inngrep i fornminner, samiske kulturminner, kulturmiljøer.
	Naturressurser	Ikke prissatt	Inngrep i områder for jordbruk, skogbruk, reindrift, fiske og havbruk, bergarter og malmer, vann.

Hovedverktøyet for gjennomføring av nyttekostnadsanalysen av de prissatte konsekvensene er beregningsverktøyet EFFEKT, hvor ferdig programmerte beregningsmoduler kalkulerer netto nytte og nyttekostnadsforhold med utgangspunkt i kostnader og relevante data om tiltaket, etablert empiri om tiltakets effekter og verdisetting av disse effektene.

Samfunnsøkonomiske analyser i forkant av at prosjekt eller tiltak iverksettes samsvarer med Finansdepartementets definisjon av evalueringsbegrepet (Finansdepartementet, 2005), men faller utenfor OECDs definisjon (OECD, 1991), som diskutert i kapittel 3.1. Det er et viktig poeng at disse analysene skal være sammenlignbare og minst mulig avhengig av aktøren som gjennomfører evalueringen. Verdivalget er fastlagt gjennom etablerte etatsstandarder, slik at det blir likt over hele landet. Dette er en forutsetning for sammenligning og prioritering mellom alternative tiltak.

3.5.3 EFFEKTEVALUERING

Transportanalyser benyttes som planleggingsverktøy for å vurdere konsekvenser av endringer i transportsystemet. Transportanalysene er gjerne bygd opp i fire trinn som inkluderer turproduksjon (antall reiser), turfordeling (hvor man reiser), reisemiddelvalg og rutefordeling (Statens vegvesen, 2012a). For komplekse analysesituasjoner er det utviklet matematiske modeller (trafikk-, transport- og godsmodeller) som hjelpemiddel til å vurdere dagens nivå og fremtidig utvikling.

Effektstudier benyttes for å avdekke eller måle virkninger av konkrete veg- og transporttiltak. På dette nivået kan evalueringer bidra til å etablere kunnskap om effekter, årsakssammenhenger og om hvordan tiltaket bør iverksettes for å oppnå best mulig effekt. Effektstudier kan i prinsippet omfatte både kvantitative og kvalitative undersøkelser, og gjennomføres i ulike faser av et prosjekt, eksempelvis ved ex-post evaluering av allerede implementerte tiltak, i feltforsøk, eller ved simuleringer og modelleringer av planlagte tiltak.

Hvordan bestemmes så nytteeffekten av et tiltak? Selv om det finnes et stort utvalg evalueringsmodeller tilgjengelig (som vist i kapittel 3.3), bygger effektevalueringer innenfor transportfaget hovedsakelig på en målorientert/ beslutningsorientert tilnærming med fokus på måling av resultater og effektivitet i relasjon til forhåndsdefinerte målsettinger (Statens vegvesen, 2012a). Metodeoppsettet baseres på det kvasi-eksperimentelle evalueringsdesignet med sammenligning av effekter før og etter implementering av tiltak, kombinert med kontrafaktiske analyser av transportutviklingen i et nullalternativ. Selv om transportevalueringer søker å oppnå et eksperimentelt forsøksdesign, er det sjelden mulig å få til i praksis, og man ender som regel opp med varianter av før- og etterundersøkelser med ulik grad av kontroll. Dziekan et al. (2013) presenterer fem vanlige forenklinger av det eksperimentelle forsøksdesignet i transportevalueringer, gjengitt og illustrert i Tabell 4.

Tabell 4: Eksempler på vanlige forsøksdesign i transportevalueringer, rangert etter vitenskapelig styrke.

<p>1) Kvasi-eksperimentelt evalueringsdesign:</p> <p>Pre-/posttest (A1, A2) med kontrollgruppe (B1, B2), men uten randomiserte utvalg</p>	<p>Før tiltak</p> <p>Etter tiltak</p> <p>Forsøksgruppe</p> <p>Kontrollgruppe</p>
<p>2) Kvasi-eksperimentelt evalueringsdesign:</p> <p>Pre-/posttest (A1, A2) med kontrollgruppe kun i etterundersøkelsen (B2)</p>	<p>Før tiltak</p> <p>Etter tiltak</p> <p>Forsøksgruppe</p> <p>Kontrollgruppe</p>
<p>3) Kvasi-eksperimentelt evalueringsdesign:</p> <p>Kun posttest (A2) med kontrollgruppe (B2) uten baseline</p>	<p>Før tiltak</p> <p>Etter tiltak</p> <p>Forsøksgruppe</p> <p>Kontrollgruppe</p>
<p>4) Før- og etterundersøkelse:</p> <p>Pre-/posttest uten kontrollgruppe (A1, A2)</p>	<p>Før tiltak</p> <p>Etter tiltak</p> <p>Forsøksgruppe</p> <p>Kontrollgruppe</p>
<p>5) Etterundersøkelse:</p> <p>Kun posttest (A2) uten baseline og uten kontrollgruppe</p>	<p>Før tiltak</p> <p>Etter tiltak</p> <p>Forsøksgruppe</p> <p>Kontrollgruppe</p>

Dziekian et al. (2013) peker på at svak metodeforståelse innenfor transportfaget gir varierende og til dels usikre resultater av evalueringsstudier av transporttiltak, og argumenterer for at kun en kombinasjon av effektevaluering og prosessevaluering kan gi sikker evidens for hvorvidt et tiltak er vellykket eller mislykket. Elvik (2008) gir eksempler på åtte vanlige varianter av pretest/posttest-oppsett i evaluering av trafikksikkerhetstiltak, fra enkle før- og etterstudier uten sammenligningsgrupper til studier som anvender varianter av empirisk Bayes metode²³.

Osland et al. (2007) har studert et utvalg målorienterte ex post effektstudier i samferdselssektoren. Et fellestrekk er at evalueringene i liten grad eksplisitt knytter seg til overordnede teorier – og dermed er de heller ikke underlagt særlig kvalitetskontroll og teoridiskusjoner. Videre pekes det på to vanlige typer transparensproblemer. Det første følger med bruken av komplekse statistiske analyser som innebærer at det er vanskelig for utenforstående å få innsikt i fenomenet som studeres. Det andre transparensproblemet er knyttet til kvalitative evalueringer der modellene som ligger til grunn i liten grad er formaliserte, slik at antagelser og resonneringer er skjult for omgivelsene.

Over tid bidrar resultatene fra stadig flere effektevalueringer til å etablere empiri om virkningen av spesifikke tiltak som grunnlag for standardiserte effektkataloger og input i de samfunnsøkonomiske lønnsomhetsvurderingene. Trafikksikkerhetshåndboka (Høye, Elvik, Sørensen, & Vaa, 2012) er et eksempel på en kilde som viser oversikt over resultatene fra relevante evalueringer innenfor et tiltaksområde.

²³ Statistisk metode for beregning av forventningsrette ulykkestall

4 EVALUERING AV ITS

Utprøving og evaluering av ITS-tiltak spiller en viktig rolle i å fremskaffe et hensiktsmessig beslutningsgrunnlag for en fremtidsrettet transportpolitikk. Tilgang på gode evalueringsmetoder har dermed avgjørende betydning for kunnskapsutviklingen. I dette kapittelet gjøres det rede for ITS-evaluering som fagfelt, med vekt på metodebruk og metodiske utfordringer.

4.1 KUNNSKAPSUTVIKLING

Evaluering av ITS kan ha flere formål, i tråd med de evalueringsteoretiske perspektivene som ble drøftet i kapittel 3. For det første er det behov for forskning som bygger kunnskap om hvordan ITS som virkemiddel påvirker trafikanter, transportsystemet og samfunnet for øvrig. Dette innebærer blant annet å dokumentere erfaringer med hvilke planlagte effekter som oppnås, hvilke utilsiktede effekter som må påregnes, hvordan tiltaket bør implementeres for å realisere sitt potensial, og i hvilken kontekst det forventes å gi best effekt. Rammebetingelser, brukeraksept og tillit til ITS-tjenestene har avgjørende betydning for virkningene som oppnås. ITS representerer et relativt nytt sett virkemidler innenfor fagfeltet, og det utvikles stadig nye tjenester og bruksområder. Det eksisterer derfor begrenset kunnskap om årsakssammenhenger og konsekvenser (Böhm et al., 2016; Meland, 2017; Odeck & Welde, 2010). Utprøving og evaluering av ITS med utgangspunkt i et læringsperspektiv, gir erfaringer som gradvis bidrar til et økende akkumulert kunnskapsgrunnlag.

Det andre formålet med evaluering er å fremskaffe kunnskap som grunnlag for *kontroll og styring*. I et slikt perspektiv vil hensikten med evalueringen være å dokumentere hvorvidt det oppnås tilfredsstillende måloppnåelse, effektivitet eller lønnsomhet i et tiltak eller prosjekt. Dette gir beslutningstakere grunnlag for å styre investeringer og prioritere mellom ulike tiltak og prosjekter innenfor transportsektoren.

Kunnskap om ITS-tiltak er etterspurt og nødvendig. Realisering av innovative tiltak forutsetter at effekter og lønnsomhet kan dokumenteres. Meland (2017) peker på at manglende innsikt i årsakssammenhenger og utilstrekkelig beslutningsgrunnlag forsinket innføringen av smarte trafikkløsninger i europeiske byer. Når myndighetene ikke tør å investere i ny teknologi uten at det foreligger tallfestet data om forventede kostnader, besparelser og nytteeffekter, blir potensialet for samfunnsnytte uforløst. Disse resultatene er i tråd med problemstillingen som drøftes i Odeck og Welde (2010). Her peker forfatterne på at fraværet av egnet evalueringsmetodikk og begrenset tilgang til data, øker risikoen for at ITS-tiltak blir nedprioritert til fordel for andre, mindre lønnsomme prosjekter i transportsektoren. Dette samsvarer også med funn fra en internasjonal studie av ulike lands praksis for beslutningstaking og investering i ITS, som konkluderte med at det

er vanskelig å fremskaffe et tilstrekkelig beslutningsgrunnlag, spesielt når det kommer til nyttevurderinger (Böhm et al., 2016; Flechl, 2017). Dette gjenspeiles også i det offentlige planarbeidet. Hughes, Falkmer, Anund og Black (2019) har vurdert 48 amerikanske statlige trafiksikkerhetsplaner med utgangspunkt i et systembasert sikkerhetsstyringsperspektiv, og konkluderer med at planene i for stor grad baserer seg på historiske perspektiver, uten evne til å respondere på et fremtidsbilde med hurtig utvikling, økt kompleksitet og mer usikkerhet.

Mitsakis, Grau, Iordanopoulos og Mintsis (2016) trekker frem en rekke potensielle nytteverdier og et mangfoldig målbilde som kan realiseres med ITS-tiltak. De etterlyser imidlertid evalueringer som håndterer helheten og synliggjør den fulle nytten av å implementere en ITS-tjeneste; i form av forståelse og kvantifisering av effekter og virkninger, tilrettelegging for fremtidige investeringsbeslutninger, samt optimalisering av eksisterende applikasjoner og tjenester.

ITS behandles i utgangspunktet på linje med øvrige veg- og trafikktiltak når effekter skal vurderes og prosjekter prioriteres. Tradisjonelt blir alle investeringer i transportsektoren målt i effekt på sikkerhet, miljø og effektivitet, mens lønnsomhet måles i form av nytte og kostnad (Böhm et al., 2016; Odeck & Welde, 2010; Statens vegvesen, 2012a, 2018). Böhm et al. (2016) beskriver en streng (rigorous) europeisk evalueringskultur som følger den eksperimentelle tradisjonen slavisk – og uten særlige refleksjoner rundt valg av evalueringsdesign. Det foreliggende metodiske grunnlaget er imidlertid tilpasset konvensjonelle transporttiltak, og når det anvendes på ITS-tiltak oppstår en del metodiske utfordringer (Bristow, Pearman, & Shires, 1997; Mitsakis et al., 2016; Newman-Askins, Ferreira, & Bunker, 2003; Odeck & Welde, 2010). Modellene har klare begrensninger når det gjelder evaluering av ITS, og problemene forsterkes ytterligere når man beveger seg fra frittstående ITS-applikasjoner til automatiserte systemer og samvirkende ITS. For å møte disse utfordringene, har ITS-evalueringer med utgangspunkt i klassiske effektstudier gjerne blitt supplert med kvalitative analyser av interne eller eksterne faktorer som påvirker resultatoppnåelsen (f.eks. teknologisk modenhet, brukeraksept og implementeringen).

Litteraturgjennomgangen belyser altså utfordringer med innføring av ITS-tiltak i transportsystemet på tre ulike nivå:

1. *Beslutningsprosessen*: Beslutninger om investering i transportsystemet favoriserer velprøvde tiltak som kan dokumentere effekter og lønnsomhet på forholdsvis kort sikt.
2. *Evalueringstradisjon og praksis*: En metodestyrt evalueringskultur basert på den eksperimentelle modellen dominerer fagfeltet. Begrenset datatilgang og manglende hensyn til

metodikkens forutsetninger og begrensninger kan gi utslag i usikre resultater og svakt beslutningsgrunnlag.

3. *Evalueringemetodikkens egnethet*: Dagens metodegrunnlag er tilpasset konvensjonelle tiltak, men er i mindre grad egnet til å håndtere en del typiske egenskaper og problemstillinger ved ITS-tiltak.

De metodiske utfordringene knyttet til ITS-evalueringer er nærmere beskrevet i de to påfølgende delkapitlene om henholdsvis samfunnsøkonomiske evalueringer og effektevalueringer av ITS-tiltak.

4.2 METODISKE UTFORDRINGER I SAMFUNNSØKONOMISKE EVALUERINGER AV ITS

Samfunnsøkonomiske evalueringer av ITS bygger hovedsakelig på økonomiske modeller hvor nytteeffekter ses i forhold til kostnader knyttet til etablering og drift av ITS-tiltaket. For tradisjonelle veg- og trafikktiltak tilbyr evalueringsrammeverket forholdsvis standardiserte prosedyrer med kalkulasjoner som forutsetter at effektene av tiltaket er kjent (empiri er etablert) og variablene som inngår i analysen kan kvantifiseres og prissettes. Nyttekostnadsanalysen har høy teoretisk validitet med en tydelig forankring i økonomisk velferdsteori, men møter noe generell kritikk knyttet til problematikk rundt verdsetting av ikke-monetære variabler (tid, miljø og ulykker) og manglende transparens (Osland et al., 2007). Når dette rammeverket brukes i evaluering av ITS-tiltak, oppstår en rekke særskilte utfordringer, både knyttet til vurdering av nytteeffekter og det økonomiske beregningsgrunnlaget som inngår i analysen. I avsnittene nedenfor gis en sammenfatning av sentrale problemstillinger basert på metodediskusjoner i Kolosz & Grant-Muller (2015), Newman-Askins, Ferreira & Bunker (2003), Odeck og Welde (2010), Stevens (2000, 2004), Bristow et al. (1997), samt norske erfaringer med ITS-evaluering dokumentert blant annet i Gjerstad and Bayer (2012), Høye, Sørensen, et al. (2011) og Thorenfeldt et al. (2011).

Manglende empiri

Den største utfordringen for samfunnsøkonomiske vurderinger av ITS-tiltak er at det foreløpig finnes begrenset erfaring med og kunnskap om tiltakenes effekter. Kunnskap om nyttevirksomheter utgjør hovedgrunnlaget for konsekvensanalysene, og data som benyttes som input i nyttekostnadsmetodikken skal ideelt sett fremkomme av faktiske implementeringer, hvor effektene er stabile over et lengre tidsrom og hvor resultater er kvalitetssikret i gjentatte studier (Elvik & Vaa, 2004). Slik evidensbasert kunnskap er i liten grad tilgjengelig for ITS-tiltak. Dette skyldes til dels at ITS er nye virkemidler med begrenset historisk empiri, men også utfordringer knyttet til effektmåling og validitet mellom ulike studier. Dette drøftes nærmere i kapittel 4.3 om

effektevaluering av ITS. Mangelfull empiri kan medføre at nytteverdiene som legges til grunn for den samfunnsøkonomiske analysen er beheftet med stor grad av usikkerhet.

Sensitivitet

En del typiske effekter av ITS-tiltak har et kvalitativt aspekt som ikke fanges opp av den tradisjonelle metodikken. Det er ikke gitt at alle tiltak vil gi nye, unike og målbare effekter. I mange tilfeller vil ITS isteden bidra til smidigere løsninger eller bedre forutsigbarhet for trafikantene. Faktorer som komfort, kvalitet på informasjon, pålitelighet, jevnere trafikkflyt, lik mobilitet for befolkningsgrupper eller utfordring av personvernet inngår vanligvis ikke i dagens konsekvensanalyser og får dermed ikke betydning når nytteeffekter skal vurderes. Den økonomiske modellen som legges til grunn for evalueringene har som forutsetning at trafikantene har full informasjon og agerer rasjonelt. Dette er en forenkling som medfører at deler av nytten ved trafikantinformasjonstiltak ikke behandles i evalueringen. Eksempelvis tar ikke transportmodellene høyde for tiltak som reisetidsinformasjon i sanntid eller effekten av at informasjon knyttet til reisemiddelvalg (ruter, tider og priser) samles på felles plattformer, til tross for at studier har dokumentert nytteverdi av informerte trafikanter (Herrala, 2007; Thorenfeldt & Bertelsen, 2014). Dette gjelder også for fordelingseffekter, hvor ITS-tiltaket kan bidra til økt nytte for spesielle befolkningsgrupper på bekostning av andre. Dersom man skal oppnå en reell sammenligning av ITS-tiltak og tradisjonelle tiltak, er det behov for sensitive metoder som også dekker de kvalitative aspektene ved virkemidlene.

Måling og verdsetting av kvalitative variabler

Det kan være vanskelig både å måle og sette verdi (en pris) på en del kvalitative effekter av ITS-tiltak. Nytteeffekter som økt komfort, redusert stress og frustrasjon, samt større forutsigbarhet i reisetid kan ha stor betydning for den enkelte trafikant – men i mange tilfeller er dette "ukjente" kvaliteter som er vanskelig å forstå verdien av for en bruker/trafikant uten tidligere erfaring med ITS-tjenesten. Tradisjonelle metoder for prissetting av variabler, som preferansestudier (Stated/Revealed Preference) og betalingsvilje (hedonistisk prissetting) har begrensninger med hensyn til verdsetting av nytteeffekter i slike tilfeller. Bruk av betalingsvilje som mål på kvalitative variabler representerer også en risiko for dobbelttelling av nytteeffekter.

Teknologisk risiko og fleksibilitet

Den kontinuerlige teknologiske utviklingen endrer betingelsene for nyttekostnadsanalysen over tid, ved at systemene blir stadig mer effektive i form av økt prosesseringskraft og tilgang til kommunikasjonsteknologi, mens kostnadene for elektronikk blir stadig lavere. Samtidig er marginalkostnaden for innføring av en ITS-tjeneste svært avhengig av eksisterende infrastruktur.

Dette gir implikasjoner både for kostnadsdelen av analysene og for levetidsbetraktningene som legges til grunn. Avansert teknologi gir større risiko for teknisk svikt og behov for raskere utskiftingstakt. I tillegg er den teknologiske levetiden gjerne kortere enn den tekniske levetiden, hvilket innebærer at systemet som ligger til grunn for ITS-tjenesten blir utdatert som følge av rask teknologisk utvikling (Böhm et al., 2016). Dette må tas hensyn til ved valg av tiltakets levetid i analysene. På den andre siden, innehar et ITS-tiltak ofte større fleksibilitet enn tradisjonelle vegtiltak. Dersom det bygges ny veg for å møte økt etterspørsel, er dette et varig tiltak med stor inngripen i omgivelsene – mens et ITS-tiltak som øker vegens kapasitet vil i større grad være reversibel dersom etterspørselen endres. Denne faktoren fanges ikke opp i dagens evalueringsmodeller.

Eksterne faktorer

Anvendelse og utbredelse av ITS-tiltak vil påvirkes av en del eksterne faktorer som jus, personvern, datasikkerhet, organisatoriske forhold og aksept hos befolkningen. Zlocki, Fahrenkrog og Lu (2016) drøfter tre ulike scenarier for markedsintroduksjon av ny teknologi; a) en frivillig eller valgbar ITS-applikasjon som inngår som del av kjøretøyet eller vegsystemet vil gi en langsom innføringsrate som er avhengig både av kundens brukeraksept og utskiftingstakten på infrastrukturen, b) en frivillig ITS-applikasjon som ettermonteres i eksisterende infrastruktur vil fortsatt være avhengig av brukeraksept, mens c) ITS-applikasjoner som blir introdusert som obligatorisk standardutstyr i kjøretøy eller vegsystemet, vil få en raskere innføringstakt og større utbredelse. De totale og endelige effektene av tiltaket vil til enhver tid være avhengig av oppnådd utbredelse i markedet. Dette gjør det nødvendig med kunnskap om brukeraksept, barrierer og markedspotensial i større grad enn for tradisjonelle vegtrafikktiltak.

Hittil er det presentert noen sentrale utfordringer knyttet direkte til den samfunnsøkonomiske vurderingen av ITS-tiltak. Analysen forutsetter imidlertid også gode input-data om nytteeffektene av tiltakene som vurderes. Denne kunnskapen fremskaffes gjennom effektevalueringer, som også kan være en krevende øvelse. Problemstillinger knyttet til effektanalysen er nærmere drøftet i neste delkapittel.

4.3 METODISKE UTFORDRINGER I EFFEKTEVALUERINGER AV ITS

Kartlegging og evaluering av nytteeffektene av ITS-tiltak er nødvendig for å etablere kunnskap om hvordan fremtidens intelligente kjøretøy og infrastruktur kan påvirke trafikanter og samfunnet for øvrig. Effektevalueringer i transportsektoren er i stor grad basert på kvasi-eksperimentelle modeller med statistisk hypotesetesting som grunnlag for å verifisere om et tiltak medfører en signifikant

effekt. Statens vegvesens ITS-veileder beskriver også et tradisjonelt evalueringsløp, hvor effektmålinger etter innføring av ITS-tiltaket sammenlignes med referansemålinger i en før-situasjon (Statens vegvesen, 2011a). Hovedelementene i evalueringen er trafikkregistreringer og trafikantundersøkelser, med fokus på blant annet endring i reisetid, trafikkvolum, hastighet og tidsluker – og det gjøres ingen videre drøftinger rundt metodevalg eller metodiske utfordringer. Det tradisjonelle metodegrunnlaget har imidlertid vist seg å ha en del begrensninger når ITS-tiltak skal evalueres. I avsnittene som følger gis en sammenfatning av noen sentrale metodiske utfordringer i effektstudier av ITS, hovedsakelig basert på metodediskusjoner i Martens & Jenssen (2012), Kulmala (2010), Jenssen (2010), Linder et al. (2007) og Newman-Askins, Ferreira & Bunker (2003), samt norske erfaringer blant annet presentert i Gjerstad og Bayer (2012), Høye, Sørensen, et al. (2011) og Thorenfeldt et al. (2011).

Teknologisk utvikling

Den eksperimentelle evalueringemetodikken forutsetter at ITS-tjenesten har oppnådd høy grad av teknologisk modenhet og at datainnsamling kan foregå over en tilstrekkelig lang periode, hvor teknologien virker stabilt og brukerne er godt kjent med tjenesten. I virkeligheten er ITS-tiltakene som regel basert på teknologi som er i rask utvikling, og systemene som evalueres har gjerne oppnådd ulik grad av teknologisk modenhet. Den generelle teknologiske utviklingen bidrar til at ITS-løsningene blir stadig mer effektive (bl.a. økt prosesseringskraft og lagringskapasitet), mens kontinuerlig utvikling av applikasjonene gir bedre funksjonalitet. Dette har betydning for effektene som oppnås og bidrar til at resultater fra studier av ITS sjelden er gyldige over tid. En litteraturstudie av førerstøttesystemet adaptiv cruisekontroll (ACC) viste for eksempel store variasjoner i trafiksikkerhetseffektene som ble avdekket i løpet av en tiårsperiode, hvor tidlige varianter av systemet hadde begrenset virkeområde (kun over 30 km/t), mens dagens systemer inkluderer funksjoner som gjør at de også fungerer i køsituasjoner og ved lavere hastigheter (Nordtømme, Jenssen, Lervåg, Hjelkrem, & Kummeneje, 2014). Et annet eksempel er knyttet til utprøvingen av selvkjørende minibusser. I forbindelse med at disse bussene ble introdusert i Norge, uttalte Olav Madland i Acando følgende til Teknisk Ukeblad²⁴: "*Alt vil ikke være løst i den første perioden. Vi legger opp til en agil utvikling, hvor vi sier at vi vil få inn de forbedringene vi ser er nødvendig fortløpende*". Denne fremgangsmåten er nok representativ for mange pilotprosjekt med ITS-tjenester, og forutsetningen om en perfekt og stabil løsning blir urealistisk. Vesensforskjeller i

²⁴ Artikkel i Teknisk Ukeblad 23. april 2016. Acando og Olav Madland sto bak innføringen av den første selvkjørende bussen fra EasyMile til Norge i september 2016.

systemenes egenskaper og teknologisk beskaffenhet over tid gjør det lite relevant å sammenligne resultater mellom studier gjennomført på ulike tidspunkt eller med ulike varianter av systemet.

Komplekse årsakssammenhenger

Det foreligger forholdsvis begrenset empiri fra faktiske implementeringer av ITS, og det er derfor vanskelig å etablere kunnskap om de kausale mekanismene som utløser umiddelbare resultater og langsiktige effekter av et ITS-tiltak. ITS representerer ofte komplekse og situasjonsavhengige årsakssammenhenger, med mange involverte aktører og gjerne synergieffekter fra samspill mellom ulike systemer. Dette gjør det også vanskelig å isolere effektene fra ett ITS-tiltak som man ønsker å studere, hvilket innebærer at den eksperimentelle metodikkens premisser om et entydig forhold mellom årsak og virkning ikke alltid er til stede. Eksempelvis drøfter Aust (2012) hvordan kombinerte effekter fra flere ITS-applikasjoner vanskeliggjør evalueringen av aktive sikkerhetssystemer i kjøretøy. I NordicWay-prosjektet som presenteres nærmere i kapittel 9.2.1, viste evalueringsarbeidet at den samme ITS-applikasjonen (varsling av glatt vegbane) utløste vidt forskjellige effekter avhengig av hvordan den ble implementert i trafikksystemet og hvem som var målgruppen for tjenesten (Bjerkan, 2017; Lervåg, 2016). I Norge ble applikasjonen brukt til forbedring av vintervedlikehold og driftsoperasjoner, mens den i Finland ble brukt som førerstøttesystem med varsling i sanntid til førere på den aktuelle strekningen. Tilsvarende erfaringer er også beskrevet i Belin, Tillgren, Vedung, Cameron og Tingvall (2010), hvor innføring av automatisk trafikkontroll (fotoboks) utløste forskjellige tiltak-respons-kjeder i ulike land. Forfatterne av denne artikkelen understreker betydningen av at den som tar beslutninger om innføring av teknologibaserte sikkerhetstiltak på veg innehar en helhetlig forståelse av systemene og deres operative funksjonalitet i den spesifikke konteksten de skal anvendes. Mitsakis et al. (2016) understreker også at evaluering av ITS-tiltak forutsetter både tilstrekkelig kunnskap om og forståelse av teknologien i seg selv, samt god innsikt i hvordan ITS-tjenestene opptrer og fungerer i samspill med trafikanter i virkelige omgivelser.

Tekniske aspekter og brukergrensesnitt

Effektene av et ITS-tiltak vil være avhengig av applikasjonens tekniske funksjonalitet, modenhet og brukergrensesnitt. ITS kan fremstå med stort sprik mellom det ideelle systemet (slik det beskrives av produsenten) og det reelle systemet (slik det virker i praksis). Dette forsterkes av at effektstudier av ITS ofte omfatter systemer som fortsatt er under utvikling (mer eller mindre modne piloter). I evalueringen må man dermed skille mellom systemets potensiale (effekter av et ideelt system) og systemets realistiske effekter (f.eks. observerte effekter) på det tidspunktet evalueringen gjennomføres. I tillegg vil systemets brukergrensesnitt (HMI) påvirke faktorer knyttet til

brukeraksept og trafikantatferd. To ellers like systemer kan med ulikt brukergrensesnitt resultere i forskjellige nytteeffekter, avhengig av brukernes aksept av og respons på tiltakene. Dette er blant annet vist i simulatorstudier av førerstøttesystemene *kjørefeltstøtte* i Moen, Lervåg og Jenssen (2009) og *hendelsesvarsling* i Lervåg, Moen og Nordtømme (2011).

Trafikantatferd og utilsiktede effekter

ITS-tiltak bidrar til å endre kjøreprosessen, stille nye krav til trafikantene og gi andre utfordringer enn det man opplever med tradisjonelle vegtrafikktiltak. Hvordan trafikanten oppfatter, tolker og reagerer på ITS-applikasjonene har stor betydning for eventuell adferdsendring. Nytteeffekten som oppnås i praksis vil derfor være avhengig av trafikantenes mentale arbeidsbelastning og respons på tiltaket. Dette er spesielt relevant for førerstøttesystemer i bil, hvor en del sikkerhetskritiske funksjoner er avhengig av en umiddelbar og riktig reaksjon fra føreren. Et dårlig utformet brukergrensesnitt vil for eksempel kunne resultere i distraksjon, økt reaksjonstid, stress, misforståelser og feiltolkninger. *Distraksjon* oppstår når håndtering av førerstøttesystemet krever visuelle eller kognitive ressurser som interfererer med gjennomføring av primæroppgaven (å holde kjøretøyet trygt på vegen). Bruk av mobiltelefon under kjøring er et eksempel på en distraherende oppgave som er blitt regulert gjennom vegtrafikkloven. *Atferdstilpasning* inntreffer når føreratferden endres som følge av ITS-tiltaket, uten at dette er tilsiktet ved utvikling eller implementering av tiltaket. Dette inkluderer blant annet faktorer som endring i oppmerksomhet, førerens plassering i kontrollprosessen og situasjonsbestemt årvåkenhet (Jenssen, 2010).

Atferdstilpasninger kan også handle om strategiske valg, eksempelvis dersom man på grunn av avanserte sikkerhetssystemer velger å utføre reiser på dager med dårlige vær- og føreforhold, hvor man ellers ville latt bilen stå. Studier av atferdstilpasning til førerstøttesystemer som for eksempel antiskrens (ESC), indikerer at omtrent 10 % av førerne kjører mer aggressivt eller under andre betingelser enn de ellers ville gjort (risikokompensasjon) fordi de har overdreven tillit til systemene (Jenssen, 2010; Rudin-Brown, Jenkins, Whitehead, & Burns, 2009). Effektevaluering av ITS-tiltak må derfor ta høyde for at utilsiktet føreratferd på sikt kan ha modererende virkning på de effektene som inngår i studien.

Etablering av baseline og kontroll av eksterne faktorer

Effektevalueringer innenfor den eksperimentelle tradisjonen forutsetter etablering av et sammenligningsgrunnlag (baseline), med innsamling av all relevant atferds- og trafikkdata i forkant av at ITS-tiltaket innføres. For tradisjonelle vegtrafikktiltak betyr dette at man sammenligner en situasjon med og uten tiltaket. Dette bildet er vesentlig mer komplekst for ITS i dagens trafikkilde, hvor mange kjøretøy og deler av infrastrukturen allerede er utstyrt med ITS-applikasjoner, og hvor

innføringen i praksis skjer gradvis. I flere tilfeller foregår den teknologiske utviklingen uavhengig av myndighetenes beslutninger, og sammenligning mot et nullalternativ uten ITS blir lite relevant. Når evalueringen foregår parallelt med utviklingen av ITS-tiltakene, medfører det en del utfordringer knyttet til selve datainnsamlingen og det kan være vanskelig å etablere et tilstrekkelig datagrunnlag. Thorenfeldt et al. (2011) påpeker blant annet at den gradvise innføringen av ITS-tiltakene *sanntidsinformasjon* og *signalstyring for kollektivtrafikk* i Trondheim gjorde det vanskelig å definere klare før- og etterperioder i evalueringsarbeidet, noe som resulterte i et begrenset og usikkert datagrunnlag. Den eksperimentelle modellen forutsetter at alle eksterne faktorer som kan påvirke resultatet av studien holdes konstant. Dette kan være vanskelig å oppnå i praksis, da effektene av ITS-tiltak ofte er svært avhengig av rammebetingelser og situasjonen rundt tiltaket. Dersom premissene for den eksperimentelle metodikken ikke er oppfylt, vil resultatene av effektevalueringen være beheftet med en viss grad av usikkerhet, avhengig av kvaliteten på datagrunnlaget og i hvilken grad man har kontroll over testbetingelsene.

Brukeraksept og markedsutbredelse

ITS forventes å ha et stort potensial i utviklingen av et bærekraftig transportsystem, og mange studier viser lovende effekter av ulike ITS-tiltak. Realisering av dette potensialet er imidlertid avhengig av at ITS-tjenestene blir tatt i bruk og oppnår en viss utbredelse i markedet. *Brukeraksept* er en viktig faktor med hensyn til å kunne predikere fremtidig opptak i markedet og dermed mulige virkninger på sikt. Trafikantenes holdninger og aksept påvirkes av erfaring og faktorer som opplevd nytte, tillit og brukervennlighet. Dette er subjektive forhold som endres over tid og gjennom opplevelser med tjenestene. ITS-tjenester som leveres av private aktører, er også avhengig av fornuftige forretningsmodeller og tilstrekkelig betalingsvilje for at tjenesten skal være levedyktig i et lengre tidsperspektiv. Mange ITS-tiltak utløser også indirekte virkninger i form av endret atferd hos kjøretøy eller trafikanter som ikke benytter seg av tjenesten. Eksempelvis viser forskning på hastighetsvalg at bilførere lar seg påvirke av hastigheten til øvrig trafikk (Rothengatter, 1988), noe som indikerer at en viss markedsutbredelse av ISA-utstyrte kjøretøy kan bidra til å senke hastigheten til øvrig trafikk. Eksplisitt forskning på eksterne virkninger av ITS vil imidlertid kreve omfattende og ressurskrevende forsøksdesign, og det er i liten grad funnet eksempler på slike studier.

Korttidseffekter, langtidseffekter og nettverkseffekter

Forskning på trafikantatferd viser at brukere av ITS-systemer endrer atferd over tid, avhengig av hvor man befinner seg i læringsprosessen (Martens & Jenssen, 2012). Etablering av en stabil føreratferd krever erfaringer med ITS-tjenestene over lengre tid og i ulike situasjoner. Studier av

langtidseffekter er imidlertid mer tid- og ressurskrevende enn studier av korttidseffekter. De fleste effektstudier av ITS-løsninger er basert på korttidsstudier som gir resultater som er representative for nye brukere i en innføringsfase. Bildet kompliseres også av at det ikke nødvendigvis er en lineær sammenheng mellom tiltaket som innføres og når effektene inntreffer. Walton (2014) peker på behovet for langsiktighet og anbefaler at evalueringer foregår parallelt med utvikling og innføring av tiltaket, for å fange opp plutselige transformasjoner og nettverkseffekter. Et eksempel på denne problemstillingen er nevnt i Gjerstad og Bayer (2012), hvor effektene av en samkjøringstjeneste først vil inntreffe når tiltaket har oppnådd et kritisk antall brukere som sikrer tilstrekkelig tilbud og etterspørsel.

4.4 ETABLERTE RAMMEVERK FOR EVALUERING AV ITS

4.4.1 FESTA-METODIKKEN

Utover 2000-tallet ble det gjennomført flere store europeiske pilotprosjekter med uttesting av førerstøttesystemer i bil (såkalte field operational tests/FOTs). I den forbindelse finansierte EU et bredt sammensatt konsortium med formål å utforme en felles metodikk for gjennomføring og evaluering av feltforsøk med avanserte førerstøttesystemer. Resultatet av dette arbeidet omtales som *FESTA-metodikken* og er dokumentert i en håndbok med praktiske anbefalinger for gjennomføring av feltstudier (FOT-NET, 2014). Metodikken skal sikre en systematisk og vitenskapelig tilnærming, samt muliggjøre evaluering av samfunnsmessige effekter ved inkludering av relevante indikatorer. Gjennom et standardisert forløp gir metodikken grunnlag for sammenligning og erfaringsutveksling mellom ulike europeiske feltstudier. Rammeverket er basert på den eksperimentelle evalueringsmodellen, i tråd med tradisjonelle veg- og trafikkevalueringer. Fremgangsmåten innebærer sammenligning av effekter på føreratferd før og etter implementering av ITS, og med vurdering av samfunnsmessige konsekvenser i et nyttekostnadsperspektiv.

En viktig del av forarbeidet til evalueringen er knyttet til definisjonen av felles prestasjonsindikatorer (KPIs), valg av målemetoder og registreringer, samt bruk av sensorer:

- *Prestasjonsindikatorer* (performance indicators) defineres med utgangspunkt i målbare variabler som uttrykk for den effekten man ønsker å studere, og kategoriseres i hovedgruppene *sikkerhet, miljø, effektivitet* og *aksept*.
- *Målinger og registreringer* (measures) uttrykker hvordan datamaterialet innhentes i henhold til kategoriene *direkte måling, utledet måling, selvrapportert måling, hendelser basert på definerte triggere* og *situasjonsvariabler*.
- *Sensorer* representerer eksempler på måleinstrumenter som kan benyttes.

Arbeidsprosessen i FESTA beskrives som et iterativt forløp som inkluderer a) *forberedelse av studien* med definisjon av funksjoner, brukersituasjoner, forskningsspørsmål og hypoteser, samt valg av prestasjonsindikatorer, målemetoder og sensorer, b) *gjennomføring av studien* med datainnsamling, samt c) *analyser* med hypotesetesting, effektanalyser og samfunnsøkonomiske nyttekostnadsanalyser.

FESTA har oppnådd bred anerkjennelse innenfor fagfeltet og anvendes i en eller annen form i de fleste større feltforsøk med førerstøttesystemer i Europa. Metodikken holdes levende og oppdateres stadig gjennom EU-finansiert støtte til nettverksarbeid, og det foregår per i dag diskusjoner i fagmiljøet om hvordan FESTA kan tilpasses feltforsøk med samvirkende ITS og høyere grad av automatiserte kjøretøy (Barnard et al., 2016).

Selv om FESTA-metodikken har bidratt til å styrke kvaliteten på gjennomføring av feltforsøk med ITS de siste årene, har den begrensninger når det kommer til evaluering av studier med samvirkende eller selvkjørende kjøretøy (Barnard et al., 2016; FOT-NET, 2016). Først og fremst er FESTA-metodikken utelukkende rettet inn mot feltforsøk med stabile prototyper eller såkalte nær-marked-løsninger. Det forutsettes at systemene som testes har høy grad av teknologisk modenhet, da de tekniske aspektene ved applikasjonene ikke behandles i evalueringen. Videre er metodikken rettet inn mot førerstøttesystemer, og har dermed begrensninger med hensyn til evaluering av andre aspekter enn den brukerfokuserte endringen i føreratferd. Ved evaluering av samvirkende ITS og selvkjørende kjøretøy er det behov for å utvide studieområdet til også å inkludere aspekter knyttet til selve kjøretøyet og samspillet med andre trafikanter og omgivelsene. Det er også interessant å undersøke hvordan mobilitet og transporttettersspørsmål påvirkes, samt eventuelle etiske problemstillinger. En av de største utfordringene er knyttet til valg av baseline; sammenligningsgrunnlaget for analysene. Å sammenligne teknologisk avanserte kjøretøy med et nullalternativ med utelukkende konvensjonelle kjøretøy er lite relevant, da transportsystemet for øvrig innehar stadig mer ITS. Å sammenligne mot den til enhver tid rådende sammensetning av trafikkbildet vil gi et flyktig grunnlag som er dårlig egnet for sammenligning av resultater over tid eller mellom ulike steder (metaanalyser). Det eksisterer per i dag ikke konsensus om hvordan baseline-problematikken skal løses, og Barnard et al. (2016) stiller spørsmål om konseptet med bruk av baseline i det hele tatt er hensiktsmessig for evaluering av fremtidens ITS-tiltak. Hvis man erkjenner at den teknologiske utviklingen av transportsystemet uansett vil finne sted, er spørsmålet om hvorvidt situasjonen var bedre før eller etter implementering lite relevant.

4.4.2 EASYWAY GUIDELINES

En rekke europeiske land deltok i EasyWay-prosjektet (2007-2013) som ble delfinansiert av EUs TEN-T-program for å stimulere til implementering av ITS og IKT-infrastruktur på veg (Tempo, 2008). I prosjektet ble det utarbeidet anbefalinger for utrulling av ITS-tjenester (Deployment Guidelines) og anbefalinger for gjennomføring av evalueringsstudier (Evaluation Guidelines), samt en mal for rapportering av evalueringsresultater. Hensikten med arbeidet var å samle erfaringsdata og dokumentere beste praksis for fremtidig implementering av ITS. Håndboken gir oversikt over og anbefalinger om praktisk gjennomføring av de ulike fasene i evalueringsprosessen; fra valg av evalueringsnivå (ex-ante/ex-post), via identifisering av brukerbehov, målsettinger og utarbeidelse av evalueringsplan, til analyse og rapportering av resultater (EasyWay EEG, 2012). Prosjektet identifiserer også en rekke relevante effekter og indikatorer for ITS-tjenester.

Evalueringsprinsippene i EasyWay er i tråd med øvrig evalueringstradisjon innenfor transportsektoren, basert på det eksperimentelle evalueringsdesignet med sammenligning av effekter i et transportsystem med og uten ITS-løsninger, og vurdering av samfunnsmessige konsekvenser i et nyttekostnadsperspektiv. EasyWay-arbeidet gir en grei oversikt over evalueringsprinsipper og god vitenskapelig praksis for den som skal gjennomføre prosjektevalueringer – men anbefalingene presenterer ikke løsninger på de metodiske utfordringene som er presentert i delkapitlene 4.2 og 4.3. Sammenlignet med FESTA har EasyWay et bredere anvendelsesområde, da det ikke er begrenset til feltforsøk eller har utelukkende fokus på førerattferd, men det gis ikke anbefalinger på samme detaljnivå som FESTA. Tilnærmingen har for øvrig mange av de samme begrensningene som FESTA-metodikken er beheftet med, og den behandler ikke problemstillinger som særskilt er knyttet til evaluering av samvirkende ITS og selvkjørende kjøretøy.

4.4.3 RAMMEVERK FOR VURDERING AV TRAFIKKSikkerhet

I tråd med økende anvendelse av teknologi i transportsystemet, er det utviklet en rekke atferdsteorier som inkluderer kjøring med førerstøttesystemer (se f.eks. Carsten, 2007; Engström & Hollnagel, 2007; Vaa, 2007). Dette har resultert i forskning på sentrale fenomener som distraksjon, atferdstilpasning, læringseffekter og brukeraksept (Bengler, 2007; Kujala, 2010; Martens & Jenssen, 2012). Slike studier har i stor grad rettet søkelyset mot trafiksikkerhetseffektene av førerstøttesystemer. Kulmala (2010) presenterer i forlengelsen av dette et forslag til rammeverk for vurdering av trafiksikkerhetseffekter ved innføring av ITS-tiltak (ex ante evalueringer). Forfatteren argumenterer for at modellering av sikkerhetseffekter må inkludere alle dimensjoner av trafiksikkerhet (eksponering, ulykkesrisiko og konsekvens), dekke både tilsiktede effekter og

effekter av atferdstilpasning, og til slutt også være kompatibelt med andre aspekter av relevante trafikksikkerhetsteorier. Basert på disse kriteriene er det presentert et rammeverk for sikkerhetsvurdering av ITS, basert på ni relevante mekanismer som påvirker sikkerheten:

1. *Direkte endring av kjøreoppgaven* som følge av førerstøttesystem i kjøretøyet. Dette kan påvirke førerens oppmerksomhet, mentale arbeidsbelastning og beslutningsvalg. Mekanismen dekker både tilsiktede og utilsiktede effekter.
2. *Direkte påvirkning av vegkantsystemer*, hovedsakelig gjennom informasjon og anbefalinger (f.eks. variable fartsgrenseskilt). Denne mekanismen utøver ingen direkte kontroll over kjøretøyet, og virkningene er mer begrenset enn for førerstøttesystemer.
3. *Indirekte endring av trafikantatferd*. Mekanismen omfatter flere former for atferdstilpasning, blant annet knyttet til faktorer som tillit, oppmerksomhet og plassering i kontroll-loopen (årvakenhet).
4. *Indirekte endring av andres atferd* (non-user behavior), som tar hensyn til at førerstøttesystemer også kan påvirke hvordan andre trafikanter oppfører seg. Eksempelvis kan saktegående selvkjørende minibusser medføre uheldige forbikjøring fra utålmodige bilister.
5. *Endring i interaksjon mellom trafikanter*, for eksempel økt oppmerksomhet som følge av varslingsystemer for hendelser eller glatt vegbane.
6. *Endring i trafikkeksponering*, forårsaket av at trafikanten gjør andre reisevalg. Eksempelvis kan førerstøttesystemer føre til at man gjennomfører reiser på dager med vanskelig føreforhold hvor man ellers ville unngått å reise.
7. *Endring av reisemiddelvalg*. Siden hvert reisemiddel er forbundet med en spesifikk ulykkesrisiko, endres også trafikksikkerheten når man skifter reisemiddel, for eksempel ved innføring av restriksjoner som vegprising, parkeringsstrategier eller reisetidsinformasjon.
8. *Endring av rutevalg*. Ulike deler av vegnettet har ulik ulykkesrisiko, dermed påvirkes også trafikksikkerheten når trafikanten velger en annen reiserute som følge av for eksempel hendelsesvarslings.
9. *Endring i ulykkeskonsekvens* som følge av skadereduserende ITS. Eksempelvis kan eCall forkorter utrykningstiden for helsepersonell ved ulykker.

Forfatteren poengterer at rammeverket dekker trafikantbeslutninger på alle nivå av kjøreprosessen (strategisk, taktisk, operasjonelt), og viser til at det finnes empiriske bevis for at disse ni mekanismene eksisterer (Kulmala, 2010). Rammeverket fremstår konkret og anvendbart for sitt bruk, og håndterer en del av de spesifikke utfordringene knyttet til effektstudier av ITS. Forfatteren

argumenterer også for at det i evalueringer bør være rom for å kombinere empiriske data med ekspertvurderinger, og at selve modellen bør være dynamisk med mulighet for kontinuerlig oppdatering etter hvert som ny kunnskap etableres. I så måte skiller fremgangsmåten seg fra de fleste andre evalueringsmodeller i transportfaget, ved at den involverer en bredere tolkning av evidensbegrepet, i tråd med Sackett et al. (1996). Selv om modellen i utgangspunktet er avgrenset til sikkerhetseffekter, fanger den opp relevante problemstillinger ved ITS som også kan være nyttig som sjekklister ved andre ITS-evalueringer.

4.4.4 RAMMEVERK FOR EVALUERING AV BRUKERAKSEPT

De fleste ITS-evalueringer involverer en eller annen form for vurdering av brukeraksept, som mål på hvordan brukerne eller den potensielle målgruppen vil respondere på ITS-tjenesten som innføres – uten at dette nødvendigvis er forankret i eksplisitt teori. Forskningsfeltet opererer med litt ulike fortolkninger av akseptbegrepet, og omfatter også varierte metoder og indikatorer for måling av teknologiaksept (Vlassenroot, Brookhuis, Marchau, & Witlox, 2010). Det skilles blant annet mellom den generelle sosiale aksepten for et tiltak i befolkningen (social acceptance) og brukeraksept fra de som har faktisk erfaring med tiltaket (user acceptance) (Van Der Laan, Heino, & De Waard, 1997). Schade and Schlag (2003) beskriver en holdningsrelatert aksept som kan måles før innføring av et tiltak og en opplevelselsbasert aksept som kan måles etter at tiltaket er implementert og testet i praksis. De fleste studier synes imidlertid å basere seg på varianter av *The Technology Acceptance Model (TAM)* (F. D. Davis, 1989; Venkatesh & Davis, 2000), med opplevd nytte (usefulness) og brukervennlighet (ease of use) som viktige forklaringsfaktorer på holdning, intensjon og bruk. Over tid er modellene testet og utvidet med flere forklaringsfaktorer, og senere presentert som *Unified theory of acceptance and use of technology* (UTAUT1, UTAUT2 og UTAUT3) (Venkatesh, Morris, Davis & F. D. Davis, 2003; Venkatesh, Thong & Xu, 2012; 2016). En videreutvikling av dette teorigrunnet er presentert i Nordhoff, Happee & van Arem (2019) som forklarer brukeraksept for automatiserte kjøretøy med utgangspunkt i 28 akseptfaktorer fordelt på syv kategorier:²⁵ Eksponering, tjenestetilbudet, sosial påvirkning og motivasjon, opplevd nytte og risiko, sosio-demografiske faktorer, reisevaner og individuelle personlighetstrekk.

Van Der Laan et al. (1997) har utviklet et spørreskjema som måler brukeraksept for ITS-tjenester ved hjelp av ni indikatorer for dimensjonene nytte (usefulness) og tilfredshet (satisfaction), se Tabell 5.

²⁵ Kategoriene er benevnt med begrepene *exposure, domain-specific, symbolic-affective, moral-normative, socio-demographics, travel behavior, personality*.

Tabell 5: Indikatorer for brukeraksept i henhold til Van Der Laan et al. (1997).

Indikatorer for nytte (usefulness)	Indikatorer for tilfredshet (satisfaction)
useful vs useless	pleasant vs unpleasant
bad vs good	nice vs annoying
effective vs superfluous	irritating vs likeable
assisting vs worthless	undesirable vs desirable
raising alertness vs sleep-inducing	

Hver indikator tildeles en verdi på en skala fra -2 (svært negativ) til $+2$ (svært positiv). Dette er blant annet egnet når brukeraksept for ulike ITS-tjenester skal sammenlignes med hverandre. Med utgangspunkt i dette skjemaet kan det beregnes gjennomsnittsverdier for faktorene nytte og tilfredshet som presenteres samlet i et kryssdiagram, med mulighet for sammenligning av verdier mellom ulike befolkningsutvalg eller med andre ITS-tjenester.

4.5 METODEDISKUSJON

Til tross for stadig økende bruk av ITS, finnes det ennå ingen helhetlig og felles akseptert evalueringsmetodikk. ITS-evalueringer møter to sentrale utfordringer: For det første representerer ITS et sett med svært forskjellige virkemidler, fra enkle teknologier til komplekse, sammensatte systemer. Samtidig påvirkes tiltakenes modenhet av at teknologien er i rask utvikling. Disse egenskapene gjør det nødvendig å tilpasse evalueringsdesign og metodebruk til det enkelte tiltaket som skal vurderes.

Samfunnsøkonomiske evalueringer er et viktig verktøy når beslutningstakere skal investere i fremtidens transportløsninger, men det foreligger ingen EFFEKT-modul for beregning av nyttekostnadsverdi for ITS-tiltak i det norske evalueringsrammeverket. I Høie (2013) presenteres et forslag til verktøy for virkningsberegning av et spesifikt utvalg ITS-tiltak, basert på nyttekostnadsmetodikk. Bruk av dette regnearket forutsetter imidlertid at man allerede har etablert empiri om tiltakets nytteeffekter og kostnader, og at den kausale sammenhengen mellom tiltaket og virkning på blant annet ulykker, fartsfordeling og en rekke andre variabler er kjent for den som skal gjennomføre evalueringen. Det finnes derimot en rekke eksempler på at det samme ITS-tiltaket kan utløse vidt forskjellige effekttyper (Belin et al., 2010; Bjerkan, 2017; Nordtømme et al., 2014), og det er vanskelig å etablere allmenngyldige standardverdier for ITS-tiltak. Osland et al. (2007) peker på risikoen for at manglende metodiske overveielser i kombinasjon med komplekse statistiske analyser (manglende transparens) kan medføre entydige og sikre vurderinger på spørsmål som er flertydige og sammensatte – som rangordning av prosjekter i nyttekostnadsanalyser. Odeck og Welde (2010) drøfter en rekke metodiske utfordringer med samfunnsøkonomiske analyser av ITS, men understreker samtidig viktigheten av å evaluere ITS-tiltak innenfor et rammeverk som gjør at

slike tiltak reelt kan vurderes og prioriteres i konkurranse med tradisjonelle tiltak. Forfatterne eksemplifiserer hvordan tilpasning av variabler og betingelser i nyttekostnadsanalysen har gjort det mulig å demonstrere stor samfunnsøkonomisk nytte for innføringen av automatiske bomstasjoner i Oslo.

Det eksisterer alternative evalueringsteknikker som i større grad enn den tradisjonelle nyttekostnadsanalysen behandler ikke-monetære nyttekomponenter (Browne & Ryan, 2011; Jordanger, Malerud, Minken, & Strand, 2007; Stevens, 2004). *Kostnadseffektivitetsanalyser (cost efficiency analysis)* vurderer virkningsgraden av et tiltak i relasjon til investerings- og driftskostnader ved å uttrykke kostnadene i monetære verdier og nyttekomponentene i form av hvor godt de oppfyller målsettingene med tiltaket. *Multikriterieanalyser/ flermålsanalyser (multi criteria analysis)* vurderer alle relevante kvantitative og kvalitative effekter av tiltaket uten å tilordne disse monetære verdier. Evalueringen baseres på et sett kriterier som legges til grunn for vektning av effekter og prioritering av alternativer. Det erkjennes likevel at til tross for at nyttekostnadsanalysen og kostnadseffektivitetsanalysen har en del åpenbare mangler og begrensninger med hensyn til evaluering av ITS-tiltak, kommer man ikke utenom behovet for å vurdere sammenhengen mellom kostnader og nyttevirkinger i prosjekter (Browne & Ryan, 2011; Stevens, 2004). Stevens (2004) konkluderer med at nyttekostnadsanalysen som metode er nyttig og nødvendig, men også utilstrekkelig for ITS-evalueringer. Forfatteren anbefaler at evalueringen utvides med en praktisk tilnærming hvor alle relevante effekter identifiseres og verdsettes med indikatorer som er hensiktsmessig for den enkelte variabel. Indikatorbaserte tilnærminger tar i større grad hensyn til de kvalitative effektene av ITS-tiltaket, men medfører gjerne et mer komplekst beslutningsgrunnlag. Når kunnskap om effekter og sammenhenger etableres over tid, kan imidlertid flere variabler prissettes og behandles innenfor rammene av den kvantitative analysen. Multikriterieanalysen anses å ha fordeler når aktørene har forskjellige interesser og ulike vurderinger av nytteeffektene av et tiltak, og når nytteeffektene har stor grad av usikkerhet eller er vanskelig å verdsette.

Det synes å være bred enighet om at dagens metodegrunnlag for samfunnsøkonomiske evalueringer av ITS ikke er tilstrekkelig, men det er ulik oppfatning av hvorvidt dette kan løses ved tilpassing av eksisterende modeller, om dagens metodegrunnlag må videreutvikles og eventuelt suppleres, eller om det er behov for et helt nytt rammeverk (Bristow et al., 1997; Kolosz & Grant-Muller, 2015; Newman-Askins et al., 2003; Stevens, 2000, 2004). På den ene siden argumenteres det for at man bør tilpasse måling og verdsetting av ITS-tiltakene til nyttekostnadsmetodikken som til tross for sine begrensninger bygger på anerkjente og velprøvde prinsipper fra økonomisk teori (Browne & Ryan, 2011; Stevens, 2004). Samtidig erkjennes det at ITS-tiltak er så komplekse og forskjellige fra

tradisjonelle vegtiltak, at det kan være behov for et helt nytt evalueringskonsept som også inkluderer eksterne faktorer som har betydning for hvilket markedspotensial tiltaket kan oppnå (Newman-Askins et al., 2003; Stevens, 2004).

Når det gjelder *effektevalueringer* av ITS-tiltak, er det tradisjonelle eksperimentet med før- og etterundersøkelser og en kontrafaktisk tolkning av kausalitetsbegrepet nesten enerådende på feltet. Dette kapittelet har vist at det eksperimentelle metodegrunnlaget er basert på en del premisser som sjelden kan oppfylles for ITS-evalueringer, og i praksis er det nødvendig med en rekke praktiske tilpasninger og antakelser for å kompensere for metodiske svakheter. Dette gir seg utslag i evalueringsstudier hvor både datagrunnlag og resultater er beheftet med usikkerhet og det kan være vanskelig å trekke sikre konklusjoner, til tross for at selve evalueringsrammeverket holder en høy vitenskapelig standard (se f.eks. erfaringer fra Gjerstad og Bayer (2012), Høye, Sørensen, et al. (2011) og Thorenfeldt et al. (2011)). Dette kan muligens også gjenspeiles i metaanalyser av ITS-effekter som ofte viser lite konsistente resultater mellom studier (Høye, Elvik, et al., 2011; Nordtømme et al., 2014; Vaa, 2006).

ITS-evalueringer synes å utfordre flere av kvalitetskriteriene knyttet til resultatenes validitet, som drøftet i kapittel 3.4.3. Uttesting og evaluering av tiltak gjennom pilotprosjekter gir dårlig mulighet for representativitet. Deltakelse i ITS-prosjektene er ofte basert på frivillighet og små utvalg, og kravet om *statistisk validitet* kan i liten grad imøtekommes. Videre utfordres den *teoretiske validiteten* av at kunnskap om virkningsmekanismer i liten grad er etablert eller uttrykt i forkant av evalueringene. Selv om nyttekostnadsanalysene representerer et unntak med et klart definert teoretisk rammeverk, betinger dette like fullt at antakelser og forutsetninger er eksplisitt og klart formulert i forkant av analysene. Fravær av et felles etablert teorigrunnlag for ITS-tiltak, gjør det vanskelig å vurdere resultatenes teoretiske gyldighet. ITS-tiltakene representerer komplekse og situasjonsavhengige årsakssammenhenger, med kombinerte effekter fra flere applikasjoner og tjenester i samme system. Dette gjør det krevende å vurdere *intern validitet* ut fra klassiske kriterier som et entydig forhold mellom årsak og virkning og kontrafaktisk kausalitet. Erfaringer viser at det samme ITS-tiltaket kan gi forskjellige effekter i ulike implementeringer, og har dermed egenskaper som i større grad samsvarer med det mekanistiske kausalitetsprinsippet; hvor effekten betraktes som en latent egenskap som kan realiseres under gunstige betingelser og omgivelser. Dette impliserer at det er større grunn til å søke lokal gyldighet i spesifikke implementeringer istedenfor allmenngyldig validitet. Til sist har gjennomgangen vist at resultatene fra effektstudier av ITS sjelden er gyldig over tid, mellom ulike steder eller kontekst. Dette gir lav *ekstern validitet* og begrenset overførbarhet mellom ulike prosjekter.

Utfordringene med å fremskaffe kunnskap om ITS-effekter, kan sees i relasjon til evidensdebatten i kapittel 3.4.5. Transportfaget legger en tydelig avgrenset fortolkning av evidensbegrepet til grunn, hvor kunnskap er ensbetydende med resultater fra det vitenskapelige eksperimentet – til tross for at transportforskning svært sjelden oppfyller metodekravene for randomiserte kontrollerte forsøk. I denne kunnskapsforståelsen går man glipp av viktig erfaringsbasert kunnskap fra eksperter som har førstehånds kjennskap til implementering og utprøving av ITS-tiltak, samt relevante perspektiver fra brukerne av ITS-tjenestene (Alvsvåg, 2009; Dziekan et al., 2013; Johansson et al., 2015; Skulberg, 2008). Kulmala (2010) representerer et unntak med å åpne for en bredere evidensforståelse og mer pragmatisk kunnskapsproduksjon i ex post evalueringer av sikkerhetseffekter av ITS. Til tross for at det har vist seg vanskelig å fremskaffe et relevant og anvendbart kunnskapsgrunnlag, står den tradisjonelle metodestyrt tilnærmingen sterkt, og det foregår lite drøfting rundt valg av alternative metoder for effektevaluering av ITS-tiltak.

4.6 BEHOV FOR VIDERE UTVIKLING AV METODEGRUNNLAGET FOR ITS-EVALUERINGER

Den raske teknologitvillingen skaper utfordringer for fagmiljø og beslutningstakere som skal foreta strategiske valg og investeringer i et fremtidsrettet transportsystem. Det utvikles nye transportløsninger med egenskaper og i et tempo som ikke passer inn i dagens beslutningsmodeller. Siden innføring av ITS-tjenester har kapasitet til å forandre både måten vi opptrer på som trafikanter og samspillet mellom kjøretøy, infrastruktur og omgivelser, endres også basisen for teorier, modeller og verktøy som brukes for å forklare eller estimere relevante variabler som trafikantatferd, trafikkavvikling, trafiksikkerhet og miljøeffekter.

Digitaliseringen stiller store krav til myndigheter og premissgivere som må ta beslutninger og innføre tiltak parallelt med at utviklingen skjer – også når beslutningsgrunnlaget er usikkert. Beslutningstakere etterspør kunnskap om effekter og lønnsomhet, og kunnskap bygges gjennom evalueringer. Det er imidlertid grunnleggende forskjeller mellom den tradisjonelle evidensorienterte kunnskapsbyggingen og evaluering av ITS-tiltak. ITS omfatter et bredt spekter av virkemidler med forskjellige egenskaper, basert på flere teknologier, rettet mot ulike målgrupper og med forventninger om et mangfoldig og komplekst effektbilde (Foss, 2017b; Kulmala, 2010; Mitsakis et al., 2016). Det har vist seg vanskelig å favne alle disse tiltakene med en felles evalueringsmetodikk. I dag vurderes ITS-tiltakene i henhold til et evalueringsrammeverk som er tilpasset tradisjonelle vegtrafikktiltak. Det er en erkjennelse i fagmiljøet at dette metodegrunnlaget er utilstrekkelig for å vurdere virkemidlene vi står overfor i fremtidens intelligente transportsystem, men det foreligger i liten grad omforente forslag om hvordan dette bør håndteres.

Utviklingen i transportsektoren skjer i stadig større grad gjennom utprøving og pilotering av nye teknologibaserte applikasjoner og tjenester. Dette utfordrer dagens evalueringspraksis på flere måter. For det første gir en gradvis innføring uklare sammenligningsgrunnlag i tradisjonelle før- og etterundersøkelser. Samtidig fremstår teknologiutviklingen som en pågående prosess, hvor både tiltaket og omgivelsene gjennomgår kontinuerlig forbedring. Vi må sannsynligvis erkjenne at vi aldri vil lande på en endelig og ideell løsning. Da må evalueringen kunne ta høyde for ulike grader av teknologisk modenhet, og samtidig fremskaffe raske resultater som grunnlag for fortløpende justering og optimalisering av både teknologi og omgivelser. Først når ITS-tiltaket er modent og evaluert i stabile implementeringer kan effekter avdekkes og realistiske nytteverdier estimeres.

Det er også grunn til å reflektere over hvilke forskningsspørsmål som er relevante. Ikke all teknologi eller ITS fungerer slik at transportmyndighetene skal velge ja eller nei til innføring. Mange ITS-løsninger har allerede stor utbredelse, det etableres sosiale plattformer for organisering av transporttjenester, og biler får stadig mer avanserte førerstøttesystemer. I mange tilfeller er det hverken relevant eller gjennomførbart å sammenligne mot et nullalternativ. Kanskje handler det i større grad om hvordan man kan legge til rette for ønsket innovasjon, innføring og anvendelse: Hvilke rammebetingelser må være til stede? Hva slags regulering er nødvendig for å unngå uønskede effekter? Hvilken rolle skal myndighetene ta for å stimulere til ønsket innovasjon og ivareta rettferdig fordeling av godene? Hvilke forretningsmodeller kan sikre en bærekraftig drift over tid? De fleste forskningsspørsmålene lar seg ikke besvare med den tradisjonelle fremgangsmåten, og det er behov for et tverrfaglig samarbeid om kunnskapsbygging som kombinerer flere innfallsvinkler og metoder, og involverer en bredere forståelse av evidensbegrepet.

Sett i forhold til evalueringsteoriene som ble drøftet i kapittel 3, er dagens metodegrunnlag og evalueringspraksis forankret i en summativ evalueringstradisjon, med vekt på effektmålinger og nyttekostnadsvurderinger. Kunnskapsbehovet vil i større grad være i tråd med den formative fremgangsmåten, hvor evalueringen foregår samtidig som tiltaket utvikles og innføres, med hensikt å forbedre tiltaket underveis i implementeringsprosessen. Det fremstår derfor som paradoksalt at evalueringsfagets generelle dreining fra tidligere summative tradisjoner til en mer formativ evalueringspraksis ikke er særlig synlig i transportsektoren eller ITS-faget. Transportområdet synes å støtte seg på en streng fortolkning av evidensbegrepet, hvor forskningsbasert kunnskap fremskaffet gjennom bruk av den (kvasi-) eksperimentelle modellen, er normen for kunnskapsproduksjon. En slik avgrensning av forskningsfeltet, risikerer å undervurdere betydningen av sosiokulturelle faktorer, med det resultat at relevant eksperterfaring og viktig kunnskap går tapt (Alvsvåg, 2009; Johansson et al., 2015; Skulberg, 2008).

Det er identifisert fem grunnleggende behov og anbefalinger knyttet til utvikling og valg av evalueringsmetodikk for ITS-tiltak:

1) Evalueringen bør ha et læringsperspektiv som bidrar til helhetlig kunnskapsbygging

Det er behov for å forstå hvordan ITS-tjenestene fungerer og påvirker både trafikanter, trafikksystemet og samfunnet for øvrig. Evalueringen må gi svar på hvordan, hvorfor, for hvem og i hvilken kontekst tiltaket virker, samt eventuelle utilsiktede effekter som kan inntreffe.

2) Evalueringen bør kunne håndtere tiltak med ulik teknologisk modenhet

Teknologien er i stadig utvikling, og ulike nivå av teknologisk modenhet og livsløpsfaser reiser ulike spørsmål som skal besvares av evalueringen. ITS-tjenesten vil som regel ikke fungere i en perfekt tilstand fra start, og det er behov for en evaluering som gir nyttig kunnskap også i situasjoner hvor tjenesten ikke fungerer optimalt eller stabilt, herunder kunnskap om hvordan teknologien bør forbedres.

3) Evalueringen bør kunne identifisere de kontekstuelle faktorenes betydning

En ITS-tjeneste vil kunne utløse ulike effekter i forskjellige situasjoner. En viktig del av evalueringen vil derfor være å kartlegge samspillet med de kontekstuelle faktorene som bidrar til tiltakets utfall, enten gjennom å forsterke ønskede effekter (suksesskriterier) eller ved å hindre måloppnåelse (barrierer). Dette kan innbefatte forhold knyttet til blant annet rammebetingelser, selve implementeringen, omgivelsene eller ulike målgrupper. Kunnskap om eksterne forhold som påvirker tiltaket har stor verdi i planleggingen av fremtidige implementeringer av tiltaket.

4) Evalueringen bør gi kunnskap om hvorfor et tiltak virker eller ikke

ITS-tjenester kjennetegnes gjerne av høy innovasjonsgrad, mange involverte aktører og komplekse sammenhenger mellom årsak og virkning. Evalueringen bør bidra til å skape forståelse for hvordan tiltaket virker, gjennom å etablere kunnskap om de kausale sammenhengene (i hvor stor grad ulike faktorer bidrar til et resultat).

5) Evalueringen bør kunne identifisere og vurdere alle relevante konsekvenser, også kvalitative

Tradisjonelle veg- og trafikktiltak har gjerne et begrenset og håndterbart antall effekter som er kjent og som i stor grad kan prissettes. ITS-tjenester har derimot potensiale til å utløse helt andre mekanismer som påvirker en rekke sosiale strukturer i omgivelsene. Evalueringen bør derfor ta høyde for å inkludere nye variabler som ikke er direkte målbare, eksempelvis personvernutfordringer, komfortfaktorer og endring i mobilitetsbehov.

5 KAN EN TEORIBASERT TILNÆRMING STYRKE ITS-EVALUERINGEN?

De foregående kapitlene har gjort rede for et bredt og mangfoldig teoretisk grunnlag for evalueringsforskning, men samtidig vist at beslutningstaking i transportsektoren følger en strengt metodestyrt tilnærming som i liten grad er tilpasset ITS-tiltakenes egenskaper og kompleksitet. I dette kapittelet stilles det derfor spørsmål om en dreining mot en teoribasert evalueringsmetodikk kan bidra til å styrke kunnskapsgrunnlaget for innføring av ITS i transportsystemet?

*Terminologien innenfor dette fagfeltet er i høyeste grad uklar og lite konsistent. Ulike begreper benyttes for å beskrive samme fenomen, like begreper benyttes for å beskrive forskjellige forhold. Betegnelsene er både avvikende og overlappende. Dette gjør det krevende å forholde seg til forskningslitteraturen og ikke minst å formidle teorien i en forståelig sammenheng. Det er ikke funnet hensiktsmessig å gjøre rede for alle tilgjengelige begreper underveis i teksten, men heller velge den benevnelsen som synes mest relevant og bruke denne konsekvent. Eksempelvis benyttes begrepet "programteori" gjennomgående i denne teksten, selv om ulike forfattere også bruker benevnelser som *intervention theory*, *intervention logic*, *intervention model*, *logic model*, *program logic*, *theory of change*, *impact path analysis*, *logical framework*, *outcomes map* og *local theory*.*

5.1 FRA METODESTYRT TIL TEORIBASERT EVALUERINGSSTRATEGI

Den innledende litteraturstudien har vist at evaluering er et vidtfavnende fagfelt med en rekke vitenskapsteoretiske tilnærminger som alle har sine styrker og svakheter, og som fremskaffer ulike typer kunnskap. Gjennom tidene har evalueringsfaget beveget seg fra et akademisk utgangspunkt med sterkt målfokus, til et mer formativt og anvendt perspektiv, hvor evalueringer i større grad bidrar til kontinuerlig læring og justering av tiltaket underveis i prosessen. Transportsektoren utnytter i liten grad dette mangfoldet av tilgjengelige evalueringsmodeller. Transportfaglige evalueringer er hovedsakelig rettet inn mot en evidensorientert vurdering av effektivitet og lønnsomhet, og i liten grad påvirket av øvrige perspektiver og diskurser fra evalueringsfaget. Fremgangsmåten er isteden metodestyrt i så stor grad at evalueringsoppskriften er operasjonalisert i regneark hvor man putter inn forhåndsdefinerte data og får ut resultatet i form av en nyttekostnadsverdi.

Den metodestyrt tilnærmingen har mange fordeler: Evalueringen er effektiv, og den kan i prinsippet utføres uten dyptgående kunnskap om tiltaket eller transportsystemet. Den viktigste kvaliteten er likevel at den gir resultater som er direkte sammenlignbare med andre evalueringsstudier som gjennomføres etter samme oppskrift. Dette innebærer at resultatene gir

direkte input til transportmyndighetenes beslutningsprosesser. Samtidig er det en vitenskapelig anerkjent fremgangsmåte som forsvarer resultatets gyldighet gjennom metodisk stringens.

Likevel tyder evalueringspraksis på at den eksperimentelle modellen ikke er tilstrekkelig når det kommer til evaluering av ITS. Effektstudier må ofte suppleres med kvalitative analyser, enten for å fange opp flere relevante problemstillinger eller for å kompensere for at betingelsene og antakelsene som ligger til grunn for metodikken ikke er oppfylt. Dersom det gjøres vesentlige tillempinger og justeringer for å få tiltaket til å passe inn i evalueringsmalen, risikerer man å ende opp med resultater som er usikre og lite konsistente.

Eksperimentelle effektanalyser og nyttekostnadsmodellen gir svar på spørsmål av typen: *Virker tiltaket? Hvor stor er den observerte effekten? Og er effekten lønnsom eller ikke?* Innføring av ITS i transportsystemet reiser imidlertid en rekke nye problemstillinger rundt utforming og valg av løsning, rammebetingelser, forretningsmodeller, suksesskriterier og barrierer. Dette gir grunnlag for andre relevante forskningsspørsmål som også kan være styrende for valg av metodikk.

Den teoribaserte tilnærmingen adresserer flere av utfordringene knyttet til ITS-evalueringer: Den kan håndtere komplekse tiltak og uklare årsakssammenhenger, den inkluderer rammebetingelser og kontekst i analysene, den kan bidra til å forbedre tiltaket underveis i evalueringsprosessen, og den åpner for å kombinere flere metoder og analyseteknikker. I praksis innebærer dette for eksempel at en teoribasert evalueringsmetodikk også kan inkludere nyttekostnadsanalyser dersom tilgangen til data ligger til rette for dette.

Figur 9 viser en sammenstilling av egenskaper ved en *klassisk metodestyrt tilnærming til effektstudier* og en *teoribasert evalueringsmetodikk*. I en overordnet teoribasert tilnærming kan det spesifikke teoribaserte metodeoppsettet (forskningsdesign og analyser) være både et alternativ og et supplement til tradisjonelle eksperimentelle analyser.

Metodestyrt og målorientert evalueringsprosess	Tilnærming	Teoribasert evalueringsprosess med utvikling av tiltakets programteori
Det vitenskapelige eksperimentet: (Kvasi-)eksperimentell og kontrafaktisk design	Forskningsdesign	Kombinerer flere design som besvarer ulike spørsmål. F.eks. eksperimentell og teoribasert (theory of change og/eller realistisk).
Kvantitativ datainnsamling og statistiske analyser	Metoder	Kvalitativ og kvantitativ datainnsamling og analyser
Gir vitenskapelig sterkt effektbevis når metodekrav er oppfylt	Resultat	Forklarer sannsynlige effekter, rammebetingelser, suksesskriterier og barrierer ved innføring

↓

+
eventuelle
supplerende analyser

Figur 9: Sammenstilling av egenskaper ved klassiske effektstudier og en teoribasert evalueringsmetodikk. I praksis vil klassiske ITS-studier ofte suppleres med tilleggsanalyser som kan bidra til å forklare resultatene eller øke studiens validitet når de underliggende metodekravene ikke kan oppfylles.

Det teoribaserte evalueringsperspektivet skiller seg fra den eksperimentelle metodikken på to sentrale punkt; den inkluderer *kontekstuelle faktorer* i evalueringsdesignet og den adresserer de *kausale sammenhengene* mellom innsats og virkning med utgangspunkt i en mekanistisk kausalteori. Dette gir stor forklaringskraft med hensyn til å besvare hva som bidrar til at et tiltak virker eller ikke virker – og i hvilke situasjoner man kan forvente å oppnå effekt (høy teoretisk og intern validitet). Dette er høyst relevant kunnskap for innføring av ITS i transportsystemet.

Med dette grunnlaget synes det hensiktsmessig å prøve ut en teoribasert tilnærming for evaluering av ITS-tiltak i doktorgradsarbeidet. Begrunnelsen for dette valget er nærmere utdypet gjennom dette kapittelet, som gjør rede for den teoribaserte evalueringens innretning, styrker og svakheter. I kapittel 6 og 7 presenteres to case-studier som viser hvordan metodikken kan anvendes i praksis.

5.2 PROGRAMTEORIEN - TILTAKETS TEORETISKE RAMMEVERK

Den teoribaserte evalueringstilnærmingen tar utgangspunkt i en konseptuell modell av sammenhengen mellom innsatsen i tiltaket og de virkninger man forventer å oppstå på kort og lang sikt – altså en beskrivelse av hvordan tiltaket er ment å operere og medføre endringer i en gitt kontekst (Chen, 1990, 2005; Funnell & Rogers, 2011; Rossi et al., 2007; Weiss, 1995). Denne

modellen omtales gjerne som tiltakets *programteori*²⁶. Programteorien representerer *tiltakets underliggende logikk*: Prosjekteierens forventningen om hvordan og hvorfor planlagte virkemidler skal føre til ønsket resultat – altså den rasjonelle grunnen for iverksettelsen av et tiltak.

En teoribasert evalueringsprosess vil innledningsvis etablere tiltakets programteori, basert på eksisterende kunnskap og forventninger om årsakssammenhenger, inkludert de underliggende mekanismene som utløser effekter på kort og lang sikt. Den konseptuelle modellen kan inneha ulik grad av detaljering og kompleksitet, avhengig av tiltakets beskaffenhet og evalueringens formål. Figur 10 viser en skjematisk fremstilling av en enkel lineær programteori. Når tiltakets kompleksitet og evalueringens detaljfokus tiltar, benyttes gjerne mer detaljerte modeller som også kan fremstilles hierarkisk eller på matriseform.



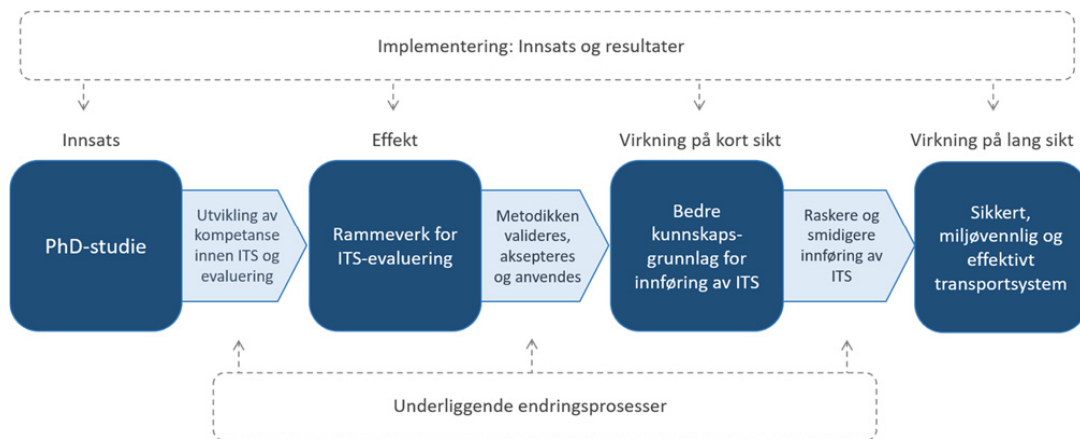
Figur 10: Skjematisk fremstilling av programteorien; den underliggende logikken bak iverksettelsen av et tiltak.

Programteorien rommer to sentrale dimensjoner som begge har betydning for tiltakets endelige utfall (Chen, 2012, 2016; Funnel & Rogers, 2011);

- *Tiltakets implementeringsteori (theory of action)*: En preskriptiv modell av hvilke betingelser (konkrete handlinger og prosesser) som må være oppfylt for at tiltaket skal avstedkomme en ønsket endring. Dette er illustrert med boksene i Figur 10. Implementeringsteorien viser hvordan prosjektet er konstruert med et logisk resonnement for å aktivere endringene i neste punkt.
- *Tiltakets endringsteori (theory of change)*: En deskriptiv modell av de underliggende mekanismene og kausale prosessene som må inntreffe for at endring skal oppnås. Dette er illustrert med pilene i Figur 10. Programteorien kan inneha ulike endringsteorier for ulike faser av tiltaket og for ulike målgrupper i prosjektet. Dette er gjerne tuftet på samfunnsvitenskapelig teori (f.eks. psykologiske, sosiale, økonomiske, politiske modeller).

²⁶ Begrepet programteori har utspring fra samfunnsvitenskapelig evalueringsforskning, hvor tiltaket som evalueres ofte består av sosiale programmer (f.eks. utdanningsreformer, helseforebygging og kriminalomsorg), mens ordet teori henspiller på forståelsen av tiltakets operasjonelle virkemåte.

Programteorien er eksemplifisert i Figur 11, som viser en forenklet skisse av programteorien for dette doktorgradsarbeidet. Figuren tar utgangspunkt i PhD-studien som et virkemiddel (tiltak) for å frembringe et forslag til evalueringsrammeverk for ITS på kort sikt (i løpet av studieperioden). Når dette arbeidet er ferdigstilt, forventes det at resultatene (evalueringsrammeverket) vil bidra til et bedre kunnskapsgrunnlag for beslutninger og implementeringer av ITS – noe som på lengre sikt skal resultere i et forbedret transportsystem. Pilene representerer de underliggende mekanismene som sørger for å realisere ønskede effekter underveis i denne prosessen.



Figur 11: Eksempel på programteori med utgangspunkt i dette doktorgradsarbeidet som et virkemiddel for å nå en langsiktig målsetting om et sikkert, miljøvennlig og effektivt transportsystem. Det er kun de to første boksene (innsats og effekt) som vil bli realisert i løpet av studieperioden.

Når programteorien er etablert, danner den utgangspunktet for etterfølgende empirisk testing av de antakelser og forventninger som ligger til grunn for modellen. Det teoribaserte rammeverket gir således en struktur for evalueringsstudien, ved å identifisere kausale relasjoner og synliggjøre hull i kunnskapsgrunnlaget. Rammeverket gir imidlertid ikke føringer for spesifikt valg av analyseteknikk, men forutsetter at datainnsamling og analyser defineres på bakgrunn av hvilke forskningsspørsmål som ønskes besvart og praktisk mulighet for datainnsamling. Den teoribaserte evalueringen kan i praksis kombineres med de fleste analyser og evalueringsmetoder, og legger dermed til rette for fleksibilitet i metodevalg og datainnsamling. I neste delkapittel gjøres det rede for to spesifikke teoribaserte fremgangsmåter, som kan benyttes alene, i kombinasjon med hverandre eller suppleres med andre evalueringsdesign.

5.3 TEORIBASERTE EVALUERINGSDESIGN

Selv om grunnkonseptet med bruk av programteori og kausale årsakskjeder er felles for all teoribasert evaluering, finnes det flere varianter av evalueringsdesign innenfor fagfeltet. Spesifikke metoder tilbyr blant annet bestemte strukturer for oppbygging av programteori og organisering av analyser. Det konkrete metodevalget foretas på bakgrunn av hva som er evalueringens hovedformål (hvilke spørsmål ønsker man å besvare?), i motsetning til den eksperimentelle tilnærmingen hvor metodevalget er styrt av vitenskapelig stringens. Eksempelvis har *theory of change*-designet stort fokus på implementeringen av tiltaket, med systematiske analyser av de lovmessige sammenhengene i årsaksvirkningskjeder og kontekst, mens det *realistiske* evalueringsdesignet i større grad vektlegger de bakenforliggende (psykologiske) aspektene som utløser individuelle atferdsendringer som en konsekvens av tiltaket (Blamey & Mackenzie, 2007; Hills & Junge, 2010; Pawson, 2013; Pawson & Tilley, 1997). Disse to nevnte metodene for teoribasert evaluering representerer sentrale fremgangsmåter som er nærmere beskrevet i påfølgende delkapittel.

5.3.1 THEORY OF CHANGE

Theory of change-evalueringer tar utgangspunkt i en programteori som er strukturert i form av innsats, reaksjon, effekt og virkning som vist i Figur 10, med utfyllende beskrivelser og forklaringer av sammenhengen (de kausale relasjonene) mellom boksene (Funnel & Rogers, 2011; Government of Canada, 2012; Hills & Junge, 2010; Sverdrup, 2002; Weiss, 1995). Begrepsbruken innenfor evalueringslitteraturen er imidlertid lite konsis, og begrepet *theory of change* brukes i praksis både som uttrykk for en spesifikk metodikk, en samlebetegnelse for variasjoner av fremgangsmåter og som betegnelse for hele eller deler av programteorien.

Evalueringer innenfor denne tradisjonen etablerer programteorien i en såkalt top-down-tilnærming, hvor man starter bakfra med målbildet og definerer hvilke virkninger man forventer å oppnå på lang sikt. Derfra jobber man seg stegvis fremover ved å identifisere hvilke betingelser som må være på plass for å oppnå resultater/effekter på de ulike nivåene. Prosessen involverer ideelt sett et bredt utvalg representanter fra prosjekteier og interessegrupper, og det er en målsetting at det teoretiske rammeverket er basert på oppnådd konsensus blant deltakerne. Når programteorien er etablert, vurderes denne med hensyn til plausibilitet (i hvilken grad modellen fremstår som logisk og rimelig), gjennomførbarhet og testbarhet (i hvilken grad det er mulig å måle/vurdere underliggende faktorer og effekter) (Blamey & Mackenzie, 2007; Bredgard, Salado-Rasmussen, & Sieling-Monas, 2016; Kubisch, 1995). Den som gjennomfører evalueringen benytter så programteorien som utgangspunkt for valg av evalueringsdesign og analyser for testing av programteoriens ulike

bestanddel og kausale sammenhenger. Tilnærmingen legger ingen begrensninger på videre metodevalg, og det anbefales å kombinere både kvantitative og kvalitative analyser.

Tiltakets kausalitet betraktes som en generativ mekanisme hvor programteoriens årsakvirkningskjeder testes med hensikt å finne de (kombinasjoner av) faktorer som har størst forklaringskraft. Kausaliteten bestemmes mekanistisk, på bakgrunn av at i) programteorien fremstår som plausibel og konsistent, ii) tiltakets innsats og aktiviteter ble implementert i henhold til plan, iii) underliggende antagelser ble bekreftet, effekter og virkninger ble observert, og til sist; iv) kontekst og eksterne faktorer ble identifisert og vurdert (Government of Canada, 2012). Blamey and Mackenzie (2007) peker på at de teoribaserte evalueringstudiene også i stor grad benytter eksperimentelle tester i enkelte analyser av kausalitet internt i modellen.

I kapittel 5.1 ble programteorien beskrevet som en kombinasjon av tiltakets *endringsteori* og *implementeringsteori*. Evalueringer med utgangspunkt i *theory of change* har gjerne stort fokus på observerte effekter og beskrivelse av handlinger og prosesser som betinger resultater i de ulike fasene. Dette gir grunnlag for en helhetlig og overordnet forståelse for hvordan tiltaket fungerer i sine omgivelser. Resultatet av evalueringen er ofte en konkret og godt dokumentert *implementeringsteori* som forklarer hvordan kombinasjonen av ulike virkemidler og ressurser bidrar til tiltakets endelige virkning (Blamey & Mackenzie, 2007; Hills & Junge, 2010). Dette gir grunnlag for bred strategisk læring med nytte for fremtidig anvendelse av tiltaket.

Theory of change anses som et spesielt egnet verktøy i evalueringer hvor det fra før foreligger begrenset empiri og hvor tiltakene har høy innovasjonsgrad og kompleksitet (involverer mange aktører, består av pakker med flere virkemidler eller adresserer flere målsettinger).

Fremgangsmåten gir svar på hvorvidt og hvordan et planlagt tiltak vil utløse en ønsket endring, kartlegger eksterne faktorer av betydning og sørger for involvering av relevante interessegrupper i prosessen (Blamey & Mackenzie, 2007).

5.3.2 REALISTISK EVALUERING

Den realistiske evalueringen utvikler sin programteori i form av såkalte CMO-konfigurasjoner (context-mechanism-outcome) som definerer virkningen av et tiltak som resultat av underliggende mekanismer og konteksten tiltaket opererer i (Pawson, 2013; Pawson & Tilley, 1997):

Outcome = Mechanism + Context

Til grunn for denne fremgangsmåten ligger en forståelse av at det samme tiltaket vil kunne utløse ulike effekter i forskjellige situasjoner og hos enkelte deltakere og målgrupper, avhengig av tiltakets

egenskaper og den aktuelle konteksten. De underliggende mekanismene beskriver hva det er ved tiltaket som utløser en reaksjon eller endring i omgivelsene, ofte i form av bakenforliggende psykologiske eller sosiale forklaringsfaktorer. Tiltakets kontekst beskriver hvilke eksterne faktorer som bidrar til å utløse eller hindre en effekt. Det fundamentale spørsmålet som søkes besvart i realistiske evalueringer er: *Hva virker, for hvem, i hvilken kontekst og på hvilken måte?*

Den realistiske fremgangsmåten definerer en rekke CMO-konfigurasjoner med basis i tilgjengelig kunnskap og empiri om tiltaket. Dette skjer gjerne ved hjelp av intervju med prosjekteier og interessegrupper, og ved gjennomgang av forskningslitteratur og relevante dokumenter. På dette viset skiller realistisk evaluering seg fra *theory of change*-tilnærmingen beskrevet i kapittel 5.3.1, fordi målet er å identifisere de mest lovende teoretiske sammenhengene, uten at dette nødvendigvis skal skape konsensus blant involverte aktører. CMO-konfigurasjonene utgjør hypoteser om hva som forventes å påvirke bestemte målgrupper i gitte omgivelser, og i neste steg velges de mest relevante hypotesene som utgangspunkt for empirisk testing. Det realistiske evalueringsperspektivet er metodenøytralt, men vektlegger vitenskapelighet i metodevalget. I praksis anbefales det å kombinere både kvantitative og kvalitative analyser i evalueringen. Blamey and Mackenzie (2007) har oppsummert den realistiske fremgangsmåten i følgende fire punkter;

Steg 1: Forstå tiltakets natur og konteksten det opererer i

Steg 2: Utvikle en rekke mini-teorier (CMO-konfigurasjoner)

Steg 3: Undersøke tiltakets virkninger for ulike målgrupper og omstendigheter

Steg 4: Justere CMO-konfigurasjonene basert på resultatene i steg 3, og velge ut kombinasjonene som best forklarer *hva som virker, for hvem, og i hvilken kontekst*

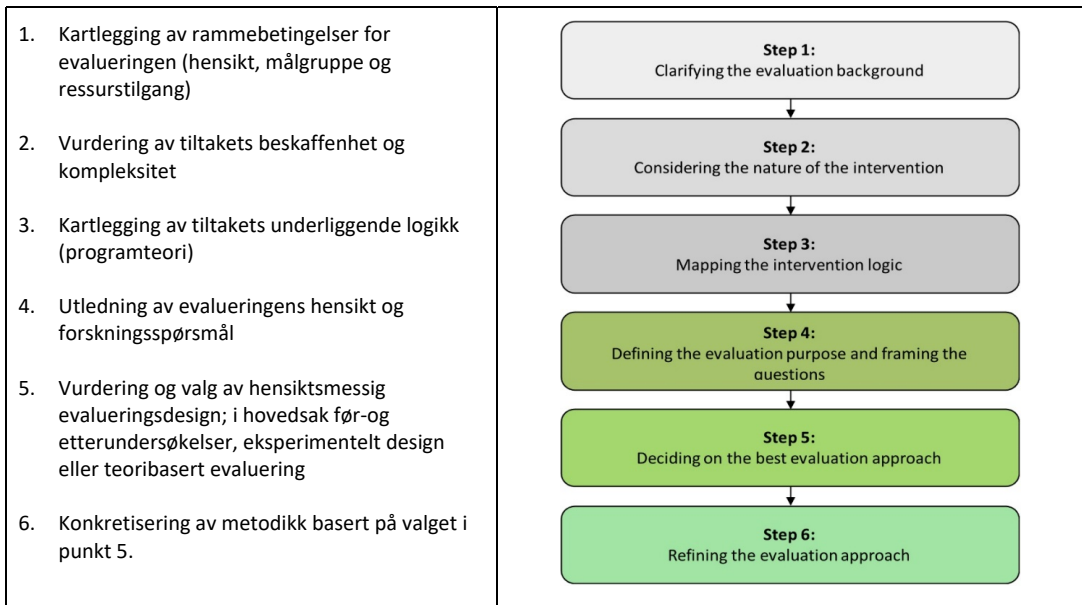
Det realistiske evalueringsperspektivet presenteres som et vitenskapsteoretisk alternativ til den eksperimentelle evalueringstilnærmingen (Bredgard et al., 2016; Pawson, 2002a, 2002b; Pawson & Tilley, 1997). Formålet med realistiske evalueringer er å finne kombinasjoner av målgruppe, innsats og kontekst som gir ønskede resultater, og ikke bevise en generell allmenngyldig effekt av tiltaket. Dette henger sammen med realistenes syn på kausalitet som er forankret i mekanistisk kausalteori og ikke kontrafaktisk gyldighet. I praksis betyr dette at evalueringen tar utgangspunkt i konkrete tilfeller hvor tiltaket har en åpenbar effekt, og anser sammenhengen mellom innsats og virkemidler som gyldig i den spesifikke konteksten (lokal kausalitet). Med dette perspektivet er overførbarheten mellom ulike implementeringer av tiltaket begrenset (lav ekstern validitet), men til gjengjeld muliggjør evalueringen en dypere forståelse av de underliggende mekanismene som trigger atferdsendring og forandring i enkelte situasjoner (høy intern validitet).

Dersom dette relateres til beskrivelsen av programteorien i kapittel 5.1, vil den realistiske evalueringen typisk legge vekt på tiltakets *endringsteori*, med analyser av psykologiske og motivasjonsrelaterte variabler. Hills and Junge (2010) illustrerer dette ved å vise til at evalueringer med utgangspunkt i *theory of change*, gjerne kartlegger hva som skjer i de ulike boksene i programteorien ("the change that happens"), mens *realistiske evalueringer* i større grad undersøker hva som skjer i rommet mellom boksene ("why and how change happens"). Mens den første tilnærmingen har sin styrke i overordnede evalueringer av komplekse tiltak på makronivå, passer den realistiske tilnærmingen bedre for tiltak som er tydelig avgrenset og med et definert influensområde på mikronivå (Blamey & Mackenzie, 2007). I et anvendt perspektiv, vil den realistiske evalueringen bidra til nyttig kunnskap om hvordan virkemidlene i et tiltak kan skreddersys for å oppnå best mulig effekt i ulike situasjoner. Det er imidlertid fullt mulig å kombinere de to sentrale teoribaserte metodene for å oppnå en mer komplett forståelse av tiltakets fullstendige virkemåte og potensiale.

5.4 ANVENDELSE AV TEORIBASERT EVALUERING I TRANSPORTFAGET

Anvendelse av teoribasert evalueringstilnærming har i den siste tiden blitt svært utbredt innenfor samfunnsvitenskapelige fagdisipliner (Funnel & Rogers, 2011; Karlsen & Jentoft, 2013; Pawson, 2013), men det finnes knapt eksempler på eksplisitt bruk av programteori som metode innenfor transportforskning. Man kan imidlertid se på teoribaserte evalueringer som en strukturert og systematisk evaluering av årsaksvirkningsskjeder, og dermed anta at tankegodset bør ligge implisitt i de fleste effektevalueringer. I mange tilfeller er likevel hensikten med iverksettelsen av tiltaket i liten grad uttrykt, med vage forventninger og uklare målsettinger. Dette gjelder kanskje spesielt innenfor ITS-området, hvor mange tiltak iverksettes med den teknologiske løsningen som utgangspunkt, og uten at denne nødvendigvis adresserer en definert samfunnsmessig utfordring (såkalt *technology push*).

Det finnes et par unntak innenfor transportfaget. I en publikasjon på oppdrag fra det britiske Transportdepartementet (UK Department for Transport) presenteres en slags veiledning for effektevalueringer som benytter en teoribasert tilnærming i utviklingen av evalueringsdesign (Hills & Junge, 2010). Tilnærmingen gir fleksibilitet i videre valg av forskningsmetode og analyseteknikk, gjennom en stegvis og systematisk prosess, som illustrert i Figur 12.



Figur 12: Generell tilnærming for effektevalueringer i henhold til Hills and Junge (2010).

Veilederen peker på den teoribaserte tilnærmingens styrke i komplekse evalueringer og/eller når hensikten med evalueringen er å etablere kunnskap om sammenhengen mellom tiltak, kontekst og nytteeffekter. Stevens (2016) har senere foreslått en tilnærming for utvikling av evalueringsmetodikk for feltforsøk med autonome kjøretøy, som bygger på nevnte anbefalinger fra UK Department for Transport. Her legges det blant annet vekt på at analysemetodene må velges på grunnlag av tiltakets programteori (intervention logic).

Belin et al. (2010) har også gått inn i bakenforliggende programteori i en studie som sammenligner innføringen av ITS-tiltaket automatisk farts kontroll (ATK) i Sverige og Australia. Forfatterne viser at selv om teknologien og målsettingen er lik i begge landene, er myndighetenes rasjonale for innføringen (programteorien) forskjellig, og dette påvirker både utformingen og implementeringen av tiltaket. I Australia er tiltaket basert på en underliggende ide om at overskridelse av fartsgrensen er en overveid og forsettlig handling, begrunnet i at trafikantene vil forsøke å kjøre så fort som mulig, så lenge kostnadene ikke overskrider nytteverdien. Dette har ført til at implementeringen av tiltaket har hatt et sterkt kontrollfokus, hvor man har lagt vekt på å gripe flest mulig og dermed øke "kostnaden" i form av opplevd deteksjonsrisiko. I Sverige har man derimot lagt til grunn at trafikksikkerhet er en viktig prioritet for trafikantene, og at fartsovertredelser trolig skyldes manglende risikoforståelse av trafikksystemet og feiltolkning av den sosiale normen for fartsvalg. Iverksettelsen av tiltaket har dermed hatt stort informasjonsfokus og lagt vekt på frivillig atferdsjustering hos trafikantene. Selv om begge casene har vist gode effekter, har programteorien

medført vesentlige forskjeller i implementeringen: I Australia benyttes et skjult, mobilt (flyttbart) system, kombinert med holdningsskapende (emosjonell) informasjon. I Sverige brukes derimot et åpenlyst system som er godt varslet med skilting, og som er fastmontert på spesielt risikofylte steder og strekninger. Kontakten med det svenske publikummet har hovedsakelig vært av informerende og forklarende karakter. Videre påvirkes utformingen av tiltaket av at det i Australia er bileier som er ansvarlig (og må stå til rette) for eventuelle fartsøvertredelser, mens det i Sverige er sjåføren som holdes ansvarlig. Forfatterne konkluderer med at selv om det tilsynelatende er det samme ITS-tiltaket som er innført i begge landene, er det så vesentlige forskjeller i tiltakets programteori, rammebetingelser og kontekst, at det neppe kan klassifiseres som samme tiltak likevel. Det vises videre til Ropohl (1999) som definerer en teknologi som et komplekst sosioteknisk system, hvor menneskets atferd, sosiale organisering, lovgivning og teknologien samhandler. I dette bildet kan ikke fotoboksen (ITS-tiltaket) vurderes som en isolert teknologi, men som et resultat av den helhetlige konteksten den opptrer i.

5.5 METODISKE VURDERINGER

Forskningsspørsmål og forklaringsevne

Den teoribaserte evalueringstilnærmingen struktureres rundt oppbygging og testing av tiltakets programteori. Programteorien er et dynamisk rammeverk som justeres og oppdateres fortløpende, etter hvert som sammenhengene i modellen bekrefte eller avkreftes. Resultatet av evalueringsstudien er dermed en mer sofistikert programteori som produkt av økt forståelse og kunnskap om tiltaket som studeres. Den teoribaserte tilnærmingen har således stor forklaringsevne, i motsetning til eksperimentelle evalueringer som kun besvarer hvorvidt tiltaket virker, basert på statistisk validitet (hvorvidt den registrerte effekten er ekte og ikke et utslag av tilfeldige variasjoner). Evalueringer tuftet på en teoribasert tilnærming vil derimot søke å forklare *hvorfor et tiltak virker eller ikke virker, for hvem det virker og i hvilke situasjoner man kan forvente en effekt*. Og hvis tiltaket ikke virker: *Hva gikk galt?* Det kan være mange årsaker til at man ikke er i stand til å påvise en målbar effekt, også i tilfeller hvor tiltaket egentlig har et stort potensial. Det kan for eksempel ha oppstått tekniske problemer, det kan ha vært forhold ved implementeringen av tiltaket som sviktet, eller utfallet kan ha blitt påvirket av eksterne forhold. Kartlegging av slike underliggende mekanismer gir grunnlag for produktiv læring og anvendbar kunnskap om både tiltaket og implementeringsprosessen.

Programteorien etableres med utgangspunkt i prosjekteiers forventninger om effekter, men kan også involvere andre interessegruppers forventninger og antakelser, og det forutsettes at effektbildet inkluderer både ønskede og utilsiktede virkninger. For komplekse tiltak kan berørte parter ha ulike

og motstridende ønsker og oppfatninger både om tiltakets antatte effekter og det planlagte målbildet. I slike tilfeller kan den teoribaserte tilnærmingen bidra til struktur og god kommunikasjon i evalueringsprosjektet (Kaplan & Garrett, 2005; Riché, 2012; van der Knaap, 2004). Utforming av programteorien vil synliggjøre avstand i verdensbildet mellom de involverte aktørene, og til en viss grad også gi grunnlag for å oppnå felles forståelse basert på eksisterende kunnskap. Visualisering av den teoretiske modellen gir videre en felles referanseramme for dialogen i prosjektet og kan også være et nyttig instrument i kommunikasjon mot omverden.

Kontekst og kausale sammenhenger

Den teoribaserte tilnærmingen avviker fra det eksperimentelle designet på to viktige områder; *inkludering av kontekst og håndtering av kausale sammenhenger*. Teoribaserte evalueringer forventer ikke universell validitet, men erkjenner derimot at det samme tiltaket vil kunne gi ulike effekter i forskjellige situasjoner (Government of Canada, 2012; Pawson, 2002b; Sager & Andereggen, 2012). En viktig del av evalueringen vil derfor være å kartlegge de kontekstuelle faktorene som bidrar til tiltakets utfall, enten gjennom å forsterke effektene (suksesskriterier) eller ved å hindre måloppnåelse (barrierer). Kunnskap om eksterne forhold som påvirker tiltaket har gjerne stor verdi for overførbarhet til andre eller fremtidige implementeringer av tiltaket. I det eksperimentelle designet anses derimot kontekstuelle faktorer som en metodisk svakhet som man forsøker å utelukke fra analysene (Campbell & Stanley, 1966).

Behandlingen av kausale sammenhenger omtales også med begrepet *attribusjon* – i hvilken grad evalueringen er i stand til å bevise forbindelsen mellom innsats og effekt. Tradisjonelle eksperimentelle studier baserer kausalitet på *suksessiv kausalteori*, altså at det er en automatikk i sammenhengen mellom årsak og virkning, og at virkningen ikke kan inntreffe uten at årsaken er til stede (Tuft, 2013). I dette landskapet bestemmes kausalitet på bakgrunn av *kontrafaktiske analyser* som innebærer en sammenligning mot hvordan situasjonen ville vært dersom tiltaket ikke ble iverksatt (nullalternativet). Bhaskar (2008) argumenterer imidlertid for at kunnskap ikke utelukkende utvikles empirisk (basert på erfaring med lovmessige sammenhenger), men også gjennom faktiske hendelser og underliggende mekanismer (tendenser) som kan observeres i den virkelige (realistiske) verden. Den teoribaserte evalueringen har en *generativ mekanistisk tilnærming* til kausalitet i tråd med den realismeorienterte forståelsen av kausalitetsbegrepet (Bhaskar, 2008). Dette innebærer at kausalitet betraktes som en iboende (latent) egenskap som kan aktiveres under gitte omstendigheter – det vil si at innsatsen/årsaken øker sannsynligheten for at en virkning inntreffer, men at dette også avhenger av andre forhold (som rammebetingelser og kontekst). Kausalitet bestemmes derfor gjennom studier av årsaksvirkningsskjeder som grunnlag for å

vurdere i hvor stor grad ulike faktorer bidrar til det observerte utfallet. I den mekanistiske forståelsen av kausalitetsbegrepet legges det vekt på at effektene som observeres er *transfaktuelle* (langvarige, ikke forbigående) istedenfor kontrafaktiske (Bhaskar, 2008).

Evidensbegrepet

Den teoribaserte tilnærmingen legger til grunn en forståelse av evidensbegrepet som inkluderer både forskningsbasert kunnskap (randomiserte kontrollerte forsøk), erfaringsbasert kunnskap (ekspertise) og brukervurderinger, i tråd med Sackett et al. (1996). Dette gir en helhetlig og solid kunnskapsproduksjon, som ikke bare konsentrerer seg om hvorvidt tiltaket har en allmenngyldig effekt, men også inkluderer erfaringer knyttet til hvilke situasjoner og rammebetingelser som vil bidra til ønsket effekt (Sanderson, 2002; Urban et al., 2014). Dette innebærer også at tilnærmingen er metodeneøytral, og legger til rette for å velge den mest hensiktsmessige analysemetoden basert på problemstilling, forskningsspørsmål, teknologisk modenhet og hvor man befinner seg i innovasjonsløpet.

Tolkningen av evidensbegrepet og hva som regnes som gyldig kunnskap, skiller seg distinkt fra den tradisjonelle forskningskulturen i transportfaget, som har en RCT-orientert evidensforståelse og dermed også en metodestyrt tilnærming til evalueringer og investeringsbeslutninger. Kritikken mot et ensidig fokus på slike eksperimentelle studier er blant annet knyttet til at man forhaster forsøk på å dokumentere effektivitet, mens tiltaket fortsatt er ustabil og lite modent, og dermed undervurderer tiltakets potensielle kapasitet i et lengre tidsperspektiv (Urban et al., 2014). Videre pekes det på at et strengt metoderegime gjør at man går glipp av nyttig og relevant kunnskap ervervet fra praktisk erfaring og utprøving (f.eks. ekspertvurderinger) (Alvsvåg, 2009; Johansson et al., 2015; Sackett et al., 1996; Skulberg, 2008). De to ulike fortolkningene av evidensbegrepet produserer forskjellig type kunnskap om tiltaket, og den teoribaserte evalueringen kan med sin metodiske fleksibilitet favne det beste fra begge tradisjonene.

Evalueringens rolle og anvendbarhet

Chen (2005) argumenterer for at den teoribaserte evalueringen med sin programteori ikke bare er beskrivende og forklarende, men også normativ og handlingsorientert i den forstand at den identifiserer hva som må til for å oppnå ønskede målsettinger. Dette berører et relevant spørsmål i evalueringsforskningen: *Når skal evalueringen foregå i tiltakets livsløpszyklus?* Tidspunktet for gjennomføring av evalueringer kan grovt sett deles inn i en idé- og konseptutviklingsfase (ex ante evaluering), en gjennomføringsfase (underveisevaluering) og en driftsfase (sluttevaluering og ex post evaluering) (Samset & Christensen, 2013). Den teoribaserte tilnærmingen med bruk av

programteori er i motsetning til en del tradisjonelle evalueringsmetoder ikke begrenset av tiltakets modenhet, og kan således anvendes gjennom hele prosjektets levetid, allerede fra konseptutviklingsstadiet (Sanderson, 2002).

Evalueringsforskningen trekker dertil frem den teoribaserte evalueringens styrke som virkemiddel for formativ kunnskapsproduksjon, det vil si dens mulighet til kontinuerlig veiledning og forbedring av tiltaket i løpet av prosjektperioden (Karlsen & Jentoft, 2013). Bredgard et al. (2016) peker på at en teoribasert tilnærming er hensiktsmessig både i *planleggingen av tiltak* (visualisering av formål og innsats), som *styringsredskap* (prioritere innsats og justere retning) og som *evalueringssredskap* (danne grunnlag for hypoteser og forsknings spørsmål). Ved å iverksette evaluering i en tidlig fase, kan resultatene fra forskningen gi nyttige innspill til beslutninger om veivalg og justeringer underveis i prosessen. Karlsen and Jentoft (2013, s. 164-165) uttaler sågar om programteorien at *"den har imidlertid kanskje det største proaktive potensialet av alle evalueringsmetoder fordi evaluators analyse kan pløyes tilbake til prosjekteier før tiltak blir satt i verk"*.

Tilnærmingens formative kapasitet leder over til spørsmålet om hva som skal være evalueringens (og evaluatorens) rolle i forskningsprosjektet. Synspunktene på dette spenner tilsynelatende fra det ene ytterpunktet hvor evaluering ses på som en *uavhengig og objektiv kontrollaktivitet* til den andre siden av skalaen hvor evaluering anses som et *verktøy for utvikling og implementering av tiltaket*. Dette temaet faller inn i en grunnleggende og til dels filosofisk akademisk debatt som berører en rekke problemstillinger blant annet knyttet til objektivitet, etikk, politikk og forskningens formål. Den teoribaserte tilnærmingen vil kunne møte kritikk med hensyn til objektivitet, både fordi den som gjennomfører evalueringen nødvendigvis vil være nært knyttet til prosjekteier i arbeidet med etableringen av programteori og fordi evalueringen vil kunne påvirke utviklingen av tiltaket underveis (Bredgard et al., 2016; Fitzpatrick et al., 2012; O'Sullivan, 2012). Ytterligere argumentasjon knyttet til debatten om objektivitet kan blant annet finnes i litteratur om følgeforskning (Baklien & Skatvedt, 2013; Furre & Horingmo, 2013; Sverdrup, 2013).

5.6 STYRKER OG SVAKHETER I DEN TEORIBASERTE TILNÆRMINGEN

Det teoribaserte evalueringsperspektivet omfatter både effekt- og implementeringsanalyser, og kombinerer kvantitative og kvalitative evalueringsmetoder. På denne måten anses gjerne teoribaserte evalueringer som et bindeledd mellom tradisjonelle målorienterte evalueringer og kvalitative prosessevalueringer (evalueringsteoriene er nærmere beskrevet i kapittel 2.3) (Bredgard et al., 2016; Pawson & Tilley, 1997). Sammenhengen mellom de nevnte evalueringsteoriene er vist i Tabell 6, gjengitt og (fritt) oversatt fra Bredgard et al. (2016).

Tabell 6: Teoribasert evaluering som brobygger mellom effekt- og prosessevalueringer. Oversatt fra (Bredgard et al., 2016).

	Eksperimentell effektevaluering	Teoribasert evaluering	Prosessevaluering
Forskningsspørsmål	Virker det?	Hva virker for hvem, hvorfor, og under hvilke omstendigheter?	Hvordan oppleves og oppfattes virkningene?
Fokus	Effekter	Effekt og prosess	Prosesser
Metode	Kvantitative metoder	Metodenøytralt; anvender både kvantitative og kvalitative metoder	Kvalitative metoder
Kausalteori	Kontrafaktisk kausalteori	Mekanistisk kausalteori	Logisk teori (raisonnement)
Ideal for kunnskapsdannelse	Kontekstuavhengig og global viten	Kontekstavhengig, men overførbar viten	Kontekstavhengig og lokal viten
Beslutningsprosess	Antar at beslutningstakere foretar rasjonelle valg av de mest effektive virkemidler for å oppnå et mål	Informerer praktikere, deltakere og beslutningstakere, samt utvikler deres programteorier	Vektlegger lydhørhet overfor berørte interessenter, og informerer beslutningstakere om hvordan tiltaket oppleves og anvendes

Den teoribaserte tilnærmingens egnethet i evalueringer diskuteres som regel i relasjon til den tradisjonelle eksperimentelle metoden med randomiserte kontrollerte forsøk (RCT), fordi denne i stor grad representerer normen for all evidensorientert metodikk. Forskningslitteraturen er forholdsvis entydig og konsistent når det gjelder de ulike evalueringsmodellenes foretreffelighet og mangler, og den teoribaserte tilnærmingen synes å ha sin styrke i følgende egenskaper (Blamey & Mackenzie, 2007; Bredgard et al., 2016; Chen, 2012; Funnell & Rogers, 2011; Government of Canada, 2012; Hills & Junge, 2010; Pawson & Tilley, 1997; Weiss, 1995):

- *Den bidrar til kunnskap om hvorfor et tiltak virker eller ikke.* Dette gjør den spesielt egnet for evalueringer med et læringsperspektiv, når formålet er å skape kunnskap som grunnlag for strategiske valg og fremtidige implementeringer.
- *Den håndterer komplekse tiltak og kausale årsaksvirkningskjeder.* Dette gjør den spesielt egnet for tiltak med høy innovasjonsgrad, med mange involverte aktører og virkemidler, samt når det foreligger begrenset empiri om tiltaket fra før. Videre er det en styrke at tilnærmingen inkluderer både planlagte og utilsiktede effekter i vurderingen, og at den tar høyde for potensielle virkninger av tiltaket i et lengre tidsperspektiv.

- *Den inkluderer kontekstuelle faktorer.* Dette gjør den spesielt egnet for tiltak hvor man antar at utenforliggende forhold spiller en viktig rolle for utfallet av tiltaket, eller hvor virkningen som oppnås avhenger av egenskaper ved målgruppen eller selve implementeringen.
- *Den er anvendelig.* Den teoribaserte evalueringen kan som regel benyttes i situasjoner hvor betingelsene for tradisjonelle metoder ikke er oppfylt. Dette bidrar til at man kan fremskaffe nyttig kunnskap også i situasjoner hvor muligheten til omfattende datainnsamling er begrenset og der det rett og slett ikke er praktisk mulig å gjennomføre eksperimentelle analyser.
- *Den har formativ kapasitet.* Den teoribaserte evalueringen kan utføres i hele tiltakets livssyklus. Dermed kan den gi løpende innspill og tilbakemeldinger som grunnlag for å foreta korrigeringer og velinformerte beslutninger allerede før tiltaket implementeres. Dette bidrar til fokusert innsats og effektiv bruk av prosjektpressurser.
- *Den bidrar til forventningsavklaring og bevisstgjøring hos involverte aktører.* Utviklingen av programteorien legger til rette for deltakelse fra interessenter og prosjekteier i evalueringsprosessen, og sikrer at involverte aktører utvikler en felles forståelse både for tiltakets og evalueringens formål og avgrensning. Både prosessen og den etablerte programteorien blir et kommunikativt instrument i prosjektet og mot omverdenen.

Den teoretiske evalueringen anses derimot å være mindre egnet til å produsere kvantitative bevis med samme vitenskapelige styrke som eksperimentelle modeller (Blamey & Mackenzie, 2007; Hills & Junge, 2010; Weiss, 1998). Manglende fokus på RCT-basert evidens innebærer at teoribaserte evalueringresultater har større nytte i et læringsperspektiv enn som grunnlag for kontroll og dokumentasjon av effektivitet. Dette kompenseres imidlertid av at den teoribaserte tilnærmingen åpner for en fleksibilitet i metodevalget, som gjør at den eksperimentelle fremgangsmåten fortsatt kan inkluderes i de deler av evalueringen hvor dette finnes hensiktsmessig. Derimot vil den tette kontakten mellom evaluator og prosjekteier, blant annet gjennom utviklingen av programteori og feedback av evalueringresultater underveis i prosessen, kunne sette spørsmålsteget ved evalueringens objektivitet. Samarbeidet mellom oppdragsgiver og den som gjennomfører evalueringen kan dermed brukes som argument mot å benytte evalueringresultatene i et kontrollperspektiv. Ytterligere kritikk mot tilnærmingen peker gjerne på utfordringen med å tilrettelegge for deltakelse og oppnå teoretisk konsensus i evalueringer med mange involverte aktører. Kritikken gjelder spesifikt for *theory of change-metodikken*, men også for realistiske evalueringer kan etablering av CMO-konfigurasjoner være en ressurskrevende øvelse når tiltakene er komplekse med flere virkemidler og målgrupper involvert.

Med styrker og svakheter tatt i betraktning, synes den teoribaserte tilnærmingen å være spesielt egnet når en eller flere av følgende betingelser inntreffer (Blamey & Mackenzie, 2007; Hills & Junge, 2010);

- *evalueringens formål er læring og kunnskapsbygging,*
- *tiltaket som evalueres er komplekst,*
- *det foreligger lite empirisk kunnskap om tiltakets kausale sammenhenger og effekter,*
- *tiltaket påvirkes av eksterne effekter eller kontekst, eller*
- *det forventes at tiltaket eller dets omgivelser vil utvikles over tid.*

I neste delkapittel diskuteres den teoribaserte evalueringens evne til å frembringe anvendbar kunnskap for den som skal legge strategiske planer, ta beslutninger og iverksette tiltak.

5.7 KUNNSKAPSPRODUKSJON I ET ANVENDT PERSPEKTIV

Teoribaserte evalueringer er ansett for å være særskilt nyttig for å frembringe strategisk kunnskap for beslutningstakere i offentlig sektor (Riché, 2012; van der Knaap, 2004). Antakelsen er nært knyttet til programteoriens mulighet til å bygge på de politiske målsettingene som ligger til grunn for investering i tiltak, og metodikkens evne til å håndtere kompleksitet og opprettholde fokus i evalueringprosessen. Der hvor eksperimentelle analyser er basert på vitenskapens krav til stringens og validitet, benytter den teoribaserte modellen en praktisk, pragmatisk tilnærming basert på logikk og erfaring.

Evalueringens evne til å forklare kausale årsakssammenhenger, og muligheten for å inkludere kontekstuelle faktorer i vurderingen, bidrar til anvendbar kunnskap om hvilke virkemidler som har effekt, for hvilke målgrupper og i hvilke situasjoner. I tillegg til å besvare hvorvidt et tiltak virker, har metodikken kraft til å forklare i hvor stor grad kombinasjonen av ulike virkemidler bidrar til det endelige resultatet, og hvilke eksterne forhold som styrker eller hindrer oppnåelsen av et ønsket resultat (Chen, 1990; Pawson & Tilley, 1997; Rossi, Freeman, & Lipsey, 2003; Sanderson, 2002). I motsatt utfall, dersom den forventede effekten uteblir, kan evalueringen også forklare hvorfor tiltaket ikke virker. Det skiller i denne sammenheng mellom *teorifeil* og *implementeringsfeil* (Bredgard et al., 2016; Dahler-Larsen, 2013). Dersom det foreligger en *teorifeil*, er det antakelsene som ligger til grunn for programteorien som ikke stemmer, og man kan anta at tiltaket ikke har den effekten man forventet i utgangspunktet. Foreligger det derimot en *implementeringsfeil*, er resultatet påvirket av eksterne forhold som kan fikses for å oppnå bedre resultater i en fremtidig implementasjon. For beslutningstakere og produkt- og tjenesteutviklere er det helt essensielt å vite

om det er selve konseptet for tiltaket som er galt eller om det er forhold ved iverksettelsen av det som hindrer at man oppnår ønsket effekt.

Urban et al. (2014) argumenterer for at evalueringer av komplekse sosiale tiltak må inneha et evolusjonsperspektiv som tar høyde for at både tiltakene og omgivelsene rundt er i stadig endring. Tiltaket er dermed ikke en statisk størrelse som skal bevises for å oppdage "den perfekte tilstanden", men en del av en livssyklus med prøving, feiling og stadig forbedring. Som en følge av dette er det urealistisk å forvente at tiltaket skal fungere på et modent nivå fra starten av og det argumenteres for at metodevalg og evalueringsdesign må baseres på hvor tiltaket er i sin utvikling. Dette er relevant for ITS-tiltak som gjennomgår en kontinuerlig teknologisk utvikling. Marchal, Van Belle, De Brouwere og Witter (2013) argumenterer også for at forskning på komplekse intervensjoner krever et dynamisk forskningsdesign, som kan tilpasses etter hvert som ny kunnskap og uforutsette problemstillinger avdekkes. Det stilles videre krav om at den som skal utføre slike studier må ha en adaptiv og fleksibel holdning, analytisk kapasitet og forskningskompetanse. Disse perspektivene harmoniserer med den teoribaserte evalueringens inkrementelle, kumulative tilnærming til kunnskapsbygging, hvor målet er kontinuerlig læring gjennom en stadig finjustering av tiltakets teoretiske rammeverk (Hills & Junge, 2010; Weiss, 1998). Når evalueringsresultatene systematisk pløyes tilbake i programteorien, vil man både synliggjøre kunnskapshull (forskningsbehov) og bidra til en helhetlig forståelse av tiltakets potensiale.

Teoribaserte evalueringer kan gjennom etableringen av programteorien bidra til å skifte fokus fra de tekniske aspektene av tiltaket til de samfunnsmessige behovene som adresseres. Innenfor innovasjonsforskning omtales dette gjerne som *technology push*-problematikk, fordi mange tiltak iverksettes med utgangspunkt i en ny teknologi man ønsker å ta i bruk, uten at effekter og målsettinger nødvendigvis er redegjort for. Ved å identifisere de kausale sammenhengene øker bevisstheten om hva man faktisk ønsker å oppnå med tiltaket. Den visuelle modellen av de ulike elementene i programteorien, bidrar også til å avklare forventninger og sikre et felles grunnlag for diskusjoner og dialog i prosjektet. På den måten er det også lettere å prioritere forskningsinnsatsen, slik at man fokuserer på de sammenhengene som fremstår med størst potensiale eller mest relevante av andre årsaker (Pawson & Tilley, 1997; van der Knaap, 2004).

Avslutningsvis må også den teoretiske evalueringens mulighet til å veilede og forme tiltaket underveis i evalueringsprosessen trekkes frem som en viktig faktor for beslutningstakere. Sanderson (2002) uttaler at *evaluation is as much about identifying good practice as about identifying the counterfactual*. I dette ligger det en oppfordring om å rette mer fokus mot *hvorfor* et tiltak virker, enn bare å avdekke hvorvidt det virker eller ikke. Evalueringens formative verdi er diskutert i de

foregående kapitlene, og tematikken er høyst relevant for prosjekteiere som kan styre investeringer og justere planer allerede før tiltaket iverksettes. På denne måten kan den teoribaserte evalueringen allerede i konseptutviklingsfasen bidra til økonomiske besparelser og hensiktsmessig ressursstyring.

Riché (2012) etterspør større bruk av teoribaserte evalueringer innenfor EUs forskningsprogrammer. Forfatteren tar utgangspunkt i evalueringer gjennomført innenfor tematikken regional utvikling (bl.a. miljø, helse, telekommunikasjon, transportinfrastruktur og byutvikling), og viser til et behov for evalueringstiltærninger som kan kombinere ulike metoder (eksperimentelle og kvalitative), samt forklare underliggende endringsprosesser og betydningen av kontekst når tiltak iverksettes. Selv om den teoribaserte tiltærningen har en innretning som ifølge forfatteren åpenbart er egnet for politikkkutforming, er den i liten grad benyttet. Dette forklares med ensrettet politisk beslutningskultur og strukturelle forhold rundt innretningen på forskningsprogrammene. For det første pekes det på en generell evidensorientert politikk som favoriserer beslutningsgrunnlag som er basert på den eksperimentelle metoden. Dette gjør at god forskning oppfattes ensbetydende med eksperimentell forskning, og valg av alternative tiltærninger må forsvares og risikerer å bli oppfattet som annenrangs forskning. Dette trekkes også frem av Böhm et al. (2016) som peker på at det er ubehagelig både for oppdragsgiver og forsker å velge alternativer til den tradisjonelle tiltærningen i ITS-evalueringer, fordi man dermed risikerer sin vitenskapelige anseelse. Videre vises det til at EU-programmene som regel mangler en tydelig uttalt programteori i oppstartsfasen, og det blir dermed opp til forskeren å gjøre nødvendig forarbeid for å avdekke og etablere tiltakets logikk og endringsprosesser (Riché, 2012). Det er imidlertid vanskelig å skaffe forskningsmidler til en slik utvikling av programteorien i en tidlig fase av evalueringsprosessen. De fleste søknadsformat stiller krav om å binde seg til et evalueringsdesign og spesifisere metoder og analyser på detaljnivå allerede i søknadsfasen (Marchal et al., 2013; Riché, 2012). Dette gjør det krevende å selge inn en teoribasert tiltærning til evalueringen i prosjektene.

Dette kapittelet har drøftet hvordan den teoribaserte tiltærningen kan være et nyttig verktøy som bidrar til å utvide kunnskapsgrunnlaget for implementering av ITS i transportsystemet. I kapittel 6 og 7 presenteres erfaringene med praktisk utprøving av en teoribasert evalueringsprosess i to ulike case-studier med evaluering av ITS-tiltak.

6 CASE A: NONSTOP KLARERING AV TUNGE KJØRETØY

*Dette kapittelet presenterer en evaluering av ITS-tiltaket **NonStop klarering av tunge kjøretøy**.*

Evalueringen er utført i regi av forskningsprosjektet NonStop, og inngår som et case-studium i dette doktorgradsarbeidet. NonStop-prosjektet ble gjennomført av SINTEF Teknologi og samfunn med Seniorrådgiver Terje Moen som prosjektleder, og Seniorrådgiver Trond Foss og PhD-kandidat Lone-Eirin Lervåg som sentrale medarbeidere. Ved prosjektets oppstart var det lagt til grunn et samarbeid mellom prosjektet og foreliggende PhD-studium, hvor prosjektet bidro med data og resultater i case-studiet, og hvor doktorgradsarbeidet bidro med utvikling av evalueringsmetodikk og innspill til evalueringsarbeidet. PhD-kandidaten har hatt hovedansvaret for gjennomføring og rapportering av følgende arbeidsoppgaver i prosjektet:

- *Utvikling av evalueringsmetodikk*
- *Behovsundersøkelse i transportbransjen*
- *Brukerundersøkelse blant kontrollører*
- *Organisering av den manuelle datainnsamlingen i demonstrator*
- *Analyser av data fra demonstrator*
- *Nyttekostnadsanalyse*
- *Følsomhetsanalyser*

Prosjektet er rapportert i Lervåg, Foss, Moen og Aakre (2015).

Innledningsvis presenteres rammene for evalueringen, med beskrivelse av tiltaket og NonStop-prosjektet. Videre gjøres det rede for eksisterende kunnskap om ITS-tiltak for klarering av tunge kjøretøy. Tidlig i prosjektet ble det utarbeidet en evalueringsplan basert på tradisjonell eksperimentell evalueringsmetodikk, med før- og etterundersøkelser av en NonStop-tjeneste som skulle implementeres og demonstreres over en tidsperiode på 12 måneder. Underveis i prosjektgjennomføringen ble det imidlertid klart at man ikke ville oppnå et system med tilstrekkelig teknisk funksjonalitet og modenhet. Disse erfaringene er kort beskrevet, og det er vist hvordan dette også fikk implikasjoner for evalueringsmetodikken i prosjektet. Det ble utarbeidet en ny evalueringsprosess, denne gangen med et teoribasert evalueringsperspektiv. Det er gjort rede for den teoribaserte evalueringsprosessen som ble gjennomført i prosjektet, med analyser og resultater, etterfulgt av konklusjonene fra evalueringen. Til slutt drøftes hvordan case-studiet har bidratt til ny kunnskap om dette ITS-tiltaket spesielt og om evaluering av ITS-tiltak generelt.

6.1 BAKGRUNN

Statens vegvesen har ansvar for kontroll av tunge kjøretøy med formål å avdekke forhold som angår trafiksikkerhet eller påvirker konkurransesituasjonen i transportnæringen. Dette inkluderer blant annet kjøretøy med overlast, sjåfører som ikke overholder kjøre- og hviletid, og kjøretøy eller materiell som er i dårlig stand. I dag er kontrollvirksomheten bygd opp rundt manuelt betjente kontrollstasjoner, og det brukes til dels mye tid på å stanse og kontrollere også de kjøretøyene som overholder gjeldende regelverk. Når det gjennomføres tungbilkontroller ledes alle passerende kjøretøy over 7,5 tonn inn til kontrollstasjonen, hvor de står i kø for å gjennomføre en vektkontroll. Resultatet fra vektkontrollen, kombinert med kontrollørens subjektive teft, avgjør hvilke kjøretøy som skal plukkes ut til videre inspeksjon. Når kapasiteten på kontrollområdet er nådd, får etterfølgende kjøretøy passere ukontrollert (lekkasje), inntil det igjen er plass til nye kjøretøy på kontrollområdet. Prosessen er lite effektiv siden kun en begrenset andel av trafikantene kontrolleres, med derav lav oppdagelsesrisiko for de som kjører ulovlig. Mannskapet som jobber på kontrollstasjonen rapporterer også om tilfeller hvor sjåfører eller transportselskap utnytter denne situasjonen for å unndra seg kontroll, for eksempel ved å kjøre til side og vente til kapasitetsgrensen på stasjonsområdet er nådd. Samtidig opplever lovlige transportører å bli utsatt for unødvendig venting på kontrollstasjonen.

Statens vegvesen etablerte i 2012 forskningsprosjektet *NonStop* som hadde til hensikt å effektivisere transportnæringen og offentlig sektor gjennom å utvikle, implementere og evaluere et automatisk system for målrettet utvelgelse av kjøretøy for tungbilkontroll. Hovedaktiviteten i dette prosjektet ble å utvikle og ta i bruk ITS-tiltaket *NonStop klarering av tunge kjøretøy*, som skulle bidra til å oppfylle følgende målsettinger:

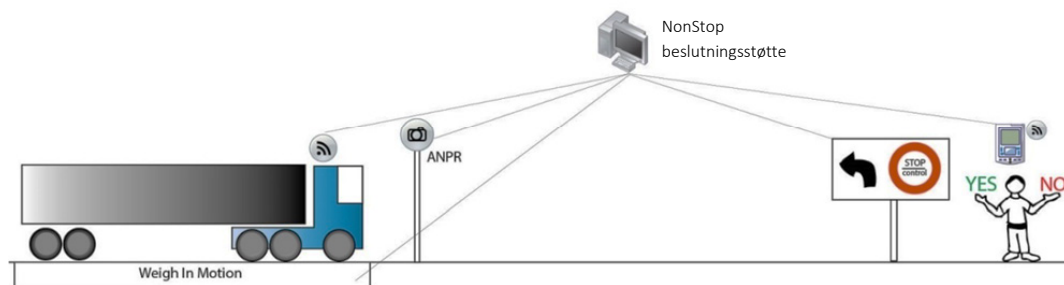
- Økt effektivitet for transportnæringen, gjennom reduksjon av antall unødvendige stopp i kontroller og større forutsigbarhet med hensyn til kontrollsituasjon
- Mer rettferdige konkurransevilkår for transportnæringen, gjennom økt sannsynlighet for å avdekke transportører som ikke overholder gjeldende lover og regler
- Mer effektiv ressursbruk for kontrollmyndighetene, ved å sikre en mer målrettet utvelgelse av kjøretøy som kontrolleres
- Samfunnsøkonomiske gevinster gjennom bedre miljø, økt trafiksikkerhet og mer kunnskap om dagens transporter

6.1.1 KORT BESKRIVELSE AV ITS-TILTAKET NONSTOP KLARERING AV TUNGE KJØRETØY

NonStop klarering av tunge kjøretøy har til hensikt å detektere, lese og kontrollere data fra individuelle kjøretøy i det øyeblikket kjøretøyet passerer registreringspunktet. Innsamlingen av data skjer mens kjøretøyet holder normal hastighet i en trafikkstrøm. Ved å etablere slike registreringspunkt i forkant av kontrollstasjoner, får kontrolloperatøren mulighet til å foreta en selektiv og målrettet forhåndsutvelgelse av hvilke kjøretøy som skal stanses for kontroll. Samtidig får kjøretøy uten indikasjoner på feil og mangler fordelen av å passere kontrollstasjonen uten videre forsinkelser.

NonStop-tiltaket som ble testet i dette studiet består av følgende ITS-tjenester (se også Figur 13):

1. **Veiing i fart (Weigh-in-Motion/WIM):** Deteksjon og innsamling av data om aksellast og totalvekt for kjøretøy.
2. **Nummerskiltgjenkjenning (ANPR):** Registrering av nummerskilt, med påfølgende oppslag i register for identifisering av kjøretøy.
3. **NonStop beslutningsstøttesystem:** Lesing og kontroll av data, inkludert oppslag i register for vurdering av tillatt vekt og aksellast, forsikringsforhold og eventuelle begjæringer.



Figur 13: Forenklet skisse av NonStop-systemet

Tiltaket utgjør en samvirkende ITS-tjeneste, hvor informasjon utveksles mellom vegkantsystemet (sensorer), en sentral ITS-stasjon (NonStop beslutningsstøttesystem) og en personlig ITS-stasjon hos kontrollør (iPad). NonStop-tjenesten henter vekt- og nummerskiltdata fra sensorer langs veien, og gjør oppslag i Statens vegvesens kjøretøyregister (AUTOSYS). Dette sammenstilles i et beslutningsstøttesystem (illustrert med en PC i figuren), hvor en definert regelmotor (som er utarbeidet i NonStop-prosjektet) styrer hvilke kandidater som flagges for kontroll og hvilke som klareres og får kjøre videre. Denne informasjonen sendes deretter til en håndholdt iPad som betjenes på kontrollstasjonen, slik at kjøretøyet status gjøres kjent for kontrolløren før kjøretøyet ankommer.

NonStop-tjenesten kan anvendes til ulike formål, avhengig av hvordan tiltaket utformes og tas i bruk. Fire relevante bruksområder ble vurdert i NonStop-prosjektet:

1. *Helautomatisk utvelgelsessystem*: Utvelgelsen av kontrollkandidater er fullstendig basert på NonStop-tjenestens indikasjoner, og innkjøring til kontrollstasjonen kan i utgangspunktet reguleres med automatiske lyssignaler som styres av informasjon fra WIM og ANPR-registreringer.
2. *Beslutningsstøttesystem*: Utvelgelsen av kontrollkandidater styres av den enkelte kontrollør, som mottar indikasjoner fra NonStop-tjenesten som en støtte til manuell utvelgelse. Informasjon fra WIM og ANPR-registreringer (NonStop-tjenesten) suppleres av kontrollørens subjektive vurderinger knyttet til tidligere erfaringer og kjøretøyets tilstand.
3. *Varslingssystem (basert på kontinuerlig datainnsamling)*: Det foretas kontinuerlig registrering av de variablene som inngår i NonStop-tjenesten, uavhengig av øvrig kontrollvirksomhet. Ved deteksjon av alvorlige overtredelser sendes et varsel/alarm til kontrollørene.
4. *Planleggingssystem/statistikkformål*: Det foretas kontinuerlig registrering av de variablene som inngår i NonStop-tjenesten, uavhengig av øvrig kontrollvirksomhet. Dette datagrunnlaget benyttes til planlegging av kontrollvirksomheten gjennom identifisering av hyppige overtredelser og passeringstidspunkt.

6.1.2 KORT BESKRIVELSE AV PROSJEKTET NONSTOP

Dette case-studiet er gjennomført som del av forskningsprosjektet NonStop i Forskningsrådets SMARTRANS-program i perioden 2012-2015 (Forskningsrådet, 2015). Prosjektet ble gjennomført med Statens vegvesen Vegdirektoratet som prosjekteier, SINTEF som forskningspartner og med Norges Lastebileierforbund og CIBER som øvrige samarbeidspartnere, se rollebeskrivelser i Tabell 7.

Tabell 7: Rollebeskrivelser i NonStop-prosjektet

Partner	Rolle i prosjekter	Kunnskap/kompetanse
Statens vegvesen, Vegdirektoratet	Prosjekteier. Systemeier og ansvarlig for å integrere kontrollsystemet med operativ kontrollvirksomhet.	Drift og vedlikehold av vegsystemet, kontrollvirksomhet
Norges Lastebileierforbund	Prosjektdeltager med ansvar for å formidle transportnæringens interesser.	Transportnæringens behov og muligheter
CIBER	Prosjektdeltager med ansvar for å integrere systemene for realisering av NonStop-tjenesten.	Applikasjons utvikling, automatisering og integrasjon mot sensorer og baksystemer.
SINTEF Teknologi og Samfunn/IKT	Prosjektledelse og FoU partner. Ansvarlig for systemspesifikasjon, demonstrasjon og evaluering av NonStop-tjenesten.	Prosjektledelse, trafikkteknikk, IKT, systemutvikling, utvikling av demonstrator og evalueringsmetodikk

De sentrale forskningsutfordringene i prosjektet var knyttet til vurdering av nyttevirkinger for transportnæringen og utøvende kontrollmyndighet, og sekundært skulle forskningsbidrag i en tidlig fase av prosessen sikre optimal utforming av systemet med hensyn til å samle inn data og informasjon som er etterspurt i forbindelse med planlegging og investering i infrastruktur.

Prosjektet var strukturert og planlagt gjennomført med utgangspunkt i følgende hovedaktiviteter:

1. **Behov- og mulighetsstudier:** Kartlegge teknologiske muligheter med sensorer og bruk av datakilder, samle erfaringer med kontrollvirksomhet i andre land, samt behovsundersøkelse i transportbransjen.
2. **Demonstrator:** Utarbeide kravspesifikasjon, utvikle beslutningsstøtteverktøy for kontrollører og system for varsling/stansing av kjøretøy, etablere demonstrator og demoperiode med datainnsamling.
3. **Evaluering:** Utvikle evalueringsmetodikk, vurdering av teknologisk modenhet, effektevaluering, samt vurdering av brukeraksept for kontrolloperatør og fører/transportsselskap.
4. **Forretnings- og forvaltningsmodell:** Vurdere modeller for lønnsom drift av nye produkter og tjenester, samt forvaltningsmodell for fremtidig utrulling i offentlig sektor.

Prosjektet og aktivitetene er nærmere beskrevet i Lervåg et al. (2015).

6.2 EKSISTERENDE KUNNSKAP OM NONSTOP KLARERING AV TUNGE KJØRETØY

Da utviklingen av NonStop-tjenesten ble innledet i 2012, fantes det allerede noen få installasjoner av sensorer for veiing av kjøretøy i fart (WIM) i tilknytning til norske kontrollstasjoner. Derimot eksisterte ingen løsning hvor de ulike ITS-tjenestene var koblet sammen til et fullverdig system, med nummerskiltgjenkjenning og beslutningsstøtte for kontrollørene. I USA, Canada og enkelte europeiske land (f.eks. Frankrike, Slovenia, Tyskland og Storbritannia) hadde man imidlertid kommet noe lengre, med utprøving og drift av løsninger for screening av kjøretøyvekt i forkant av kontrollstasjoner siden tidlig på 2000-tallet (FHWA, 2007; Nova Scotia Transportation and Infrastructure Renewal, 2008). Teknologien som benyttes i disse installasjonene (sensorer og dataregistre) varierer mellom land og testområder, samtidig som ITS-tjenestene også kan ha ulike bruksområder; eksempelvis forhåndsutvelgelse av kjøretøy til vektkontroll og/eller datainnsamling for planlegging og oppfølging av kontrollvirksomheten. Noen land (f.eks. Nederland, Frankrike og Storbritannia) lager også statistikk over individuelle transportselskaper, slik at aktører som til stadighet begår overtredelser blir fulgt opp mer spesifikt av kontrollmyndighetene, for eksempel med oppsøkende virksomhet i bedriften og økt kontrollfrekvens (Jacob & Feypell-de La Beaumelle, 2010). Senere er det blant annet i Frankrike, Tsjekkia og Taiwan også jobbet med pilotprosjekter hvor utøvende kontrollmyndighet håndhever vektbestemmelsene direkte og automatisk basert på WIM (Jacob & Cottineau, 2016), slik vi i dag anvender automatisk fartskontroll (ATK) til håndheving av fartsgrenser. Direkte håndheving stiller høye krav til teknisk funksjonalitet og presisjon, men innebærer økt produktivitet fordi det blir sanksjonert overfor transportørene også når kontrollstasjonen er ubemannet (f.eks. ved at gebyr sendes per post).

De første systemene som ble etablert, hadde til hensikt å redusere slitasjen på asfaltdekket og vegkonstruksjonen forårsaket av overlastede kjøretøy, samt å redusere kostnadene knyttet til selve kontrollvirksomheten (Aakre, Engen, & Roche-Cerasi, 2014; FHWA, 2007). Etter hvert har effektbildet blitt utvidet, med større fokus på sekundære effekter som produktivitet og konkurransevilkår for næringslivet, miljøutslipp, trafiksikkerhet, samt å generere data for statistikk- og planleggingsformål (Aakre et al., 2014; Groot, 2005; Jacob & Feypell-de La Beaumelle, 2010; Poulikakos, Heutschi, Arraigada, Anderegg, & Soltic, 2010). Selv om ITS-tjenester for klarering av tunge kjøretøy stadig utvikles og testes i større utstrekning, finnes det få eksempler på faktiske effektstudier, og det eksisterer nesten ikke vitenskapelig dokumentasjon av konkrete effekter og virkninger over tid. Publiserte forskningsresultater er i hovedsak rettet inn mot teknisk funksjonalitet, med fokus på sensorenes presisjon og nøyaktighet (Fitzgerald, Sevillano, O'Brien, & Malekjafarian, 2017; Haugen, Levy, Aakre, & Tello, 2016; Jacob & Cottineau, 2016;

Jones, 2008). I den grad øvrige effekter beskrives, er dette gjerne som et utfall av logiske resonnement, men i liten grad knyttet til faktiske målinger og undersøkelser (Jacob & Feypell-de La Beaumelle, 2010). Oppdatert kunnskap om utbredelse og status på WIM-installasjoner og ITS-tjenester for klarering av tunge kjøretøy finnes i hovedsak på respektive vegmyndigheters nettsider, og i mindre grad i akademiske kilder.

I påfølgende avsnitt oppsummeres eksisterende kunnskap om ITS-tjenester for klarering av tunge kjøretøy. Planlegging av NonStop-prosjektet foregikk i 2012, og baserer seg således på litteratur om tilsvarende systemer fra før dette tidspunktet. Dette omfatter i hovedsak erfaring med relevante WIM-installasjoner i Frankrike (Stanczyk, Geroudet, Thiounn, & Millot, 2008; Stanczyk & Klein, 2012), Storbritannia (Jones, 2008) og USA (NORPASS, 2012; Stephens, Carson, Hult, & Bisom, 2003). Resultatene fra litteraturstudien i prosjektet er blant annet presentert i Aakre et al. (2014) og i prosjektrapporten (Lervåg et al., 2015).

6.2.1 POTENSIELLE EFFEKTER OG VIRKNINGER

Nøyaktighet og presisjon

ITS-tjenester for klarering av tungbiler er bygd opp av flere bestanddeler (sensorer, dataregister, IKT-systemer) som alle innehar en teknisk kvalitet og funksjonalitet som påvirker hvorvidt systemet genererer pålitelige data. Systemets nøyaktighet og presisjon styrer i hvilken grad systemet er i stand til å plukke ut de riktige kontrollkandidatene. Dette er ikke en *effekt* i seg selv, men en viktig *forutsetning* for at ITS-tiltaket skal yte som forventet og realisere potensielle effekter. Det finnes en del forskningsbasert dokumentasjon på nøyaktighet og presisjon for ulike typer WIM-teknologi og installasjoner (f.eks. Fitzgerald et al., 2017; Haugen et al., 2016), og for hvordan dette gir utslag i form av deteksjons- eller feilrater i kontrollvirksomheten (f.eks. Groot, 2005; Jones, 2008; Stanczyk et al., 2008; Stanczyk & Klein, 2012).

Kravet til nøyaktighet for WIM-installasjoner er spesifisert i COST 323 (2001). Toleransegrensen varierer avhengig av systemets bruksområde. Dersom WIM-systemet brukes til vektkontroll (direkte håndheving) tillates en feilmargin inntil 5%, mens installasjoner som benyttes til forhåndsutvelgelse av kjøretøy for vektkontroll aksepterer en feilmargin i størrelsesorden 7-10 %. Her i Norge er det utført testing av to typer sensorteknologi for WIM-installasjoner fra fire ulike leverandører (Haugen et al., 2016). En av testene ble gjennomført ved Sandmoen kontrollstasjon som del av NonStop-prosjektet (N=511 kjøretøy), mens de øvrige testene ble gjennomført på Verdal kontrollstasjon (N=31, 33 og 83) i 2014/2015. Resultatene viste feilmarginer i størrelsesorden 2-15%, men de samme sensortypene ga ulike resultater i forskjellige installasjoner.

Studien konkluderte med at kalibreringsprosedyrene er tid- og kostnadskrevede, nøyaktigheten reduseres over tid, og sensorene har forholdsvis kort levetid. Tilsvarende WIM-system som benyttes i Frankrike er testet over en toårsperiode (N=254) og oppgitt å ha en nøyaktighet som innebærer feilrate innenfor 4% (Stanczyk & Klein, 2012). Det jobbes stadig med utvikling av nye sensorer og algoritmer for å forbedre WIM-systemenes nøyaktighet og presisjon (Fitzgerald et al., 2017; Jacob & Cottineau, 2016).

I Frankrike var det i 2015 etablert totalt 29 kontrollpunkter med WIM og nummerskiltgjenkjenning for kontinuerlig registrering av overlast for tunge kjøretøy (Jacob, 2015). Dette omfatter årlig veiing av i størrelsesorden 20 millioner kjøretøy, hvorav totalt 50.000 kjøretøy har blitt stoppet og kontrollert på statisk vekt i tillegg. Dataene viser at omtrent 10 % av tungtrafikken på franske veier kjører med overlast. Før WIM-punktene ble etablert, var deteksjonsraten på kontrollstasjonene 25 %, mens man i etterkant av WIM-punktene registrerte overlast hos 96 % av kjøretøyene som ble stoppet. Erfaringer fra Frankrike er blant annet rapportert i Stanczyk og Klein (2012) og Stanczyk et al. (2008) som viste deteksjonsrater på 96 % ved henholdsvis Boulou (N=145) og Le Muret (N=280).

Storbritannia har benyttet en kombinasjon av WIM og nummerskiltgjenkjenning for målrettet kontroll av kjøretøy med overlast siden 2004/2005. I tillegg gjøres det bruk av historiske data på transportørnivå, slik at hvert kjøretøy som ankommer vurderes ut ifra tidligere atferd²⁷. Jones (2008) har dokumentert systemets nøyaktighet og presisjon, basert på testing i seks måneder med 240 daglige varslinger om overvekt, hvorav 6 % av disse kjøretøyene ble tatt inn for kontrollmåling. Resultatene dokumenterte at feilmarginen er lavere enn 10 % for alle kjøretøy totalt, og bedre enn 5 % for ¾ av alle kjøretøyene som ble kontrollmålt. 90 % av alle kjøretøyene som ble valgt ut til kontroll hadde en faktisk vektoverskridelse som resulterte i bøtelegging. Til sammenligning er den gjennomsnittlige deteksjonsraten på andre kontrollstasjoner i Storbritannia bare 24 %.

Generelt synes studiene å rapportere om forholdsvis gode erfaringer med systemenes deteksjonsrate, selv om det i liten grad er gjort rede for øvrige aspekter knyttet til nøyaktighet og presisjon. De oppgitte deteksjons- eller feilratene er imidlertid ikke sammenlignbare mellom ulike studier, da det legges litt forskjellige kriterier til grunn for vurderingene. Ulike land opererer med forskjellige grenseverdier for hva som regnes som overvekt, gjerne en prosentdel av maksimalt tillatt kjøretøyvekt, som gjenspeilet nivået for når den respektive kontrollenheten iverksetter sanksjoner overfor transportøren. I de franske og britiske studiene er denne grenseverdien oppgitt å

²⁷ Operatørene får beregnet en risikoscore basert på tidligere atferd (Operator Compliance Risk Scores)

være hhv. 5 % og 8 % overskridelse av tillat maksimalvekt. Videre bør feilratene i utgangspunktet differensieres i to typer feil; 1) kjøretøy som stanses uten å ha overlast (falsk positiv) og 2) kjøretøy med overlast som passerer kontrollstasjonen uten å bli markert (falsk negativ). De refererte studiene gir kun indikasjoner på feil av typen *falsk positiv*, hvor man registrerer hvor mange av de forhåndsutvalgte kjøretøyene som egentlig skulle fått passere uhindret. Dette gir derfor et ufullstendig bilde av systemenes faktiske nøyaktighet og presisjon.

Effektivitet i kontrollvirksomheten

Målrettet utvelgelse av kandidater for tungbilkontroll gir kontrollørene mulighet til å rette oppmerksomheten mot de kjøretøyene som har størst sannsynlighet for å begå lovovertridelser. Som beskrevet i forrige avsnitt, viser evaluering av ITS-tjenester for klarering av tunge kjøretøy lovende resultater med hensyn til å oppnå en høyere deteksjonsrate. Når kontrollørene bruker mindre tid på rutinepreget vektkontroll, gir dette mulighet til å rette fokus mot mer komplekse kontroller, som eksempelvis kabotasje og overholdelse av kjøre-/hviletid. Effektivisering av kontrollvirksomheten medfører også økonomiske besparelser knyttet til personell og utstyr på kontrollstasjonen. Det er imidlertid ikke funnet studier som har dokumentert de økonomiske effektene for kontrollvirksomheten, kun rapporter som argumenterer for at dette vil være en logisk konsekvens. Karim, Ibrahim, Saifizul og Yamanaka (2014) fokuserer blant annet på hvordan WIM-data kan anvendes for planlegging og effektivisering av fremtidig kontrollvirksomhet. Studien viser til at Malaysia, som mange andre utviklingsland, har et stort antall lovbrudd, og i dag begrenses kontrollmulighetene av kapasiteten til kontrollmannskapet. Screening av kjøretøy på ett sted over en måned (januar 2010) viste at 21,5 % av alle kjøretøy som passerte hadde overlast, med typiske utslag på spesifikke dager, tidspunkt og type kjøretøy. Muligheten til å fokusere ressursinnsatsen anses derfor å gi god effektivitet.

Beskyttelse av asfaltdekke og vegkonstruksjon

Kjøretøy med overlast påfører vegen store skader, med dertil påfølgende kostnader til drift og vedlikehold (Jacob & Feypell-de La Beaumelle, 2010; O. Larsen, 2012). Vegslitasje estimeres med utgangspunkt i kjøretøyenes akseltrykk og betegnes i form av såkalte ESAL²⁸-kilometer. Dette innebærer at høy vekt på enkelte akslinger kan ha større betydning for vegslitasje enn kjøretøyets totale vekt. ITS-tiltak som kan bidra til å øke deteksjonsraten på svært tunge kjøretøy – og som også kan ha en oppdragende effekt i transportbransjen, har potensiale til å redusere vegslitasjen vesentlig. En pilotstudie fra Montana²⁹ i USA har studert hvordan bruk av WIM for overvåking og utvelgelse

²⁸ Equivalent Standard Axle Loadings

²⁹ State Truck Activities Reporting System, STARS

av kjøretøy for vektkontroll påvirker overlast og vegslitasje (Stephens et al., 2003). Det ble gjennomført en før- og etterstudie i fem testområder, med vegslitasje målt og uttrykt som funksjon av en ESAL-faktor på månedlig basis. De fem testområdene ble valgt ut på grunn av store problemer med vegslitasje i utgangspunktet, og resultatene ble sammenlignet med en kontrollgruppe bestående av 11 andre målingspunkt. Resultatene viste at målrettet utvelgelse av kjøretøy for vektkontroll basert på WIM-data bidro til følgende effekter; flere tunge kjøretøy som overholdt vektreglementet, reduksjon i andel kjøretøy med overlast (22%), reduksjon i gjennomsnittlig størrelse på overlast, reduksjon i skader på asfaltdekket (samtidig som det ble registrert økt skadeomfang på veger uten målrettet vektkontroll). Effektene vedvarte noen måneder etter at testperioden var over, men det ble ikke registrert langtidseffekter på kjøretøyvekt.

Endret atferdsmønster i transportbransjen

Evalueringer fra Frankrike (Stanczyk & Klein, 2012) og Nederland (FHWA, 2007; Jacob & Feypell-de La Beaumelle, 2010) har vist at streng håndhevelse av vektbestemmelsene kan ha oppdragende effekt blant transportaktørene. I Frankrike ble det rapportert en signifikant nedgang på over 20 % i andel kjøretøy med overlast etter innføring av WIM-punktene, til tross for generell trafikkøkning i samme periode³⁰. Effekten var enda større for kjøretøy med veldig stor overlast. I begge disse landene brukes ITS-tjenesten for å identifisere transportselskap som jevnlig bryter reglene, og/eller som har alvorlige tilfeller av overlast. Transportselskap som havner på oppfølgingslisten får advarsler og besøk av inspektører, og de blir utsatt for flere kontroller langs vegen. Studiene viser at slike kontrollsystemer har positiv effekt på adferdsmønsteret hos transportselskapene, med prosentvis reduksjon i både overskridelser av aksellast og total overlast på kjøretøyene. Den nederlandske studien viser også økt effektivitet (deteksjonsrate) i vektkontroller og rapporterer at transportselskapene investerer i nye, hensiktsmessige kjøretøy, samt egne vektsystemer på terminalene for bedre å ivareta internkontroll (FHWA, 2007). Dersom statistikk over passeringshistorikk gjøres tilgjengelig for transportoperatøren, kan dette datamaterialet danne grunnlag for bedre planlegging av godsfordeling og utnyttelse av kjøretøymateriell over tid. Det er imidlertid ikke funnet forskning som dokumenterer dette, utover de nevnte rapportene.

ITS-tjenestene som beskrives i dette avsnittet er ikke direkte overførbare til Norge, fordi man har ulik organisering av kontrollvirksomheten (henholdsvis trafikkpoliti og vegmyndighet), i tillegg til nasjonale forskjeller i personvernlovgivningen mellom landene.

³⁰ Registreringer fra 2009-2010, ikke oppgitt antall kjøretøy som inngår i undersøkelsen

Produktiviteten i transportbransjen

Reisetid og forutsigbarhet i godsleveranser har stor økonomisk betydning for næringslivet. Dagens kontrollregime medfører unødvendige forsinkelser for kjøretøy som har papirer og vekt i orden, men som likevel stanses i rutinekontroller. Videre vil en lav risiko for å bli oppdaget kunne friste useriøse aktører til å kjøre ureglementert (f.eks. med overlast for å unngå å sette inn et kostbart ekstra kjøretøy).

Jacob and Feypell-de La Beaumelle (2010) viser til estimeringer fra Frankrike som indikerer at et vogntog (med 5 akslinger) som opererer med 20 % overlast oppnår en årlig urettmessig besparelse i størrelsesorden 25.000 €. Økonomisk gevinst av å jukse har en ugunstig konkurransevridende effekt i transportbransjen og for samfunnet for øvrig. ITS-tiltak som øker deteksjonsraten på kjøretøy som bryter lover og regler, og som samtidig reduserer reisetid og sørger for bedre forutsigbarhet for lovlige aktører, vil bidra til økt produktivitet og mer rettferdige konkurransevilkår i transportbransjen.

Amerikanske beregninger har anslått at hver sjåfør sparer 5 minutter for hver gang han får kjøre uhindret forbi kontrollstasjonen, og at dette medfører besparelser for transportindustrien i størrelsesorden 9,60 USD for hver passering³¹. Det er imidlertid ikke funnet forskningsrapporter som dokumenterer denne effekten. Brown et al. (2007) presenterer en analyse av økonomiske aspekter knyttet til ITS-tiltak for næringstransporten som innebærer elektronisk klarering av kjøretøy ved vektstasjoner, samt innføring av web-baserte løsninger for dokumentkontroll. Analysen er basert på intervjuer med 60 aktører fra transportbransjen, med kartlegging av deres erfaringer fra deltakelse i et amerikansk pilotprosjekt (Commercial Vehicle Information Systems and Networks, CVISN). Resultatene indikerte store fortjenester for transportbransjen i et tiårsperspektiv, med investeringene tilbakebetalt allerede i løpet av første driftsår. De viktigste nytteeffektene som ble trukket frem var spart reisetid, økt effektivitet og økt mulighet for samarbeid mellom myndigheter og private transportaktører.

I Canada ble det gjennomført en spørreundersøkelse i transportbransjen i forbindelse med utprøving av WIM-system på kontrollstasjoner over en periode på fem måneder (Nova Scotia Transportation and Infrastructure Renewal, 2008). Resultatene viste høy aksept blant transportørene, med 87 %

³¹ Beregninger fra 2016 presentert på nettsidene til Washington State Department of Transportation (<http://www.wsdot.wa.gov/CommercialVehicle/CVISN/technology.htm>). Oppdaterte tall for 2018 indikerer fortsatt redusert forsinkelse på **5 min/passering** og besparelse for transportoperatøren på **\$9.95/passering**.

som svarte at de var enig i at vegstrekningen ble tryggere med WIM, 86 % som opplevde at systemet målte presist, og 87 % som opplevde at systemet bidro til spart reisetid.

Miljø

Kjøretøy med overlast genererer negative miljøeffekter i form av ekstra støy, vibrasjoner og utslipp (Nova Scotia Transportation and Infrastructure Renewal, 2008; Poulidakos et al., 2010). ITS-tiltak som bidrar til målrettet vektkontroll av tunge kjøretøy har potensiale til å redusere miljøulemper på to måter; ved å redusere antall kjøretøy med overlast, og ved å redusere omfanget av akselerasjon og bremsing hos klarerte kjøretøy som ikke må stanse på kontrollstasjonen. Måling av miljøeffekter involverer imidlertid komplekse modeller av mange faktorer, og det er ikke funnet studier som kartlegger miljøeffektene av ITS-tiltak for klarering av tungbiler. Det finnes noen få eksempler på studier som estimerer en miljøeffekt basert på drivstoffbesparelse som resultat av redusert reisetid for kjøretøy som opptrer i henhold til regelverket (ERTRAC, 2013; Groot, 2005), uten at det reflekteres over hvorvidt redusert overlast kan medføre at det må brukes flere kjøretøy til å transportere samme mengde gods.

Trafikksikkerhet

Kjøretøy som er overlastet eller har feil vektfordeling kan forårsake bremsesvikt, lengre bremsestrekning og dårligere stabilitet som gjør det vanskelig å manøvrere kjøretøyet på veien. Siden ulykker opptrer forholdsvis sjelden, er det vanskelig å dokumentere trafikksikkerhetseffekter. Til tross for manglende ulykkesstatistikk, er det forventet at overlastede kjøretøy har større sannsynlighet for å bli involvert i hendelser og ulykker (Jacob & Feypell-de La Beaumelle, 2010). Trafikksikkerhetshåndboken (Høye et al., 2012, kapittel 4.30) peker på at ulykker med tunge kjøretøy ofte er mer alvorlige enn andre ulykker, men det er ikke funnet forskning som dokumenterer sammenhengen mellom ITS-tiltak for klarering av tungbiler for vektkontroll og effekter på trafikksikkerhet. Det finnes heller ikke entydig dokumentasjon på sikkerhetseffektene av ordinær vektkontroll. Trafikksikkerhetshåndboken (Høye et al., 2012, kapittel 5.3) viser til at grunnlaget for å tallfeste virkningen av utekontroll på antall ulykker er meget usikkert, men antyder at økning av kontrollhyppigheten sannsynligvis vil bidra til å redusere ulykkestallene, og at virkningen er større for tunge kjøretøy enn for lette kjøretøy.

Trafikkstyring og trafikkavvikling

Kontinuerlig datainnsamling om tungtransporten genererer verdifulle datasett og informasjon som kan benyttes til trafikkstyringsformål. Det er imidlertid ikke funnet eksempler på at ITS-tiltak for klarering av tunge kjøretøy benyttes systematisk til dette formålet. Tunge kjøretøy kan skape avviklingsproblemer på veger med stor stigning på grunn av lavere akselerasjon. Som nevnt i

avsnittet om trafikksikkerhet, antas det også at kjøretøy med overlast er overrepresentert i hendelser og ulykker. Dette har uheldige konsekvenser for fremkommeligheten på veien, da hendelser som involverer tunge kjøretøy ofte medfører store forsinkelser og ulemper for øvrig trafikk. ITS-tiltak som bidrar til å redusere antall kjøretøy med overlast på veien vil følgelig kunne ha en positiv effekt for trafikkavviklingen, men det er ikke funnet forskningsresultater som underbygger denne sammenhengen.

Litteraturgjennomgangen viser at effektene av å ta i bruk ITS-tiltak for klarering av tunge kjøretøy først og fremst øker deteksjonsraten for overlast og at dette har oppdragende effekt i transportbransjen. ITS-tiltak for klarering av tunge kjøretøy antas også å påvirke transportsystemet gjennom to sentrale dimensjoner: For det første blir kontrollvirksomheten mer målrettet fordi ressursene brukes på kjøretøy med stor sannsynlighet for lovbrudd, istedenfor kjøretøy som har utstyr og papirer i orden. Dernest øker kapasiteten på kontrollstasjonen, slik at det blir mindre lekkasje av kjøretøy som passerer ukontrollert på grunn av begrensninger i fysisk areal eller personressurser.

6.2.2 HVILKE EVALUERINGSMETODER BENYTTES?

Det er begrenset erfaring med aktuelle metoder for effektevaluering av ITS-tiltak for klarering av tunge kjøretøy. Der hvor tiltakene er evaluert, er dette i stor grad basert på eksperimentelle studier som sammenligner deteksjonsrater i en periode før og etter innføring av ITS-tiltaket (Jones, 2008; Stanczyk et al., 2008; Stanczyk & Klein, 2012). Disse studiene er drøftet nærmere i avsnitt 6.2.1. Deteksjonsraten brukes som mål på effektivitet i kontrollvirksomheten, og uttrykkes enten som andel kjøretøy med overlast eller andel kjøretøy som blir bøtelagt. Grensene for hva som regnes som overvekt og når sanksjoner iverksettes varierer imidlertid i de ulike studiene, og må tas hensyn til dersom man ønsker å sammenligne resultater mellom ulike prosjekter. I Stephens et al. (2003) dokumenteres slitasje på vegkonstruksjonen uttrykt i ESAL-km, som beskrevet i avsnitt 6.2.1. Metodikken som benyttes er tradisjonelle før- og etterundersøkelser i fem testområder med målrettet kontrollvirksomhet, sammenlignet med en kontrollgruppe bestående av 11 andre målingspunkt. Tiltaket ble imidlertid innført på de fem stedene med størst vegslitasje, og dermed også størst potensiale for endring. Studien er altså utført uten randomiserte eller representative utvalg, og overføringsverdien av resultatene blir noe begrenset. I Frankrike og Nederland har enkelte studier dokumentert endret atferdsmønster i transportbransjen, i form av indikatorene *reduisert antall kjøretøy med overlast* og *reduisert gjennomsnittlig størrelse på overlast* (Jacob & Feypell-de La Beaumelle, 2010; Stanczyk & Klein, 2012). Det er sparsomt med detaljer om metodikken som ligger til grunn for disse analysene, utover at resultatene er basert på kontinuerlig

logging av data på enkelte registreringspunkt over tid, med utarbeidelse av statistikk for individuelle transportselskaper. Hvert land har imidlertid ulike ordninger for spesifikk oppfølging og sanksjoner mot transportfirma, og man må kunne anta en sterk sammenheng mellom type sanksjon og størrelsen på effekt. Felles for alle disse nevnte analysene, er at de baserer seg på vektregistreringer fra WIM-punkt i en begrenset testperiode. I tillegg finnes det noen eksempler på spørreundersøkelser som kartlegger subjektive opplevelser av nytteeffekter i transportbransjen (Brown et al., 2007; Nova Scotia Transportation and Infrastructure Renewal, 2008).

I 2012 ble det gjennomført en ex post evaluering av et nytt felles kontrollområde for toll, vegvesen og politi på den norske grensestasjonen på Svinesund (Finne, Bull-Berg, & Samset, 2012). Det må bemerkes at anlegget er noe mer enn en vanlig kontrollstasjon, da man i tillegg til teknisk kontroll av kjøretøy, også foretar fortolling og smuglingskontroll. Hensikten med studien var å vurdere den offentlige investeringen med hensyn til følgende evalueringskriterier; effektivitet i prosjektet, måloppnåelse, utilsiktede virkninger, relevans, levedyktighet og samfunnsøkonomisk lønnsomhet. Evalueringen benyttet en målorientert evalueringsmodell (OECD, 2002) som inkluderer en programteoretisk tilnærming i evalueringsdesignet. Denne metodikken er senere videreutviklet som del av forskningsprogrammet Concept for evaluering av store, statlige investeringer i infrastruktur (Volden & Samset, 2017). Datagrunnlaget i evalueringsstudien innbefattet befarings på kontrollområdet, intervjuer med 20 involverte aktører, samt dokumentstudier. Rapporten viser at det var vanskelig å finne indikatorer som kunne belyse effekter og virkninger på en enkel og entydig måte. Eksempelvis var økt fremkommelighet en av de trafikale målsettingene ved tiltaket, men det eksisterte ikke gode data på hvor store forsinkelsene var eller hvor ofte de inntraff før tiltaket ble iverksatt, og det ble heller ikke samlet inn den type data i etterkant. Det ble heller ikke konkludert på om investeringen var samfunnsøkonomisk lønnsom, med begrunnelse i to hovedutfordringer: For det første var det vanskelig å etablere et nullalternativ og kontrafaktisk utviklingsspor for sammenligning. Dessuten var det heller ikke mulig å sette økonomisk verdi på nyttevirkningene. Til tross for mangelfullt datagrunnlag, gir rapporten en grundig kvalitativ diskusjon av de seks evalueringskriteriene – og argumenterer for at prosjektet alt i alt synes vellykket.

Zhang, Haas og Tigh (2008) presenterer et rammeverk for nyttekostnadsanalyser av WIM-installasjoner i Canada, basert på tre nyttevariabler; reduserte forsinkelser i transportbransjen, økt trafiksikkerhet og kapasitetsforbedring på kontrollstasjonen. Det er gitt et eksempel på bruk av beregningsmodellen for et scenario med én kontrollstasjon og et trafikkvolum på 500 kjøretøy som daglig må innom kontrollstasjonen. Gjennomsnittlig ventetid er satt til 5 min/kjøretøy, og andel klarerte kjøretøy er 70 %. På kostnadssiden benyttes erfaringstall fra investeringer og drift av WIM-

systemer i Canada, med en beregnet levetid på 10 år. Dette gir nyttekostnadsverdier for variablene *reduisert reisetid, økt trafikkikkerhet og økt kontrollkapasitet* på henholdsvis 1) $N/K = 4.235$, 2) $N/K = 5.077$ og 3) $N/K = 3.046$. Resultatene fra litteraturstudien tyder på at det er vanskelig å anslå trafikkikkerhetseffektene av WIM-tiltak, da det knapt finnes dokumentasjon på slike effekter. Derimot synes det lettere å estimere reisetidsbesparelser som funksjon av deteksjonsrate og antall klarerte kjøretøy. Beregningsmodellene som ble presentert i Zhang (2008) viser at WIM-tiltak som eliminerer unødvendig forsinkelse er en lønnsom investering i de fleste tilfeller, bortsett fra ved svært lave trafikkvolum. Zhang et al. (2008) refererer også en tidligere case-studie hvor nyttekostnadsverdien av spart reisetid knyttet til en WIM-tjeneste i Brunswick, Canada ble estimert for en 5-årsperiode (M. Davis, 2003). Den aktuelle kontrollstasjonen oppnådde en forhåndsklarering av 76,5 % av det totale trafikkvolumet (655 kjt/døgn), hvilket ga en total N/K -verdi på 3,69. Trafikktall og reisetid som ligger til grunn for denne beregningen må imidlertid anses å være svært usikker, da datainnsamlingen kun omfatter en periode på 24 timer. Studien peker imidlertid på to sentrale erfaringer fra studien: For det første krevde WIM-installasjonen grundig oppfølging med vedlikehold og kalibrering, fordi sesong- og temperaturendringer hadde stor påvirkning på systemets nøyaktighet og presisjon. Dernest ble det trukket frem at nyttekostnadseffekten av WIM-tiltaket er stedsspesifikk, og dermed ikke automatisk overførbart til andre kontrollstasjoner. Eksempelvis nevnes det at resultatene vil være avhengig av lokale trafikkforhold og hvilken presisjon som kreves av WIM-systemet (kostnader knyttet til kalibrering).

New York State Department of Transportation gjennomførte i perioden 2003-2012 et forskningsprosjekt med hensikt å utvikle og teste et automatisert og målrettet kontrollsystem, basert på ITS-tjenestene WIM og nummerskiltgjenkjenning (Clough Harbour & Associates, 2012). Prosjektet erfarte at det var krevende å oppnå tilstrekkelig teknisk funksjonalitet og kvalitet, og det ble brukt mer ressurser enn forutsatt på å få de ulike komponentene til å fungere godt sammen i et samvirkende IKT-system. Opprinnelig var det planlagt at systemet skulle testes live i en fire uker lang pilotperiode, men dette ble ikke gjennomført. Omfattende forskning gjennom en tiårsperiode kulminerte i en detaljert spesifisering og endelig design av sensorer og komponenter i et elektronisk screeningsystem, uten at man kom så langt at man fikk testet eller evaluert nytteeffektene.

I Canada ble det gjennomført utprøving og evaluering av bruk av et WIM-punkt for elektronisk screening av kjøretøy i tilknytning til en kontrollstasjon (Nova Scotia Transportation and Infrastructure Renewal, 2008). Også her opplevde man at det ble brukt betydelige ressurser på kalibrering og justering av IT-komponenter for å oppnå tilfredsstillende teknisk funksjonalitet. Kjøretøy- og vektdata fra WIM-punktet ble samlet inn i en periode på fem måneder etter at WIM-

punktet ble installert i 2007 (N=59.899), og sammenlignet med data fra tilsvarende periode to år tidligere (N ikke oppgitt). Resultatene fra før- og etteranalysen ble supplert med en spørreundersøkelse i transportbransjen (N=381). Evalueringen tok utgangspunkt i følgende nytteeffekter:

- *Trafikksikkerhet*: Reduksjon i konfliktpunkt ved inn- og utkjøring på kontrollstasjonen (færre kjøretøy som må svinge innom kontrollstasjonen, gir færre potensielle konfliktpunkt).
- *Effektivitet og produktivitet i transportbransjen*: Sparte transportkostnader som funksjon av tids- og drivstoffbesparelser.
- *Økt kontrollkapasitet*: Antall identifiserte lovovertrедelser, treffrate (i forhold til antall kjøretøy på vektkontroll) og sanksjonsrate (i forhold til trafikkvolum).
- *Miljø*: Gjennomsnittlig utslipp som funksjon av akselerasjon, nedbremsing og kjøretøy på tomgang. I praksis ble dette vurdert kvalitativt med utgangspunkt i antall klarerte kjøretøy som fikk passere kontrollstasjonen uhindret.
- *Brukeraksept*: Kartlegging av erfaring og aksept basert på spørreundersøkelse i transportbransjen.

Studien konkluderer med positive resultater på alle felt, i hovedsak basert på at antall kjøretøy som får passere kontrollstasjonen uhindret har økt fra 30-40 % lekkasje i førsituasjonen til 77 % klarerte kjøretøy etter at WIM-punktet ble etablert (Nova Scotia Transportation and Infrastructure Renewal, 2008). Videre forklares en nedgang i sanksjoner med at tiltaket har hatt en oppdragende effekt i transportbransjen. Dette synes imidlertid å være konkludert på særdeles tynt grunnlag, da studien er gjennomført uten noen form for kontrollgruppe – og rapporten også beskriver samtidige endringer i kontrollvirksomheten (antall ansatte og kontrolldøgn), samt opplærings- og holdningskampanjer i transportbransjen i løpet av perioden.

Kunnskapsgjennomgangen viser at ITS-tiltaket *klarering av tunge kjøretøy* kan forventes å ha en positiv effekt på kontrollvirksomhetens produktivitet og konkurranseforholdet i transportbransjen, ut ifra et logisk resonnement. Forskningsresultater gir først og fremst indikasjoner på at systemene bidrar til økt deteksjonsrate og flere klarerte kjøretøy som får passere uhindret. De studiene som finnes tar i hovedsak utgangspunkt i en eksperimentell evalueringsmetodikk, med registreringer av vektdata i en begrenset periode før og etter innføring av ITS-tiltaket. Det er flere forhold som vanskeliggjør overføringsverdien mellom ulike lokasjoner. For det første har de ulike

installasjonene varierende funksjonalitet og består av ulike typer sensorer og komponenter, som gjør det umulig å knytte effektene fra ulike studier til ett generelt ITS-tiltak for klarering av tunge kjøretøy. Den endelige effekten som oppnås, er avhengig av kvalitet, nøyaktighet og samspill mellom mange ulike komponenter i et IKT-system. Også anvendelsen av tiltaket varierer mellom forskjellige installasjoner, fra overvåking av trafikkstrøm for planleggingsformål, til screening av kontrollkandidater – og til og med direkte håndhevelse av vektreglementet. Konteksten for iverksettelsen av tiltaket vil dermed ha betydning for hvilke effekter som realiseres. Felles for studiene som presenteres, er også at de er knyttet til innovasjonsprosjekter som har til hensikt å utvikle og pilotere teknologibaserte løsninger for å styrke kontrollvirksomheten. I tillegg til at man ofte står overfor teknologisk umodne løsninger, resulterer dette også i at utvalget i studiene i liten grad er randomisert. Ofte iverksettes slike tiltak på steder hvor man har et særlig problem, eller hvor man antar at forholdene ligger godt til rette for å oppnå gode resultater. I hovedsak er studiene også gjennomført uten bruk av kontrollgrupper og uten sammenligning mot et kontrafaktisk utviklingsløp.

6.3 OPPRINNELIG EVALUERINGSPLAN I NONSTOP-PROSJEKTET

Evaluering av tiltaket *NonStop klarering av tunge kjøretøy* var en av de sentrale forskningsutfordringene i NonStop-prosjektet, og det ble tidlig utarbeidet en evalueringsmetodikk med fokus på kartlegging av effekter og vurdering av potensielle virkninger for kontrollmyndighet, transportører og samfunnet for øvrig (Lervåg et al., 2015). Evalueringen søkte å besvare to sentrale forskningsspørsmål i prosjektet:

- *Hva er virkningen av å ta i bruk NonStop-tiltaket?*
- *Hvordan bør NonStop-tiltaket utformes og implementeres for å oppnå best mulig effekt for kontrollmyndighet og transportbransjen?*

Gjennomgang av eksisterende litteratur om ITS-tiltak for klarering av tunge kjøretøy ga på dette tidspunktet ingen entydig oppskrift på evalueringsmetodikk (se avsnitt 6.2.2), og evalueringsdesignet ble utarbeidet med utgangspunkt i foreliggende rammeverk for ITS-evalueringer, bl.a. FESTA-metodikken (FOT-NET, 2014) og anbefalinger fra europeiske ekspertgrupper (EasyWay), samt tradisjonelle effektevalueringer innenfor transportsektoren (Statens vegvesen, 2011a, 2012a). Allerede i prosjektbeskrivelsen lå også en implisitt forventning fra oppdragsgiver om at evalueringen skulle bygge på et eksperimentelt design med før- og ettermålinger av effekter og en nyttekostnadsanalyse av tiltaket.

Det ble utarbeidet et eksperimentelt evalueringsdesign med målinger på to kontrollstasjoner gjennom en demonstrasjonsperiode på 12 måneder, samt landsdekkende spørreundersøkelser blant kontrollører og i transportbransjen. Evalueringen inkluderte en teknisk og funksjonell vurdering av systemet, en vurdering av effekter og nytteverdi for ulike målgrupper, samt en vurdering av brukeraksept. Den tekniske evalueringen inkluderte en innledende mulighetsstudie som grunnlag for teknologivalg i prosjektet, samt en behovsstudie blant kontrollører for å avdekke ITS-tiltakets relevans og innspill til systemets utforming. Systemet skulle valideres gjennom tekniske målinger og kalibrering av nøyaktighet og presisjon. Effektevalueringen var planlagt gjennomført med data om kjøretøy, vektstatistikk, reisetidsmålinger og kontrollhistorikk i forkant av og gjennom hele demonstratorperioden, med en påfølgende nyttekostnadsanalyse for vurdering av samfunnsøkonomisk lønnsomhet. I tillegg skulle transportbransjens erfaringer og grad av brukeraksept kartlegges og dokumenteres ved hjelp av spørreundersøkelser både i forkant og etterkant av innføringen. Et sammendrag av den opprinnelige evalueringsplanen er gjengitt i Tabell 8. De undersøkelsene som faktisk ble gjennomført er nærmere presentert i avsnitt 6.5.

Tabell 8: Opprinnelig evalueringsplan for NonStop-prosjektet

Tema	Formål	Evalueringsmetode	Datainnsamling
Funksjonell og teknisk ytelse	Bekreftede at alle funksjonelle og tekniske krav er oppfylt	Mulighetsstudie teknologi	Litteraturstudie og dokumentgjennomgang
		Behovsstudie kontrollør	Fokusgruppeintervju
		Måling av nøyaktighet og presisjon	Dokumentgjennomgang og feltstudie (kalibrering)
Effekter og nytteverdi	Bekreftede at investering og drift er lønnsomt sammenlignet med dagens kontrollvirksomhet	Før-/ettermåling av effekter for kontrollmyndighet, transportbransjen og samfunnet.	Demonstrator med datainnsamling i 12 mnd på to kontrollstasjoner med WIM/ ANPR. Statistikk fra øvrige kontrollstasjoner (kontrollgruppe)
		Nyttekostnadsanalyse	Input fra effektstudie (demonstrator) Kostnadsdata fra Statens vegvesen
Brukeraksept	Bekreftede at kontrollmyndighet og transportaktører opplever systemet som brukervennlig, nyttig, effektivt, nøyaktig, pålitelig og rettferdig	Behovsundersøkelse transportbransjen	Spørreundersøkelse sjåførere og transportselskap før innføring
		Etterundersøkelse transportbransjen	Spørreundersøkelse sjåførere og transportselskap etter innføring
		Etterundersøkelse kontrollører	Intervju og spørreundersøkelse kontrollører etter innføring

Det ble også definert en rekke effekter med tilhørende målbare indikatorer (KPI) som grunnlag for effektmålinger og nyttekostnadsanalyse i prosjektet. Tabell 9 gir oversikt over aktuelle indikatorer og hypotesene man ønsket å teste i prosjektet.

Tabell 9: Definerte indikatorer og hypoteser i NonStop-prosjektet (KPI 8 og KPI 10 var ikke forventet å gi målbare effekter i prosjektperioden)

Indikator	Beskrivelse	Hypotese
KPI 1	Antall kjøretøy med overlast/ totalt antall kontrollerte kjøt (deteksjonsrate)	<i>KPI 1 øker etter innføring av NonStop</i>
KPI 2	Antall lovlige kjøt som kontrolleres/totalt antall kontrollerte kjøt (falsk positiv)	<i>KPI 2 avtar etter innføring av NonStop</i>
KPI 3	Antall kjøt som passerer uten vurdering pga kapasitetsproblemer (lekkasje)	<i>KPI 3 avtar etter innføring av NonStop</i>
KPI 4	Andel kontrollører som opplever NonStop som nyttig verktøy i utvelgelse av kontrollkandidater (brukeraksept)	<i>KPI 4 øker etter innføring av NonStop</i>
KPI 5	Forsinkelse i kontrollområdet for kjøt som ikke skal kontrolleres	<i>KPI 5 avtar etter innføring av NonStop</i>
KPI 6	Andel førere av lovlige kjøt som opplever redusert ventetid (brukeraksept)	<i>KPI 6 > 0 etter innføring av NonStop</i>
KPI 7	Andel førere og transportselskap som aksepterer NonStop (gjennomsnittsverdi brukeraksept)	<i>KPI 7 øker etter innføring av NonStop</i>
KPI 8	Andel kontrollpliktige kjøretøy i trafikkstrøm (oppdragende effekt)	<i>KPI 8 avtar etter innføring av NonStop (på lang sikt)</i>
KPI 9	Andel førere som er villig til elektronisk deling av informasjon fra kjøt (brukeraksept)	<i>KPI 9 øker etter innføring av NonStop</i>
KPI 10	Antall kjøt med overlast involvert i relevante hendelser/ulykker	<i>KPI 10 avtar etter innføring av NonStop</i>

Følgende premisser ble lagt til grunn for å anta kausal sammenheng mellom tiltakets virkemidler og forventede effekter:

- Generell grad av lovlidighet i transportbransjen er uforandret gjennom prosjektperioden (ingen eksterne faktorer som påvirker endring i holdning og atferd hos transportørene).
- Kontrollregimet er uforandret gjennom prosjektperioden (ingen vesentlige endringer i f.eks. økonomi som påvirker kontrollvirksomheten).
- Det forekommer ikke vesentlige endringer i lovverk/forskrifter knyttet til transportvirksomheten.
- NonStop-systemet fungerer stabilt, med moden teknologi og god brukervennlighet.

6.4 ERFARINGER FRA PROSJEKTGJENNOMFØRINGEN

NonStop-prosjektet hadde som målsetting å *utvikle, implementere og evaluere* et automatisk system for målrettet utvelgelse av kjøretøy for tungbilk kontroll. Utviklingen av systemet skulle i utgangspunktet benytte allerede etablerte systemer for nummerskiltgjenkjenning, vektsensorer og eksisterende dataregistre, som i prosjektet ble koblet sammen og utvidet med et NonStop beslutningsstøttesystem, og gitt et hensiktsmessig brukergrensesnitt mot kontrollørene. Implementeringen av systemet var planlagt gjennomført på to separate kontrollstasjoner, med innsamling av relevante før-data etterfulgt av en 12 måneder lang demonstrasjonsperiode. Underveis i prosjektet støtte man på en rekke utfordringer som medførte endringer både i tiltakets funksjonalitet og den fysiske installasjonen. Dette fikk også vesentlige implikasjoner for hvordan systemet kunne evalueres. Erfaringene i prosjektet er nærmere drøftet i Foss, Lervåg og Moen (2016). I avsnittene som følger gis en oversikt over de viktigste utfordringene som også påvirket evalueringsprosessen.

Utviklingen av ITS-tiltaket ble mer ressurskrevende enn forutsatt

Da ITS-tiltaket skulle etableres, ble eksisterende ANPR- og WIM-tjenester av ulike årsaker ikke gjort tilgjengelig for prosjektet, til tross for at dette var forutsatt ved oppstart. Prosjektet måtte dermed kjøpe og installere nytt utstyr, noe som medførte både forsinkelser i tid og større ressursbehov til teknologiutvikling – på bekostning av øvrige aktiviteter i prosjektet. En konsekvens av dette var blant annet at man tidlig nedskalerte demonstratoren til én kontrollstasjon. Sandmoen kontrollstasjon som ligger på hovedinnfartsåren (E6) til Trondheim fra sør, ble valgt som testlokasjon (ÅDT = 20.000, med 14 % tungtrafikk, fartsgrense 70 km/t). Kalibrering og kvalitetssikring av WIM-systemets nøyaktighet og presisjon ble en kritisk faktor for utviklingen av ITS-tiltaket.

ITS-tiltaket oppnådde en lavere grad av funksjonalitet og teknologisk modenhet enn forutsatt

Spesifikasjonen av NonStop beslutningsstøttesystem ble utarbeidet basert på fokusgruppeintervju med aktører fra den operative kontrollvirksomheten, kontrollmyndigheter og transportbransjen. Systemet skulle gi beslutningsstøtte basert på indikasjoner på tre sentrale faktorer:

- a) Beslutningsstøtte for vekt kontroll (aksellast og totalvekt bil og henger)
- b) Beslutningsstøtte for andre forhold (begjæringer, kontrollseddel og forsikring)
- c) Beslutningsstøtte basert på historikk (indikasjoner fra tidligere kontroller)

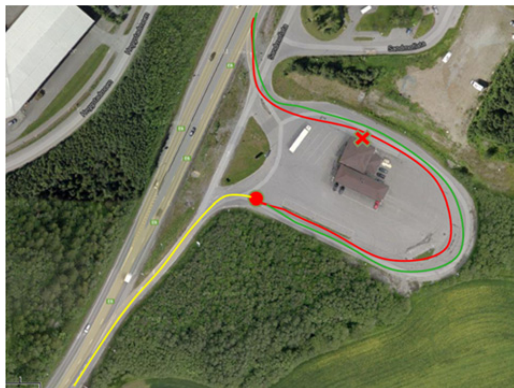
På grunn av omprioriteringer i prosjektet ble det ikke lagt vekt på å løse forskningsutfordringer knyttet til personvern eller etablere nødvendig tilgang til register for kontrollhistorikk i løpet av

prosjektperioden. Statens vegvesens nye digitale kontrollsystem (VaDIS) var under utvikling, men ble ikke lansert før januar 2015. NonStop-systemet ble dermed utviklet med funksjonalitet for a) vektkontroll og b) andre forhold, men ikke med indikasjoner basert på c) kontrollhistorikk. Systemet for nummerskiltgjenkjenning viste seg å ikke oppfylle ITS-tjenestens krav til pålitelighet og stabilitet, og man erfarte en del problemer med å korrelere data fra WIM og ANPR til riktig kjøretøy. Videre var kommunikasjonsnettlet mot vegkantstasjonen (ITS Roadside Station) tidvis ute av drift, delvis på grunn av problemer med å kombinere privat og offentlig IT-nett, men også fordi den fysiske installasjonen ble skadet (fiberledningene på testområdet ble gnagd over av mus). Dette medførte noe tap av data, men også forsinkelser i prosjektet.

Rammebetingelser for fysisk utforming av kontrollstasjonen ble ikke oppfylt

En av de viktigste effektene av NonStop-tjenesten er muligheten til å la klarerte kjøretøy passere kontrollstasjonen uhindret, slik at man unngår forsinkelser for de kjøretøyene som ikke skal kontrolleres nærmere. For å oppnå dette må kontrollstasjonen utformes slik at man i forkant informerer førerne om hvem som skal innom kontrollstasjonen, samt ha fysiske arealer som muliggjør segregering av trafikken. Det ble foreslått å gjennomføre demonstrasjonsperioden på Sandmoen kontrollstasjon med tre alternative løsninger i prioritert rekkefølge; i) variabelt skilt som informerte om spesifikk kontrollstatus 300 m før innkjøring til kontrollstasjonen, evt manuell dirigering ved innkjøringsrampen, ii) et eget forbikjøringsfelt ("fasttrack") inne på kontrollområdet, eller iii) utvelgelse ved den statiske vekten. Løsningene ble vurdert i samråd med de operative kontrollørene, og det første alternativet ble forkastet begrunnet risiko for uohensiktsmessig atferd dersom kjøretøyene ble varslet for tidlig, samt sikkerhetshensynet til kontrollørene. Løsningen med fasttrack viste seg vanskelig å gjennomføre i praksis, og det endte med at alle kjøretøy måtte kjøre en sløyfe rundt selve kontrollstasjonen, og at de eventuelt ble plukket ut av køen underveis på bakgrunn av indikasjoner fra beslutningsstøttesystemet (se Figur 14). Det betyr i realiteten at "non stop"-delen av ITS-tiltaket ikke kunne tilbys transportørene i løpet av prosjektperioden.

Figur 14: Kontrollstasjonen sett ovenfra. Den gule linjen angir avkjøring fra E6. Den røde linjen angir køen for kjøretøy som skal kontrolleres. Den grønne linjen angir "fasttrack" for kjøretøy som ikke er plukket ut til kontroll, og som får kjøre videre. Det røde krysset angir plasseringen av den statiske vekten på kontrollstasjonen.



Grensesnittet mellom teknologi og kontrollør (HMI) var teknologisk umodent

Teknologiutviklingen i prosjektet ble som allerede beskrevet mer tids- og kostnadskrevende enn forutsatt, og ressursene til utvikling av NonStop-tjenestens grensesnitt mot kontrollørene ble dermed svært begrenset. Istedenfor at kontrollørene fikk hver sin iPad med NonStop beslutningsstøttesystem, ble det utviklet en demoversjon på én iPad. Brukergrensesnittet var imidlertid ikke tilfredsstillende, og skjermen på enheten viste seg vanskelig å bruke i sol, sterkt lys eller ved mye regn. Det var ikke tid eller ressurser tilgjengelig til å videreutvikle grensesnittet slik at systemet kunne håndteres av kontrollørene i en normal arbeidssituasjon i prosjektet. Dette ble i demoperioden løst ved at NonStop-beslutningsstøttesystemet (iPad) ble håndtert av en forsøksleder som ga muntlig innspill om kjøretøyenes vektstatus til kontrolløren.

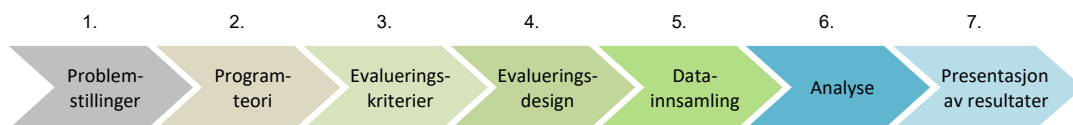
Forholdene som er beskrevet i dette avsnittet utgjorde vesentlige implikasjoner for evalueringsstudien i prosjektet. For det første ble ressurser allokert fra implementering og evaluering til teknologiutvikling, for at prosjektet i det hele tatt skulle kunne etablere et fungerende NonStop beslutningsstøttesystem. Dette medførte reduksjoner i planlagt datainnsamling, med en nedskalert demonstrator og færre brukerundersøkelser. Dessuten oppnådde NonStop-tiltaket en vesentlig lavere grad av teknologisk modenhet enn det som lå til grunn for det planlagte evalueringsdesignet. Det var derfor ikke hensiktsmessig eller mulig å gjennomføre evalueringen slik den først var planlagt, og det ble gjort tilpasninger av både omfang og metodikk. Dette er nærmere beskrevet i avsnitt 6.5.

6.5 UTVIKLING AV EN TEORIBASERT EVALUERINGSMETODIKK

Ved prosjektets oppstart var det forventet å kunne sette i drift et teknologisk modent NonStop-system med beslutningsstøtte for vektkontroll, andre forhold (begjæringer, forsikring) og kontrollhistorikk. ITS-tjenesten skulle ha et fullverdig grensesnitt mot kontrolloperatørene og et tilpasset opplegg for varsling og innhenting av kjøretøy til kontrollområdet. I løpet av utviklingsfasen ble det imidlertid klart at systemet oppnådde en vesentlig lavere grad av teknologisk modenhet, og tiltaket var å anse som en prototyp klar for begrenset testing i virkelige omgivelser. I tillegg var de tilgjengelige ressursene for implementering og evaluering betydelig redusert. Dette gjorde det nødvendig å foreta grunnleggende endringer i evalueringsplanen og en helt ny tilnærming ble valgt.

Det ble utviklet en ny evalueringsprosess basert på et *teoribasert evalueringsperspektiv*. Den viktigste begrunnelsen for valget av tilnærming var behovet for å kunne håndtere ITS-tiltaket i en tidlig fase av innovasjonsutviklingen (teknologisk umodent) og at de kausale sammenhengene

(årsakvirkningsforhold) fremsto som forholdsvis uklare. Det var behov for en evaluering som kunne fremskaffe kunnskap om tiltakets fremtidige potensiale, også i tilfeller hvor den aktuelle installasjonen ikke fungerte optimalt. Evalueringsprosessen er illustrert i Figur 15, og gjennomføringen av de syv trinnene er beskrevet i delkapitlene som følger.



Figur 15: Evalueringsprosessen i NonStop-prosjektet

6.5.1 PROBLEMSTILLINGER

Innledningsvis ble det gjennomført to undersøkelser for å kartlegge brukerbehov i relevante målgrupper; 1) fokusgruppemøte med interessenter fra vegmyndigheter, kontrolloperatører og transportindustrien, og 2) spørreundersøkelse blant ulike aktører i transportbransjen (N=1018). Dette dannet grunnlaget for å definere problemstillingene som ble lagt til grunn for evalueringsarbeidet.

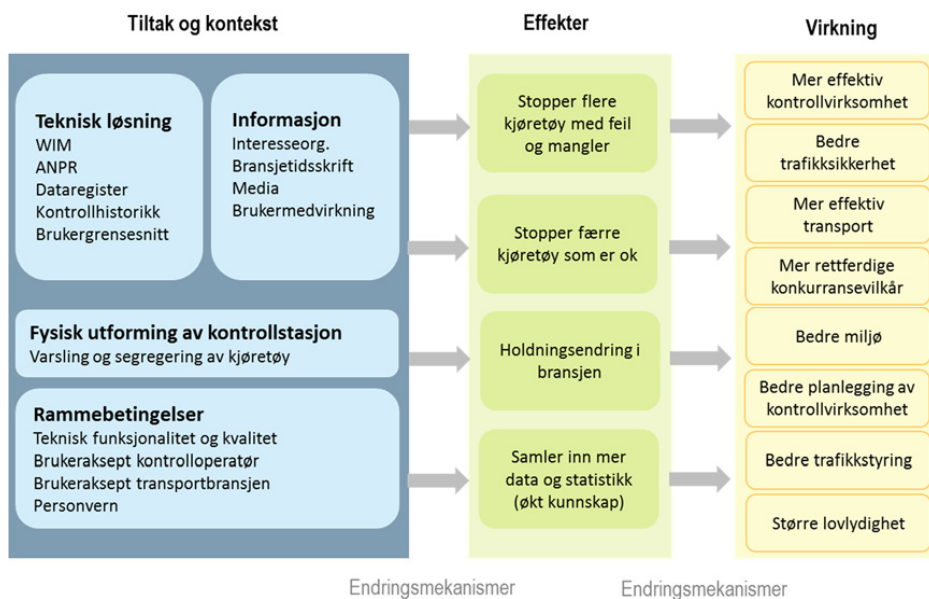
Undersøkelsene bekreftet de innledende antakelsene om at dagens kontrollvirksomhet oppleves som lite effektiv med hensyn til å sørge for god trafikkikkerhet og rettferdige konkurransevilkår i transportbransjen (Lervåg et al., 2015). Eksempelvis rapporterte 69 % av tungbilførerne at de hadde blitt kontrollert i løpet av de siste tre måneder, og så mange som 25 % oppga at de hadde blitt stoppet flere ganger på samme dag i løpet av denne perioden. Det ble oppgitt at disse kontrollene avdekket feil og mangler som resulterte i anmeldelse eller bruksforbud hos 5 % av de kontrollerte førerne. Dette impliserer at minst 95 % av kontrollaktivitetene var resultatløse og dermed ineffektiv bruk av kontrollressursene – i tillegg til at det forårsaket unødvendig venting for transportørene. Tidsbruken i forbindelse med tungbilkontroll vil naturlig nok variere med type kontroll. I Statens vegvesens utekontrollveiledning (Statens vegvesen, 2011b) anslås gjennomsnittlig tidsforbruk for en typisk tungbilkontroll å være 15 minutter, mens en enkel vektkontroll (inkludert vurdering av last og dokumenter) estimeres til 8 minutter. Tungbilførerne har anslått tidsforbruket i forbindelse med en typisk inspeksjon å være i størrelsesorden 10-20 minutter. Resultatene fra spørreundersøkelsen i transportbransjen viste en tydelig og gjennomgående oppfatning av at det foregår mye lovstridig atferd som ikke blir avdekket av dagens kontrollvirksomhet. Basert på resultatene fra de to undersøkelsene, ble følgende problemstillinger definert og lagt til grunn for evalueringsarbeidet:

- 1) Dagens kontrollvirksomhet har for dårlig effektivitet, både med hensyn til antall kjøretøy som kontrolleres og ved at virksomheten påfører transportbransjen unødvendige forsinkelser.
- 2) Det er urettferdige konkurransevilkår i transportbransjen. Lav oppdagelsesrisiko og svake sanksjoner har medført en utbredt oppfatning av at overholdelse av lover og regler gir dårlig lønnsomhet.
- 3) Tungbiltransport er forbundet med trafiksikkerhetsutfordringer knyttet til feil og mangler ved kjøretøyet eller lasten, og i enkelte tilfeller også useriøse aktører som bevisst bryter lover og regler.

6.5.2 PROGRAMTEORI

ITS-tiltakets programteori ble utarbeidet som en konseptuell modell av sammenhengen mellom tiltakets virkemidler og forventede effekter. De relevante sammenhengene ble identifisert med grunnlag i eksisterende kunnskapsgrunnlag (se kapittel 6.2) og prosjektets problemstillinger (se kapittel 6.5.1).

NonStop-tiltakets programteori er oppsummert og fremstilt visuelt i Figur 16.



Figur 16: NonStop-tiltakets programteori

I kapittel 5.2 ble programteorien beskrevet som en sammenstilling av tiltakets endringsteori og implementeringsteori. NonStop-tiltakets implementeringsteorien utgjør et logisk resonnement av hvilke betingelser som må være oppfylt for at tiltaket skal avstedkomme de endringene man ønsker å oppnå. Dette er illustrert ved de reaksjonene som skjer i selve boksene i figuren. Endringsteorien beskriver de underliggende mekanismer og kausale prosesser som må inntreffe for at man skal oppnå ønsket resultat og virkning. Dette er representert ved pilene i Figur 16, og nærmere beskrevet som *mekanismer* i Tabell 10. Tabellen beskriver hvordan tiltaket forventes å utløse en rekke mekanismer som bidrar til ønskede endringer, både i form av umiddelbare resultater og virkninger på lengre sikt.

Tabell 10: Tiltakets underliggende logikk; forventet sammenheng mellom innsats og virkninger

Tiltak og kontekst: NonStop klarering av tunge kjøretøy					
Tekniske løsninger: <i>WIM, ANPR, Dataregister, Kontrollhistorikk, Brukergrensesnitt</i>					
Fysisk utforming av kontrollstasjon: <i>Varsling og segregering av kjøretøy</i>					
Informasjonsformidling: <i>Interesseorganisasjoner, Bransjetidsskrift, Media, Brukermedvirkning</i>					
Rammebetingelse	Mekanisme	Effekt	Mekanisme	Virkning	
Teknisk funksjonalitet og kvalitet	Brukeraksept kontrolloperatør	Målrettet utvelgelse av kjøretøy på kontrollstasjon	Stopper flere kjøretøy med feil og mangler	Økt treffrate	Mer effektiv kontrollvirksomhet
				Redusert ulykkesrisiko (kjøretøy)	Bedre trafikkikkerhet
				Økt oppdagelsesrisiko	Mer rettferdige konkurransevilkår
					Større lovlidighet på sikt
	Brukeraksept transportbransjen	Målrettet utvelgelse av kjøretøy på kontrollstasjon	Stopper færre kjøretøy som er ok	Redusert vegslitasje pga overlast	Bedre miljø (inkl. vegslitasje)
				Mer tid til kjøretøy med feil og mangler	Mer effektiv kontrollvirksomhet
				Mindre forsinkelser for transportbransjen	Mer effektiv transport
				Økonomisk gevinst ved å opptre lovlidig	Mer rettferdige konkurransevilkår
	Personvern	Kontinuerlig overvåking av trafikkstrøm	Samler inn mer data om transport og kontroller	Redusert start/stopp/ tomgang på kontrollstasjon	Bedre miljø
				Redusert ulykkesrisiko (atferdsendring)	Bedre trafikkikkerhet
				Flere som kjører lovlidig gir felles spilleregler	Mer rettferdige konkurransevilkår
				Flere som kjører lovlidig gir mindre vegslitasje	Bedre miljø
			Oppdragende effekt	Større lovlidighet på sikt	
			Økt kunnskap om godstransport på veg	Bedre planlegging av kontrollvirksomhet	
				Bedre trafikkstyring	
			Forbedret kontrollvirksomhet	Større lovlidighet på sikt	

NonStop-tiltakets programteori danner grunnlaget for etterfølgende empirisk testing av de antakelser og forventninger som ligger til grunn for modellen, og den gir således en struktur for evalueringsstudien ved å identifisere sammenhenger og synliggjøre hull i kunnskapsgrunnlaget. I dette trinnet skiller evalueringsmetodikken seg vesentlig fra den opprinnelige evalueringsplanen, som var basert på en eksperimentell tilnærming til effektvurdering. Det eksperimentelle metodedesignet tar utgangspunkt i observerte effekter/virkninger, og forutsetter at tiltaket i løpet av prosjektperioden vil produsere målbare endringer. I en slik tilnærming må det kontrolleres for eksterne faktorer, samtidig som situasjonsavhengige forhold holdes utenfor analysene (f.eks. forhold knyttet til kontrollstasjonens fysiske utforming).

6.5.3 EVALUERINGSKRITERIER

Evalueringskriteriene gjenspeiler formålet med evalueringen og hvilke kriterier den observerte nytteverdien av tiltaket skal vurderes mot. I NonStop-prosjektet ble evalueringskriteriene valgt med utgangspunkt i forventninger om hva som ønskes oppnådd, basert på prosjektets målsettinger (se kapittel 6.1) og uttrykt i form av *virksomheter* i programteorien (kapittel 6.5.2). Tabell 11 gir en oversikt over definerte evalueringskriterier med tilhørende forskningsspørsmål/hypoteser i NonStop-prosjektet. Disse variablene er styrende for datainnsamlingen, og strukturerer i neste omgang evalueringens vurderinger og konklusjoner.

Tabell 11: Evalueringskriterier i NonStop-prosjektet

Evalueringskriterier	Hypoteser
<i>Mer effektiv kontrollvirksomhet</i>	Andel kontrollerte kjøretøy med feil og mangler øker Andel kontrollerte kjøretøy som er ok reduseres Færre kjøretøy slipper forbi kontrollstasjonen ukontrollert (reduert lekkasje) Kontrollsituasjonen oppleves mer effektiv av kontrollører (brugeraksept) Bedre planlegging og tilrettelegging av kontrollvirksomheten ved bruk av statistikk
<i>Mer effektiv transport</i>	Kjøretøy som er ok bruker mindre tid på kontrollstasjonen; - antall stopp reduseres - kontrollen gjennomføres raskere Kontrollvirksomheten oppleves mer effektiv av aktører i transportbransjen (brugeraksept)
<i>Bedre trafiksikkerhet</i>	Flere kjøretøy med feil og mangler kontrolleres/antall feil som avdekkes øker
<i>Mer rettferdige konkurransevilkår</i>	Økt oppdagelsesrisiko: - andel feil som avdekkes øker - andel kontrollerte kjøretøy med feil og mangler øker - færre kjøretøy "slipper forbi" kontrollstasjonen pga. kapasitetsproblemer (reduert lekkasje) Andel kontrollerte kjøretøy som er ok reduseres
<i>Økt lovlidighet på sikt</i>	Fanges ikke opp i løpet av prosjektperioden, men holdningsendringer på kort sikt kan gi indikasjoner
<i>Bedre miljø</i>	Fanges ikke opp i prosjektperioden (forventer ikke målbare miljøeffekter). Kan muligens gi indikasjoner basert på antall stopp, lastfordeling/vekt o.l.

6.5.4 EVALUERINGSDESIGN

Som tidligere beskrevet erfarte man i NonStop-prosjektet flere utfordringer som fikk implikasjoner for utviklingen av et passende evalueringsdesign. Det ble forholdsvis tidlig klart at NonStop-tjenestens tekniske funksjonalitet og kvalitet ville medføre begrensninger for metodevalget. I tillegg fikk omprioritering av prosjektressurser mellom de ulike arbeidspakkene i prosjektet innvirkning på hva som var praktisk mulig å gjennomføre av undersøkelser og demonstratoraktiviteter. I denne fasen av evalueringsprosessen ble det derfor nødvendig å prioritere evalueringsressursene. Det innebar blant annet at NonStop-demonstrasjonen ble avgrenset til ett geografisk område (Sandmoen kontrollstasjon) og at implementeringen ble nedskalert fra fullskala feltforsøk (12 mnd) til en pilotstudie. Det var imidlertid fortsatt en ambisjon om at NonStop-tiltaket skulle oppnå tilstrekkelig teknisk funksjonalitet og kvalitet til at det var mulig å gjøre en eksperimentell før- og etteranalyse av utvalgte variabler knyttet til deteksjonsrate og tidsforbruk.

Evalueringsdesignet som ble utviklet for NonStop-prosjektet er basert på *theory of change*-metodikken (beskrevet i kapittel 5.3.1), hvor programteoriens ulike bestanddeler testes med hensikt å bekrefte kausale sammenhenger. Dette innebærer at man finner de faktorene som best kan forklare sammenhengen mellom virkemidlene (tiltaket) og effekt/virkning. Dette gir grunnlag for bedre kunnskap og økt forståelse av hvordan tiltaket fungerer i den aktuelle kontekst. Med utgangspunkt i tilgjengelige evalueringsressurser og eksisterende kunnskap om sannsynlige effekter, ble det valgt å fokusere evalueringen på *effektene for transportbransjen*. Analysene av sammenhenger mellom tiltakets årsak og virkning kombinerer både kvalitative og kvantitative metoder, og evalueringsdesignet legger vekt på de observerte resultatene og beskrivelse av de handlinger og prosesser som betinger effekter i de ulike fasene. Fremgangsmåten gir svar på hvordan det planlagte NonStop-tiltaket vil utløse en ønsket endring for den aktuelle målgruppen, og kartlegger eksterne faktorer av betydning.

NonStop-prosjektets evalueringsdesign er presentert i Tabell 12.

Tabell 12: Evalueringsdesign i NonStop-prosjektet

Mer effektiv transport		Metode
Endring/resultat:	Stopper og kontrollerer færre kjøretøy som er ok	Logging av kjøretøy- og kontrolldata i demonstrator
Betingelser:	Måltrettet utvelgelse (deteksjonsrate)	Logging av kjøretøy- og kontrolldata i demonstrator
	Tilstrekkelig brukeraksept hos kontrollør	Spørreundersøkelse (kvalitativ/kvantitativ)
	Tilstrekkelig teknisk funksjonalitet og kvalitet	Eksperimentelle komparative studier
Mer rettferdige konkurransevilkår		Metode
Endring/resultat:	Holdningsendring i transportbransjen	Spørreundersøkelse (kvantitativ)
	Stopper og kontrollerer flere kjøretøy med feil og mangler	Logging av kjøretøy- og kontrolldata i demonstrator
Betingelser:	Økt oppdagelsesrisiko (økt deteksjonsrate, redusert lekkasje)	Logging av kjøretøy- og kontrolldata i demonstrator
	Mindre forsinkelser for kjøretøy som er ok	Logging av kjøretøy- og kontrolldata i demonstrator
	Tilstrekkelig brukeraksept hos transportaktører	Spørreundersøkelse (kvantitativ)
	Tilstrekkelig teknisk funksjonalitet og kvalitet	Eksperimentelle komparative studier

Evalueringen ble utformet med hensikt om å avdekke to sentrale effekter: I hvilken grad vil NonStop-tiltaket bidra til a) mer effektiv transport, og 2) mer rettferdige konkurransevilkår i transportbransjen? Den metodiske fremgangsmåten innebar i hovedsak logging av kjøretøy- og kontrolldata for kartlegging av deteksjonsrate og reisetid før og etter innføring av tiltaket, samt spørreundersøkelser for kartlegging av brukeraksept hos kontrolloperatører og i transportbransjen. Systemets tekniske funksjonalitet og kvalitet skulle testes gjennom eksperimentelle komparative studier, med vektdata fra WIM-punktet kontrollert mot vektdata fra den statiske vekten på kontrollstasjonen. Detaljer og vurderinger som ligger til grunn for datainnsamlingen er nærmere beskrevet i kapittel 6.5.5.

6.5.5 DATAINNSAMLING

I forbindelse med planlegging av demonstratoraktivitetene ble det foretatt en rekke metodiske vurderinger som ledet til anbefalinger for datainnsamlingen. Dette er gjort rede for i Lervåg et al. (2015), og oppsummert i avsnittene nedenfor.

Type kontrollsituasjon

Kontrollsituasjonen bør være representativ for normale kontrollaktiviteter, og samtidig være så lik som mulig i før- og ettersituasjonen. Generelle tungtransportkontroller³² omfatter 80 % av kontrollaktivitetene på Sandmoen, men type overtredelse som vektlegges og dermed tidsforbruk på utvelgelse, venting og gjennomføring av kontroll varierer i stor grad mellom hvert kjøretøy. Tilstandsundersøkelser³³ av tekniske og administrative forhold er i større grad standardisert, og dermed bedre egnet til sammenligning av data i en før- og ettersituasjon.

- ✓ *Det anbefales å gjennomføre datainnsamling knyttet til både tungtransportkontroll og tilstandsundersøkelser. I den utstrekning det er mulig bør ytre omstendigheter bli kartlagt (f.eks. antall kontrollører, type kontroll, tid på døgnet), slik at faktorer som kan forklare avvik i resultatene blir tatt hensyn til i evalueringen.*

Tid på døgnet

Det gjennomføres tungbilkontroller både på dagtid og på kveld/natt. Det antas at data som samles inn på ulike tider av døgnet vil være av såpass ulik art (forskjellige kontrolltyper), at det er hensiktsmessig å sammenligne samme tidsperioder i en før- og etter-situasjon. Effekten av NonStop-systemet med hensyn til å redusere forsinkelser og øke effektivitet vil sannsynligvis være størst på dagtid, når trafikkvolum og dermed kapasitetsproblemene på kontrollstasjonen er størst. Det kan imidlertid tenkes at de mest useriøse aktørene spekulerer i å passere kontrollstasjonene på kveld og natt når det forekommer færre kontroller. For den praktiske gjennomføringen av datainnsamling i prosjektet kan fravær av dagslys medføre problemer for videoregistrering og analyse av bilder fra ANPR-kameraet. Samtidig vil uttesting av løsningen under vanskelige forhold gi svar på om systemet er tilstrekkelig robust for utrulling i større skala.

- ✓ *Ut ifra hensynet til ressursforbruk anbefales at datainnsamling avgrenses til å gjelde kontroller på dagtid. Dette bør imidlertid kombineres med en mer begrenset og automatisk datainnsamling over hele døgnet i en lengre periode for statistikkformål.*

³² Tungtransportkontroll er ikke styrt mot et spesifikt fokusområde, men avhenger av kontrollørens spesialkompetanse og synlige indikasjoner på kjøretøyene som vinkes inn. En tungtransportkontroll inneholder alltid dokumentkontroll, vekt og minst en av følgende kontrolltype; farlig gods, teknisk/miljø, lastsikting, avgifter, dimensjoner/spesialtransport, hjulutrustning og/eller bilbeltebruk.

³³ Tilstandsundersøkelser er i større grad spesifisert ift. vekt, bremses, dimensjoner, dokumenter, farlig gods og/eller lastsikring. Det er krav om at hver region skal gjennomføre et visst antall tilstandsundersøkelser hvert år, og for Sandmoen omfatter dette 400 tungbiler/år.

Bruksområde

NonStop-tjenesten kan ha ulike bruksformål, avhengig av hvordan tiltaket utformes og anvendes. Det er i prosjektet definert fire aktuelle bruksområder, presentert i avsnitt 6.1.1; et *helautomatisk utvelgelsessystem* hvor kjøretøyene selekteres utelukkende basert på teknologien, et *beslutningsstøttesystem* hvor utvelgelsen styres av kontrolløren som mottar indikasjoner fra NonStop-tjenesten som et supplement til manuell utvelgelse, et *varslingssystem* hvor deteksjon av alvorlige overtredelser utløser et varsel/alarm til kontrollørene, eller kontinuerlig overvåking av trafikkstrømmen for *statistikkformål og planlegging* av fremtidig kontrollvirksomhet. Evalueringen må ta hensyn til konteksten tiltaket anvendes i, da effektene som kan forventes oppnådd vil variere med bruksområdet. Eksempelvis vil man kunne forvente umiddelbare effekter på reisetid dersom NonStop-tiltaket brukes som helautomatisk utvelgelsessystem eller beslutningsstøttesystem, mens et varslingssystem og/eller planleggingssystem i hovedsak gir effekter på treffraten av lovstridige kjøretøy. Tabell 13 presenterer fordeler (+) og ulemper (-) med å benytte de ulike bruksområdene som case i evalueringen.

Tabell 13: Fordeler og ulemper knyttet til datainnsamling for ulike bruksområder av NonStop-tiltaket

Helautomatisk utvelgelsessystem	Beslutningsstøttesystem	Varslingssystem	Planleggingssystem/ statistikkformål
(+) Viser NonStops fulle potensiale med hensyn til tidsbesparelse/ effektivitet og redusert forsinkelse.	(+) Viser effekten av en realistisk bruk av NonStop på kort sikt	(+) Datainnsamling kan foretas automatisk uten å forstyrre ordinær kontrollvirksomhet	(+) Datainnsamling kan skje automatisk uten å forstyrre ordinær kontrollvirksomhet
(+) Viser hvilke forhold som faller utenfor/ ikke fanges opp av NonStop	(+) Antatt større aksept for dette bruksområdet blant kontrollørene	(+) Viser systemets effekt mht. å fange opp ureglementerte aktører	(+) Gir nyttig informasjon til kontrollmyndighet og fagmiljø
(-) Forutsetter at kontrollører er villig til å følge NonStop-indikasjoner 100 % i datainnsamlingsperioden, også når det medfører at kjøretøy som ellers ville blitt stoppet får kjøre forbi ukontrollert	(-) Effektene som avdekkes er ikke rene effekter av NonStop, men også avhengig av kompetanse hos den enkelte kontrollør	(-) Fanger ikke opp NonStop-effekter knyttet til tidsbesparelse og redusert ventetid	(-) Fanger ikke opp NonStop-effekter knyttet til tidsbesparelse og reduksjon av ventetid

- ✓ Det anbefales å samle inn data tilknyttet alle bruksområder. Helautomatisk kontrollsystem og beslutningsstøttesystem krever at datainnsamling foregår under kontrollvirksomhet både i en før- og ettersituasjon, mens varslingssystem og planlegging/statistikk ivaretas gjennom en kontinuerlig, automatisk datainnsamling gjennom hele demonstrasjonsperioden.

Organisering av kontrollplassen

Tre alternativer for utforming av kontrollområdet ble drøftet; i) *utvelgelse og varsling av kjøretøy langs hovedvegen, slik at de som ikke skal kontrolleres får passere uhindret*, ii) *utvelgelse umiddelbart inne på kontrollstasjonen, slik at de som ikke skal kontrolleres får passere i en egen "fasttrack"*, og iii) *utvelgelse ved den statiske vekten*. Disse vurderingene er tidligere beskrevet i avsnitt 6.4. NonStop-systemets effekt med hensyn til effektivitet, tidsbesparelse og redusert forsinkelse forventes å være størst når utvelgelsen skjer på hovedvegen, eventuelt ved bruk av en dedikert fasttrack. Kontrollørene uttrykker imidlertid bekymringer knyttet til arbeidsforhold og sikkerhet ved de alternativene hvor utvelgelsen skjer i høy hastighet.

- ✓ *Ut ifra en samlet vurdering med vektlegging av hensynet til HMS, anbefales det at utvelgelsen av kontrollobjekter skjer umiddelbart etter innkjøring på kontrollplassen.*

Omfang av datainnsamling

I utgangspunktet er det ønskelig med mest mulig datainnsamling over en lengre periode, men dette må også sees i lys av tilgjengelige ressurser. Behovet for datamengde vil variere med antall usikre variabler i de ulike testsituasjonene, og av spredningen i innsamlet materiale.

- ✓ *Det anbefales at datainnsamlingen i NonStop-prosjektet har et omfang på minimum fem dager i før-situasjonen og fem dager i etter-situasjonen, samtidig med pågående kontrollvirksomhet. I tillegg anbefales kontinuerlig innsamling av WIM/ANPR-data over hele døgnet i minimum tre uker, uavhengig av kontrollaktivitetene.*

Gjennomført datainnsamling

Det ble i forbindelse med demonstratoraktivitetene og den øvrige evalueringen foretatt ytterligere justeringer av datainnsamlingen, i hovedsak begrunnet i ITS-tiltakets manglende teknologiske modenhet og tilgjengelige evalueringsressurser. Demonstratoraktivitetene ble gjennomført med følgende omfang:

- *Type kontrollsituasjon:* Generell tungtransportkontroll
- *Tid på døgnet:* Kontrolldata og reisetidsregistreringer på dagtid, vektdata (WIM) for statistikkformål over hele døgnet
- *Bruksområder:* Beslutningsstøtte/varslingssystem og statistikkformål
- *Organisering av kontrollplassen:* Utvelgelse ved statisk vekt
- *Omfang datainnsamling:* Kontroll- og reisetidsregistreringer i 8 dager, i praksis ingen etterperiode, overvåking av tungtrafikk med WIM i 42 døgn.

Tabell 14 viser en oversikt over den totale datainnsamlingen som faktisk ble gjennomført i prosjektet.

Tabell 14: Gjennomført datainnsamling i NonStop-prosjektet

Metode	Omfang	Formål
Teknisk testing og kalibrering av utstyr	Flere ganger underveis i utviklingen av NonStop-systemet	Kartlegging av teknisk funksjonalitet og kvalitet, input til forbedringer underveis
Fokusgruppeintervju	Møter med interessegrupper okt. 2012 (N=14) og april 2013 (N=13).	Gi input til utforming av NonStop beslutningsstøttesystem
Spørreundersøkelse i transportbransjen	Før-undersøkelse med 1018 respondenter fra hele landet våren 2013	Kartlegge behov og brukeraksept basert på beskrivelse av systemet
NonStop demonstrator	- WIM/ANPR-data fra ca 30 000 kjøretøy som passerte registreringspunkt over en periode på 42 døgn i juni/juli 2014 (ikke kontinuerlige registreringer). - Manuelle registreringer av ventetid på kontrollstasjonen for ca 2.900 kjøretøy i åtte dager med tungtransport-kontroll.	Grunnlag for effektanalyser. Kartlegging av teknisk funksjonalitet og kvalitet. Grunnlag for vurdering av brukeraksept hos kontrollører.
Spørreundersøkelse blant kontrollørene	Etter-undersøkelse med 17 kontrollører ved Sandmoen kontrollstasjon høsten 2015	Kartlegge behov og brukeraksept basert på erfaringer med systemet

Det ble gjennomført flere runder med teknisk testing i forbindelse med kalibrering av sensorer og utvikling av den teknologiske løsningen i prosjektet. Dette foregikk ved at vektdata fra WIM-punktet ble sammenlignet med vektdata fra den statiske vekten på Sandmoen kontrollstasjon. Endelig dokumentasjon av teknisk funksjonalitet og kvalitet på NonStop-tjenesten ble basert på målinger av 511 kjøretøy i demonstratoren (Haugen et al., 2016). Det ble gjennomført intervju med aktører fra ulike interessegrupper (bl.a. kontrollmyndighet, Norges lastebileierforbund, Statens vegvesen, systemutviklere og teknologiaktører) i to fokusgruppemøter, høsten 2012 og våren 2013. Hensikten med denne aktiviteten var å få innspill til utforming av NonStop-tjenesten, blant annet med tanke på teknologivalg, type avvik som registreres, grenseverdier for indikasjoner, organisering av kontrollstasjonen og grensesnitt mot kontrollører. Det ble gjennomført en landsdekkende spørreundersøkelse blant aktører i transportbransjen³⁴ våren 2013. Formålet med undersøkelsen var å kartlegge transportbransjens behov og interesser med hensyn til utvikling av et fremtidig kontrollsystem. Resultatene fra denne undersøkelsen dokumenterer også før-situasjonen med hensyn til erfaringer med kontrollvirksomhet, holdninger og brukeraksept. Demonstratoren ble gjennomført i juni/juli 2014 med WIM/ANPR-data fra tunge kjøretøy på E6 i retning fra sør mot

³⁴ Undersøkelsen ble gjennomført på web (epost) av medlemmer i bransjeorganisasjoner som Norsk lastebileierforbund, Maskinentreprenørenes forbund, NHO Transport, NHO Logistikk og transport og Transportarbeiderforbundet, i tillegg til ansatte hos store aktører som TINE og ASKO.

Trondheim. Det ble foretatt manuelle målinger av ventetid for alle tunge kjøretøy på Sandmoen kontrollstasjon gjennom åtte arbeidsdager med tungbilkontroll. Gjennomføringen av målingene er nærmere beskrevet i Lervåg et al. (2015). Den opprinnelige planen med registreringer i en før-situasjon uten NonStop-tjenesten og etter-situasjon med NonStop-tjenesten, måtte fravikes – da tjenesten ikke var tilstrekkelig moden (blant annet med hensyn til brukergrensesnitt og organisering av kontrollområdet) til å påvirke kontrollsituasjonen i praksis. I stedet ble det lagt vekt på registrering av reisetidsmålinger som grunnlag for å vurdere potensiell tidsbesparelse for et fremtidig fungerende NonStop-system. I praksis betyr dette at NonStop fungerte som et baksystem med dataregistrering gjennom demonstratorperioden, men indikasjoner og fortløpende resultater var i utgangspunktet kun synlig for forsøksleder og ikke umiddelbart tilgjengelig for kontrollørene³⁵. Systemet ble imidlertid tilstrekkelig demonstrert og testet til å gi grunnlag for en tidlig vurdering av brukeraksept blant kontrollørene. Dette ble gjennomført som del av en spørreundersøkelse høsten 2015, etter at analyser og resultater fra demonstratoren ble gjort kjent for de involverte kontrollørene.

Det ble ikke gjennomført en planlagt oppfølgingsstudie i transportbransjen i etterkant av implementeringen, da NonStop-tjenestens funksjonelle tilstand og demonstratorens varighet ikke var egnet til å frembringe observerbare effekter for transportørene. Det ble heller ikke gjennomført en planlagt landsdekkende behovsundersøkelse blant kontrollører (av organisatoriske årsaker i Statens vegvesen), men denne ble erstattet med fokusgruppemøter som sikret kontrollørenes medvirkning og ivaretagelse av deres interesser i utformingen av NonStop-tjenesten.

6.5.6 ANALYSER

Analyser av teknisk funksjonalitet og kvalitet

Målet med de tekniske analysene var å identifisere hvorvidt kravene til NonStop-tjenestens funksjonelle ytelse ble oppfylt. Det ble i løpet av prosjektperioden gjennomført flere tester og kalibreringer av WIM-sensorene mot den statiske vekten på kontrollstasjonen, og vektdata fra WIM syntes å ha en tilfredsstillende kvalitet under demonstratoren med en gjennomsnittlig feilmargen på 2,73 % (N=511) (Haugen et al., 2016; Lervåg et al., 2015). Teknologien knyttet til nummeregjenkjenning (ANPR) viste seg imidlertid å være beheftet med større grad av usikkerhet. I tillegg til feilregistreringer (personbiler fikk tildelt vektdata fra vogntog) og bortfall av data fra ANPR, erfarte man periodevis betydelige forsinkelser på bilder og informasjon fra NonStop

³⁵ De siste fire dagene oppholdt forsøksleder seg sammen med kontrollørene med NonStop-tjenesten tilgjengelig, og det ble tidvis gjennomført demonstrasjoner og utvekslet informasjon om indikasjoner fra NonStop-systemet. Kontrollvirksomheten ble likevel gjennomført som normalt, og reisetidsregistreringene ble ikke signifikant påvirket.

beslutningsstøttesystem til ITS-stasjonen (iPad) hos kontrolløren. Av 2600 manuelle registreringer av kjøretøy på kontrollstasjonen i løpet av demonstrasjonsperioden, oppnådde man å koble 74,1 % av disse til dataloggen fra NonStop-systemet. NonStop-tjenestens totale ytelse er avhengig av kvalitet i alle komponenter, og på grunn av systemets forholdsvis lave grad av teknologisk modenhet, ble det ikke gjennomført videre detaljerte analyser av funksjonalitet og kvalitet.

Analyser av brukeraksept hos transportbransjen

Det ble gjennomført en omfattende undersøkelse av transportbransjens behov og interesser for klarering av tunge kjøretøy (N= 1018 respondenter). Analysene avdekket en høy etterspørsel etter et effektivt og målrettet ITS-tiltak for utvelgelse av kontrollkandidater, med forventninger om at et slikt system vil få positiv effekt og oppdragende virkning på transportbransjen. Transportaktørene opplever at deres egen bransje er preget av svært urettferdige konkurransevilkår, og at det er dårlig lønnsomhet forbundet med lovlydig atferd. Det er en felles oppfatning at dagens kontrollvirksomhet er for lite effektiv med hensyn til å avdekke tilfeller av juks og useriøse aktører. Det synes imidlertid ikke å være helt samsvar mellom NonStop-tjenestens indikatorer for utvelgelse av kontrollkandidater (i hovedsak vekt), og de forholdene som oppleves som særlig problematiske for transportbransjen; som *kjøretøyenes tekniske tilstand, kabotasjevirkosomhet* og juks med *kjøre- og hviletid*. Både tungbilførerne og transportbedriftene har høye forventninger til NonStop-tjenesten og mener at en innføring på landsbasis vil gi dem positive effekter i form av mer rettferdige konkurransevilkår, reduserte forsinkelser, bedre forutsigbarhet i reisetid, økt trafiksikkerhet, høyere grad av lovlydighet og bedre arbeidsmiljø. I tillegg uttrykker en hovedvekt av førerne at de er villige til å dele data fra kjøretøyet elektronisk i bytte mot en mer effektiv kontrollsituasjon (58-75 % avhengig av type data; det er størst skepsis til deling av sjåfør-ID, og minst skepsis til å dele kjøretøydata om bremsetilstand og vekt). Utfyllende informasjon om metodikk og resultater er presentert i Lervåg et al. (2015).

Analyser av brukeraksept hos kontrollørene

Det ble gjennomført en kartlegging av behov og brukeraksept hos involverte kontrollører i etterkant av demonstratoren (N=17). Analysene avdekket et reelt behov for å effektivisere dagens kontrollvirksomhet, og en ITS-tjeneste som kan bidra til mer målrettet utvelgelse av kontrollkandidater ønskes velkommen (Lervåg et al., 2015). Kontrollørene erfarer selv at dagens kontrollvirksomhet har beskjeden effekt med hensyn til å sikre god trafiksikkerhet og like konkurransevilkår for transportindustrien. Det er et sterkt ønske om å inkludere flere datatyper i beslutningsstøttesystemet, enn det som ble implementert og demonstrert i NonStop-prosjektet. I tillegg til vekt/aksellast og forsikringer/begjæringer, er det også behov for å ta i bruk kjøre-/hviletid

og bremsetilstand som indikatorer for utvelgelse av kontrollkandidater. Kontrollørene forventer at konseptet *NonStop klarering av tunge kjøretøy* vil få jevnt over positive virkninger for deres arbeidshverdag – spesielt når det gjelder tidsbruk på kontroller og effektivitet med hensyn til å avdekke juks. Dette må oppfattes som et forholdsvis sterkt signal fra kontrollørene. Selv om utvalget er forholdsvis beskjedent (17 kontrollører), tyder både svarprosent (100 %) og høy grad av frivillige svar i fritekstfelt at dette er et tema som engasjerer kontrollørene. Alle tilbakemeldinger tyder på en høy grad av aksept blant kontrollørene, til tross for at den applikasjonen som ble testet i demonstratoren var forholdsvis umoden og med en del unøyaktighet knyttet til funksjonalitet og datakvalitet. Siden systemet ikke oppnådde et tilstrekkelig brukergrensesnitt (iht. spesifikasjonen som ble utarbeidet i prosjektet), ble det heller ikke funnet hensiktsmessig å analysere systemets brukervennlighet.

Analysen av nytteeffekter

Analysene av nytteeffekter er basert på demonstratoraktivitetene i prosjektet. Tabell 15 viser trafikk tallene som ligger til grunn for analysene (Lervåg et al., 2015). Disse dataene representerer en før-situasjon på kontrollstasjonen og gir grunnlag for input i senere nyttekostnadsanalyse.

Tabell 15: Trafikktall (tunge kjøretøy) på Sandmoen kontrollstasjon i demoperioden

Dato	Tidspenode kontroll	Antall kjøretøy på E6 i perioden³⁶	Antall kjøretøy innom kontrollstasjonen	Antall kjøretøy kontrollert
03.06.2014	08:00-15:30	985	291	21
04.06.2014	08:00-15:30	871	405	31
05.06.2014	08:00-15:30	842	385	24
06.06.2014	08:00-15:30	1098	511	18
10.06.2014	08:00-15:30	1084	388	20
11.06.2014	08:00-15:30	735	308	22
13.06.2014	08:00-15:30	976	383	27
18.06.2014	06:00-13:30	911	219	22
TOTALT		7500	2890	185
Gjennomsnitt pr dag	7,5 timer	938	361	23

³⁶ Trafikkvolumet er beregnet basert på registrerte trafikk tall i demonstratoren. Teknisk svikt av ulike årsaker medførte en gjennomsnittlig registreringstid på WIM-punktet på drøyt 5 timer pr kontrolldag, mens registreringene på kontrollstasjonen representerer hele kontrollperioden på 7,5 timer pr kontrolldag. Dette er det tatt høyde for i beregningen av trafikkvolum.

Kontrollrate: Registreringer og påfølgende beregninger viste at det i snitt passerte 938 tunge kjøretøy på E6 i løpet av en vanlig arbeidsdag med tungbilkontroll. Av disse ble 361 kjøretøy (39 %) sluset gjennom kontrollstasjonen og den statiske vekta på Sandmoen, hvorav 23 kjøretøy ble plukket ut til nærmere kontroll. Dette ga en kontrollrate på 6,4 % av alle tunge kjøretøy som var innom kontrollstasjonen, og 2,5 % av alle tunge kjøretøy som totalt passerte kontrollpunktet på E6.

Lekkasje: Alle tunge kjøretøy som passerte på E6 ved tungbilkontroll ble vinket inn på kontrollområdet, inntil kapasiteten på kontrollplassen (enten med hensyn til areal eller antall kontrollører) var oppfylt. Skiltet som varsler kontroll ble da slått av, og påfølgende kjøretøy fikk passere uforstyrret på E6. Differansen mellom antall passeringer på E6 og antall registreringer på kontrollstasjonen utgjorde 577 kjøretøy per kontrolldag. Denne lekkasjen av kjøretøy som passerte strekningen i kontrollperioden tilsvarte 62 % av tungtrafikken i perioden.

Forsinkelse for transportbransjen: På kontrollplassen ble alle kjøretøyene veid som grunnlag for utvelgelse av kandidater til videre kontroll. For de som ble sendt videre uten ytterligere inspeksjon, kunne ventetiden på kontrollområdet vært unngått med et NonStop-system som foretar utvelgelsen i forkant av kontrollstasjonen. I løpet av en typisk kontrolldag passerte i gjennomsnitt 338 kjøretøy gjennom kontrollstasjonen uten å bli ytterligere kontrollert (utover vekt). Dette er kjøretøy som opplevde unødvendig ventetid og som tok opp unødvendig kapasitet på kontrollstasjonen. Kjøretøyenes tidsforbruk på kontrollplassen er presentert i Tabell 16.

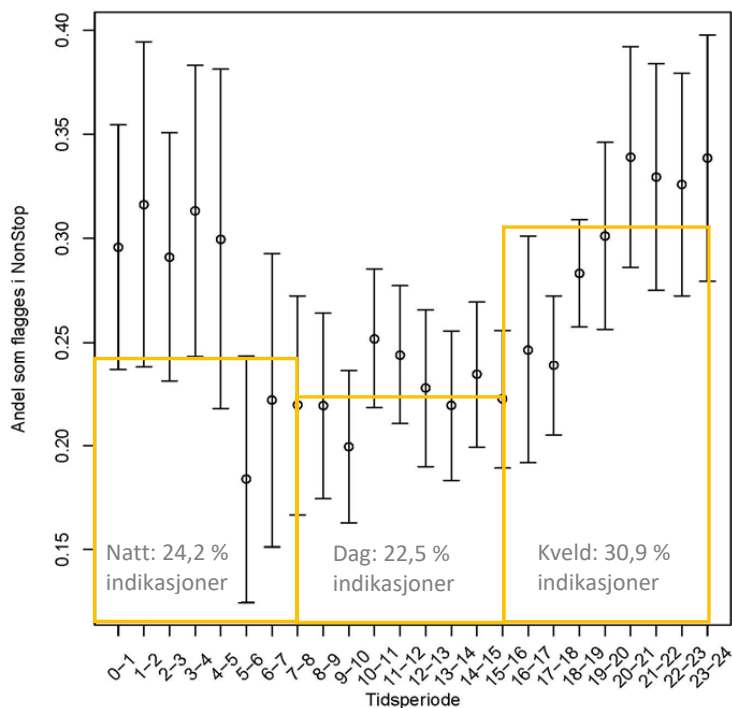
Tabell 16: Tidsforbruk ved tungbilkontroll på Sandmoen kontrollstasjon i demoperioden. Den store variasjonen i maksimal ventetid skyldes at kjøretøy som får påvist feil og mangler blir stående til disse er rettet opp (midlertidig bruksforbud).

Ventetid kontrollstasjon, fra innkjøring til utkjøring.	Gjennomsnitt	Median	Min. ventetid	Maks. ventetid
<i>Alle kjøretøy</i>	4,37 min	2,10 min	24 sek	5,31 timer
<i>De som ikke plukkes ut til kontroll</i>	2,33 min	2,00 min	24 sek	8,67 min
<i>De som faktisk kontrolleres</i>	39 min	26 min	1,57 min	5,31 timer

For kjøretøy som ble sendt videre ukontrollert, utgjorde ventetiden en unødvendig forsinkelse på i gjennomsnitt 2,33 minutt. I periodene hvor kapasiteten på kontrollstasjonen var fylt, økte ventetiden til 8-9 minutt for disse kjøretøyene. For transportbransjen representerer dette tapt reisetid.

Kunnskap om tungtransportens døgnfordeling: Det ble gjennomført en statistisk analyse av døgnfordelingen av kjøretøy som ble flagget for nærmere kontroll av NonStop-systemet. Analysen er basert på data fra omtrent 30 000 kjøretøypasseringer på E6 ved kontrollstasjonen i

demoperioden (Lervåg et al., 2015). Utgangspunktet for beregningen var antall kjøretøy med indikasjoner på overlast eller begjæring/forsikring, som andel av totalt antall passerte tunge kjøretøy per time, se Figur 17.



Figur 17: Døgnfordeling av kjøretøy med indikasjoner på overlast eller andre forhold, med tilhørende konfidensintervall. Analyser for tidsintervall lik 1 time.

Figuren viser en tendens til at en større andel av kjøretøyene som passerte på kvelds- og nattestid har feil/mangler. Selv om det er store usikkerheter i disse estimatene (lav presisjon på NonStop-dataene) er trenden relativt klar, og resultatene tyder på en større grad av ulovlig aktivitet i tidsperioden kl 20-24. En grov inndeling i tidsintervallene *dag* (kl 08:00-16:00), *kveld* (kl 16:00-24:00) og *natt* (kl 00:00-08:00), viser at andel ulovlig aktivitet var signifikant (0.05-nivå) større på kveldstid ($p=30,9\%$) enn på dag ($p=22,5\%$) og natt ($p=24,2\%$). Testene viser imidlertid ikke signifikante forskjeller mellom dag og natt, noe som sannsynligvis skyldes at datamaterialer er for mye aggregert (for lange tidsperioder). I henhold til figuren ser det ut til at det også er en høy andel ulovlige passeringer på nattestid, men at dette avtar i 05-tiden om morgenen.

Nyttekostnadsanalyser

Det ble gjennomført en ex ante nyttekostnadsanalyse av ITS-tiltaket *NonStop klarering av tunge kjøretøy* basert på nytteeffekten av *reduerte forsinkelser for transportbransjen*. Forutsetningene

som er lagt til grunn for analysen er presentert i Tabell 17. Datagrunnlaget for nyttekomponentene ble frembragt i NonStop-demonstratoren, mens kostnadselementene er basert på data fra Statens vegvesen, samt svenske erfaringer (Lervåg et al., 2015).

Tabell 17: Grunnlagsdata for nyttekostnadsanalysen

Variabel	Verdi	Kilde
<i>Felles prisnivå</i>	2014	Tidspunkt for demonstrator
<i>Levetid ITS-installasjon</i>	7 år	Forventet levetid for ITS-installasjonen
<i>Analyseperiode</i>	7 år	Tilsvarende forventet levetid på ITS-installasjonen
<i>Kalkulasjonsrente</i>	4 %	Finansdep. R-109/14: <i>Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv.</i>
<i>Skattekostnad</i>	0,20	Finansdep. R-109/14: <i>Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv.</i>
<i>Realprisutvikling yrkestrafikk</i>	1,30 %	Finansdep. R-109/14: <i>Prinsipper og krav ved utarbeidelse av samfunnsøkonomiske analyser mv.</i>
<i>Tidsavhengige driftskostnader</i>	641 kr/t	Antar 60 % lastebiler, 40 % vogntog, prognose 2014
Kontrolldata Sandmoen:		
<i>Antall kontrolldøgn per år</i>	92	Antar 75 % kontroller på dag og 25 % på natt, ref. Kontrolleder Sandmoen og arbeidslister for 2014.
<i>Antall kjøretøy³⁷ dagkontroll</i>	338	NonStop-demonstrator
<i>Antall kjøretøy nattkontroll</i>	169	Antar 50 % av kjøretøy på dag, ref. Kontrolleder Sandmoen
<i>Gjennomsnittlig ventetid</i>	2,33 min/kjt	NonStop-demonstrator
Kostnader:		
<i>Investeringskostnader</i>	1.300.000 kr	Statens vegvesen/svenske erfaringer
<i>Driftskostnade per år</i>	150.000 kr	Statens vegvesen/svenske erfaringer

Nytteverdien i analysen er avgrenset til *sparte tidskostnader for yrkestrafikk*. Andre potensielle nytteeffekter knyttet til f.eks. effektivisering av kontrollvirksomheten, trafikkikkerhet, rettferdige konkurransevilkår og redusert slitasje på vegnettet er holdt utenfor beregningene, og må eventuelt anses som tilleggsnytte. Beregningene tar ikke hensyn til trafikkutvikling i perioden, da det antas at det i dag er kapasiteten på kontrollstasjonen og ikke trafikkvolumet som er dimensjonerende for antall kontrollerte kjøretøy per år.

Nyttekostnadsberegningene av NonStop-tiltaket er presentert i Tabell 18.

³⁷ Med *antall kjøretøy* menes kjøretøy som sluses gjennom kontrollstasjonen uten å kontrolleres ytterligere.

Tabell 18: Nyttekostnadsanalyse for NonStop klarering av tunge kjøretøy, effekter for transportbransjen

Aktører	Komponenter	Endring (1.000 kr diskontert)
Transportbransjen Det offentlige	Nytte av spart reisetid	5.625
	Investering	-1.300
Samfunnet for øvrig	Drift og vedlikehold	-1.210
	Restverdi	0
	Skattekostnad	-260
SUM	Nytte	5.625
	Kostnad (budsjettkrone)	-2.770
	Netto nytte	2.855
	Netto nytte/budsjettkrone	1,1

Beregningene viser at det er god samfunnsøkonomisk lønnsomhet knyttet til innføring av en ITS-tjeneste for klarering av tunge kjøretøy. Etablering av en teknologisk moden NonStop-tjeneste på Sandmoen kontrollstasjon er beregnet å gi netto nytte (overskudd etter at kostnader er dekt inn) på kr 2 855 000,- i form av spart reisetid for transportbransjen i løpet av systemets levetid. Netto nytte er et uttrykk for hva samfunnet får igjen for investeringen, og prosjektet er lønnsomt når netto nytte > 0. Hvis netto nytte er negativ, har tiltaket en avkastning som er lavere enn kalkulasjonsrenten (Statens vegvesen, 2012a). Sett i forhold til investeringen, gir NonStop-tiltaket en netto nytte per budsjettkrone (NNB) lik 1,1. I praksis betyr dette at samfunnsnyten knyttet til sparte tidskostnader for transportbransjen alene er mer enn dobbelt så stor som kostnadene ved innføring og drift av systemet (110 %) ³⁸. I tillegg kommer andre nyttekomponenter som trafiksikkerhet, miljø og mer rettfærdige konkurransevilkår.

Usikkerhet og følsomhet i nyttekostnadsanalysen

Det ble gjort en kvalitativ vurdering av faktorer som bidro til usikkerhet i evalueringen, og følgende forutsetninger ble trukket frem som viktige for å kunne realisere de potensielle virkningene av et ITS-tiltak for klarering av tunge kjøretøy:

- Teknologien og samspillet mellom de ulike systemkomponentene må fungere godt for at systemet skal oppnå tilstrekkelig funksjonalitet og kvalitet.
- Beslutningsstøttesystemet må oppnå god brukeraksept blant kontrollørene, slik at det faktisk blir tatt i bruk.
- Systemets levetid ble satt til 7 år i nyttekostnadsanalysen; kortere eller lengre levetid vil påvirke lønnsomheten

³⁸ Tidligere var det vanlig å operere med nytteverdi per kostnad (N/K). N/K-verdien i denne beregningen er 2,03.

- Kontrollstasjonen må fysisk utformes slik at utvelgelsen skjer i forkant, og slik at kjøretøy som ikke skal kontrolleres kan passere området uforstyrret.
- Eventuelle nyttevirksomheter for trafiksikkerhet, kontrollmyndighet og miljø kommer i tillegg til beregnet nytte for transportbransjen.
- Kostnadssiden i analysene er avgrenset til investering og drift av en ferdig utviklet NonStop-tjeneste. Utviklingskostnaden antas å utgjøre en veldig liten del av kostnadssiden når den skal fordeles på et uvisst antall fremtidige kontrollstasjoner.
- Grunnlagsdataene i analysen er beheftet med en viss grad av usikkerhet, og ved tvil om verdisetting av variabler er det gjennom hele analysen forsøkt å gjøre konservative valg for å unngå at evalueringen gir urealistisk høye forventninger til systemet.

Det ble videre gjennomført følsomhetsanalyser, ved å variere verdien av de ulike nytte- og kostnadselementene (hhv. spart reisetid, antall kontrollerte kjøretøy, antall kontroll dager per år, investeringskostnad og driftskostnad). En detaljert oversikt over hvordan usikkerhet i estimeringen av de ulike variablene påvirker NNB-verdien (netto nytte per budsjettkrone) er presentert i Lervåg et al. (2015), som konkluderer med at lønnsomheten i prosjektet er forholdsvis robust. Også når verdiene til enkeltkomponentene i analysen varieres, er netto nytteverdi gjennomgående positiv, hvilket tilsier at prosjektet er lønnsomt i et samfunnsøkonomisk perspektiv.

6.5.7 RESULTATER

Resultatene fra analysene underbygger en rekke av forventningene og betingelsene som ble lagt til grunn i utviklingen av NonStop-tiltakets programteori. Tabell 19 viser hvilke sammenhenger som er studert i prosjektet. Grønne felt indikerer at evalueringsstudien har gitt resultater som styrker den aktuelle sammenhengen, mens gult felt viser at man har fått resultater som indikerer en sammenheng, uten at dette er målt direkte. Rødt felt (gjelder kun faktoren *teknisk funksjonalitet og kvalitet*) viser at ITS-tiltaket ikke oppfylte denne betingelsen, mens de forholdene som ikke ble behandlet i analysene er krysset ut. Programteorien har en logisk oppbygning, med årsaksvirkningsskjeder fra venstre mot høyre i tabellen. Realisering av virkningene i siste kolonne, forutsetter at rammebetingelser og mekanismer i tidligere kolonner blir oppfylt/bekreftet. I tilfellet NonStop, lykkes man ikke med å oppfylle betingelsen om teknisk funksjonalitet i første kolonne. Dermed var det heller ikke mulig å oppnå påvisbare effekter. Likevel har evalueringen frembragt kunnskap om sammenhengen mellom øvrige faktorer, slik det fremgår i tabellen.

Tabell 19: Oversikt over hvilke deler av programteorien som ble adressert i evalueringsstudien. Grønne felt viser at studien har gitt resultater som styrker sammenhengen, gule felt indikerer en sammenheng uten at dette er målt direkte, mens rødt viser at betingelsen ikke var oppfylt i studien. Hvite felt (kryss) viser de forholdene som ikke er behandlet i evalueringsstudien.

Tiltak og kontekst: NonStop klarering av tunge kjøretøy					
Tekniske løsninger: WIM, ANPR, Dataregister, Kontrollhistorikk, Brukergrensesnitt					
Fysisk utforming av kontrollstasjon: Varsling og segregering av kjøretøy					
Informasjonsformidling: Interesseorganisasjoner, Bransjetidsskrift, Media, Brukermedvirkning					
Ramme-betingelse	Mekanisme	Effekt	Mekanisme	Virkning	
Teknisk funksjonalitet og kvalitet	Brukeraksept kontrolloperatør	Målrettet utvelgelse av kjøretøy på kontrollstasjon	Økt treffrate	Mer effektiv kontrollvirksomhet	
			Redusert ulykkesrisiko (kjøretøy)	Bedre trafiksikkerhet	
			Økt oppdagelsesrisiko	Mer rettferdige konkurransevilkår	
			Redusert vegslitasje pga overlast	Større loyldighet på sikt Bedre miljø (inkl. vegslitasje)	
		Målrettet utvelgelse av kjøretøy på kontrollstasjon	Stopper færre kjøretøy som er ok	Mer tid til kjøretøy med feil og mangler	Mer effektiv kontrollvirksomhet
				Mindre forsinkelser for transportbransjen	Mer effektiv transport
				Økonomisk gevinst ved å opptre loyldig	Mer rettferdige konkurransevilkår
				Redusert start/stopp/tomgang på kontrollstasjon	Bedre miljø
	Brukeraksept transportaktører	Økt oppdagelsesrisiko	Holdningsendring i transportbransjen	Redusert ulykkesrisiko (atferdsendring)	Bedre trafiksikkerhet
				Flere som kjører loyldig gir felles spilleregler	Mer rettferdige konkurransevilkår
				Flere som kjører loyldig gir mindre vegslitasje	Bedre miljø
				Oppdragende effekt	Større loyldighet på sikt
	Personvern	Kontinuerlig overvåking av trafikkstrøm	Samler inn mer data om transport og kontroller	Økt kunnskap om godstransport på veg	Bedre planlegging av kontrollvirksomhet
				Forbedret kontrollvirksomhet	Bedre trafikkstyring Større loyldighet på sikt

I tråd med prioriteringene underveis i prosjektet, la evalueringen vekt på å avdekke i hvilken grad innføring av ITS-tiltaket klarering av tunge kjøretøy ville gi positive nytteeffekter for transportbransjen, i form av reduserte forsinkelser (ventetid) på kontrollstasjonen og dermed en økonomisk gevinst av å opptre loyldig. Dette vil på lengre sikt bidra til mer effektive transportoppdrag og rettferdige konkurransevilkår i bransjen. Resultatene tyder også på at kontrollvirksomheten effektiviseres, med forbehold om at det i denne studien ikke ble gjort nærmere analyser av kontrollstatistikken utover vektdata. Realisering av disse effektene forutsetter imidlertid at ITS-tjenesten oppnår tilstrekkelig god teknisk funksjonalitet og kvalitet. Den oppdaterte programteorien (Tabell 19) gir et godt grunnlag for å identifisere kunnskapsgap

(hvite/gule felt), og viser hvor det er behov for videre forskning for å oppnå kunnskap om de fullstendige virkningene av å ta i bruk ITS-tiltaket.

Tabell 20 gir en kort oppsummering av de faktiske resultatene fra evalueringsstudien.

Tabell 20: Oppsummering av evalueringsresultater

Forventet resultat (effekt)	I hvilken grad ble resultatet bekreftet?
<i>Stopper og kontrollerer flere kjøretøy med feil og mangler</i>	Det ble ikke gjort egne målinger av deteksjonsraten eller antall kjøretøy med feil og mangler, da systemets tekniske funksjonalitet ble vurdert å være utilstrekkelig. Analysene viser imidlertid et stort potensial for å oppnå en mer målrettet utvelgelse; i dag passerer 62 % av kjøretøyene kontrollområdet når kapasiteten allerede er fylt (lekkasje), og kontrollraten for alle kjøretøy som passerer er kun 2,5 %. Dersom etatens egne tidsestimater legges til grunn (SVV, 2011), vil 8 min besparelse per vektkontroll av klarerte kjøretøy gi en beregnet tidsgevinst på 45 timer i løpet av ett år på Sandmoen kontrollstasjon – altså 6 hele kontroll dager á 7,5 time. Kontrollørene uttrykker selv tro på at NonStop-tjenesten vil øke deteksjonsraten (brukeraksept).
<i>Stopper og kontrollerer færre kjøretøy som er ok</i>	Analysene viser stort potensiale for å redusere antall kontroller av kjøretøy som har utstyr og papirer i orden. I dag sendes 94 % av alle kjøretøyene som er innom kontrollstasjonen videre, uten ytterligere kontroll enn vekt. Dette tilsvarer 36 % av alle tunge kjøretøy som passerer på E6 i kontrollperioden. I snitt opplever disse en forsinkelse på 2,33 min for hver stans. Transportbransjen uttrykte en sterk etterspørsel etter et mer målrettet system for klarering av tungbiler, og analyser av uttrykte forventninger indikerer høy brukeraksept i denne målgruppen.
<i>Holdningsendring i transportbransjen</i>	Det ble ikke gjort spesifikke målinger av holdningsendringer i transportbransjen.
<i>Samler inn mer data og statistikk om transport og kontroller</i>	Analysen av tungtransportstrømmen over en lengre periode viser at det er overvekt av kjøretøy med indikasjoner på feil/mangler på kvelds- og nattetid. Dette er relevant informasjon for planlegging av fremtidig kontrollvirksomhet. Videre fremskaffet selve datainnsamlingen (telling på kontrollstasjonen) kunnskap om kontrollrate og antall passeringer som var ukjent for kontrolloperatører/vegmyndighet. Undersøkelse av brukeraksept viste at transportaktørene er villig til automatisk deling av data i bytte mot mer effektive kontroller.

6.6 WHAT IF? POTENSIELLE RESULTATER FRA EN EKSPERIMENTELL EVALUERING

Dersom man i denne evalueringen hadde gått videre med det opprinnelig planlagte eksperimentelle evalueringsdesignet med før- og etterundersøkelser, er det to mulige utfall; evalueringen ville blitt stanset, eller den ville konkludert på usikkert datagrunnlag.

Utfall 1: Evalueringen ville blitt stanset

Evalueringen ville blitt stanset etter datainnsamlingen i demonstratoren, fordi det ikke var tilrådelig å gjennomføre de planlagte analysene med foreliggende utvalgsstørrelse og datakvalitet. Som

tidligere beskrevet (i kapittel 6.5.5) ble datainnsamlingen begrenset blant annet av brukergrensesnittet på ITS-tjenesten og den fysiske organiseringen av kontrollområdet. En videreføring av et evalueringdesign med før- og etterregistreringer hadde resultert i bortkastet bruk av prosjektets tid og evalueringsressurser, og ville ikke frembragt ny kunnskap på feltet. Det anbefalte utfallet ville altså vært å avbryte evalueringstudien på dette tidspunktet, begrunnet i ITS-tjenestens manglende teknologiske modenhet.

Utfall 2: Evalueringen ville konkludert på usikkert datagrunnlag

Dersom man likevel hadde gått videre med den eksperimentelle evalueringen, kunne en tillemping og justering av datamaterialet muliggjort analysene. I løpet av demoperioden var NonStop-tjenesten tilgjengelig hos forsøksleder hele tiden, mens kontrollørene ble gjort oppmerksom på indikasjoner fra NonStop-systemet i perioder. I praksis kunne altså datainnsamlingen bli delt i en form for før- og etterundersøkelse (med og uten NonStop-tjeneste), med den begrensede tekniske funksjonaliteten som ITS-tjenesten hadde oppnådd på dette tidspunktet. Datamaterialet ble testet som et ledd i kvalitetssikringen av senere analyser, for å avdekke hvorvidt resultatene var blitt påvirket av forsøksleders tilstedeværelse og demonstrasjon av NonStop-tjenesten. Disse analysene viste at det ikke ble målt signifikant forskjell i ventetid i de periodene hvor NonStop-tjenesten ble demonstrert av forsøksleder, se Tabell 21.

Tabell 21: Gjennomsnittlig ventetid i perioder med og uten demo av NonStop-systemet

Gjennomsnittlig ventetid på kontrollstasjonen	NonStop-demo (N=1411)	Uten NonStop (N=1224)	Sign. (0,05-nivå)
<i>Alle kjøretøy</i>	4,31 min	4,43 min	Nei (0,83)
<i>De som ikke plukkes ut til kontroll</i>	2,30 min	3,36 min	Nei (0,33)

Dersom dette datagrunnlaget hadde blitt benyttet i en eksperimentell forsøksdesign ville analysene gitt grunnlag for å konkludere at tiltaket ikke har noen påvisbar effekt eller virkning, hverken for kontrollvirksomheten eller transportbransjen. Selv om dette er gyldig for tiltaket i sin nåværende (umodne) form, ville det både resultert i misvisende kunnskap og bortkastede evalueringsressurser. Den teoribaserte evalueringstudien som faktisk ble gjennomført, viste at NonStop-tiltaket har stort potensiale for å redusere forsinkelser for transportbransjen og effektivisere kontrollvirksomheten – og at det er høy samfunnsøkonomisk lønnsomhet forbundet med innføring av systemet. Det forutsetter imidlertid at man greier å oppnå tilfredsstillende teknisk funksjonalitet og kvalitet i utviklingen av ITS-tjenesten.

6.7 KONKLUSJONER FRA EVALUERINGEN

Evalueringen har avdekket at NonStop-tjenesten som konsept har et stort potensial med hensyn til å oppnå samfunnsøkonomisk lønnsomhet. Allerede fra første implementering (Sandmoen kontrollstasjon) vil den forventede nyttegevinsten i form av *reduerte forsinkelser for transportbransjen* alene være dobbelt så stor som kostnadene ved å ta systemet i bruk. Evalueringen har også demonstrert nytteeffekter i form av *effektivitet for kontrollvirksomheten*, selv om dette ikke er prissatt eller inkludert i de samfunnsøkonomiske beregningene. Når det gjelder effekter for trafiksikkerhet og miljø, vil dette avhenge av systemets endelige funksjonalitet, kvalitet og bruksområde. Evalueringen har også vist at systemet per i dag er teknologisk umodent og at det i løpet av prosjektperioden ikke var mulig å oppnå en tilfredsstillende kvalitet for implementering i større skala.

Tabell 22 gir en oversikt over konklusjonene fra evalueringsarbeidet, relatert til evalueringskriteriene som ble definert i kapittel 6.5.3.

Tabell 22: Konklusjoner fra evalueringsarbeidet

Evalueringskriterier og konklusjoner	
Mer effektiv kontrollvirksomhet	<i>Evalueringen gir grunn til å tro at NonStop-tiltaket vil effektivisere kontrollvirksomheten, både med hensyn til at oppmerksomhet og ressurser flyttes fra kjøretøy som er i orden til kjøretøy med feil/mangler, og at lekkasjen av kjøretøy som passerer uregistrert i kontrollperioden reduseres. Et grovt estimat anslår at NonStop kan frigjøre kontrolltid tilsvarende seks kontroll dager i løpet av ett år på Sandmoen kontrollstasjon, gitt dagens kontrollnivå. I tillegg vil NonStop-systemet fremskaffe statistikk som øker kunnskapen om transportvirksomheten – slik at fremtidig kontrollvirksomhet kan målrettes mot steder og tidspunkt med høy grad av ulovlig aktivitet. Det er imidlertid ikke gjort målinger på treffrate for andre forhold enn vektdata, men målrettet kontrollvirksomhet, sammen høy brukeraksept blant kontrollører, gir grunn til å anta at tiltaket på sikt vil bidra til mer effektiv kontrollvirksomhet.</i>
Mer effektiv transport	<i>Evalueringen har avdekket at NonStop-tiltaket kan gi en vesentlig tidsbesparelse for transportbransjen. Dersom utvelgelsen av kjøretøy skjer i forkant av kontrollstasjonen representerer dette en gjennomsnittlig besparelse på 2,33 min per kjøretøy som klareres. Med dagens kontrollomfang utgjør dette totalt 1100 timer/år for Sandmoen kontrollstasjon alene, hvilket utgjør en besparelse på kr 705 000,- pr år (2014-nivå) for transportbransjen.</i>
Bedre trafiksikkerhet	<i>Trafikksikkerhetseffekten vil avhenge av NonStop-tjenestens teknologiske modenhet og hvilken grad av presisjon som oppnås med hensyn til å plukke ut kjøretøy med feil og mangler. Dette forholdet ble ikke inkludert i evalueringen.</i>
Mer rettferdige konkurransevilkår	<i>Evalueringen gir indikasjoner på at NonStop-tjenesten vil bidra til mer rettferdige konkurransevilkår, gjennom tidsbesparelse for de kjøretøyene som opererer iht. lovverket. Kunnskapsgrunnlag som målsetter kontrollvirksomheten og det at færre kjøretøy slippes forbi kontrollstasjonen i perioder med kapasitetsproblemer (reduert lekkasje) vil også bidra til økt oppdagelsesrisiko og større grad av rettferdighet. En reell effekt vil imidlertid være avhengig av at systemet oppnår høy teknologisk modenhet og god presisjon, slik at treffsikkerheten ved utvelgelse av kontrollkandidater øker.</i>
Økt lovlydighet på sikt	<i>Dette forholdet ble ikke inkludert i evalueringen, da det ikke kunne forventes målbare effekter i løpet av prosjektperioden.</i>
Bedre miljø	<i>Dette forholdet ble ikke inkludert i evalueringen, da det ikke kunne forventes målbare effekter i løpet av prosjektperioden.</i>

6.8 NY KUNNSKAP OM ITS-TILTAKET

Evalueringen av NonStop-tjenesten har avstedkommet ny kunnskap om ITS-tiltaket *NonStop klarering av tunge kjøretøy*. ITS-tiltakets teknologiske modenhet er åpenbart en viktig forutsetning for å oppnå ønskede effekter i transportsystemet; hvis systemet ikke virker (godt nok) gir det heller ingen effekt å ta det i bruk. Dette gjelder alle systemets komponenter, fra de ulike sensorer via baksystem til grensesnittet (HMI) mot systemets brukere. I NonStop-prosjektet erfarte man at systemet som ble utviklet ikke oppfylte kravene til teknisk funksjonalitet og modenhet, og det var dermed ikke mulig å måle faktiske effekter i en reell implementasjon. Evalueringen har likevel bidratt til å fremskaffe kunnskap som er nyttig for fremtidig innføring av ITS-tiltaket.

Effekter

Evalueringen har dokumentert et betydelig uforløst potensial for både transportindustri og kontrollmyndigheter, gjennom datainnsamling på Sandmoen kontrollstasjon og overvåking av trafikkstrømmen som passerer kontrollstasjonen. Dagens kontrollvirksomhet påfører lovlydige transportaktører unødvendig ventetid, samtidig som den ikke greier å forhindre vesentlig lekkasje av kjøretøy som burde vært nærmere kontrollert. Analysene har bekreftet innledende hypoteser og avdekket størrelsesorden på disse variablene. Innføring av tiltaket *NonStop klarering av tunge kjøretøy* forventes altså å øke produktiviteten i transportbransjen gjennom å redusere forsinkelser på kontrollstasjonen, og samtidig bidra til mer rettferdige konkurransevilkår gjennom økt oppdagelsesrisiko for transportaktører som bryter lover og regler for å øke egen inntjening. Kontrollvirksomheten effektiviseres gjennom mer målrettet utvelgelse av kjøretøy og et bedre kunnskapsgrunnlag for fremtidig planlegging av kontroller.

Rammebetingelser og kontekst

Evalueringen har identifisert noen eksterne faktorer som har stor betydning for ITS-tiltakets evne til å utløse potensielle effekter. Først og fremst har den fysiske utformingen av kontrollstasjonen direkte innvirkning på effektene for transportbransjen. Dersom tiltaket skal gi redusert reisetid for kjøretøy som opptrer lovlydig, må klarerte kjøretøy kunne passere kontrollstasjonen uten nevneverdige forstyrrelser. Dette innebærer at varsling og innhenting av de kjøretøyene som faktisk skal kontrolleres foregår uten at øvrige kjøretøy må svinge innom kontrollstasjon. Størrelsen på trafikkvolumet som kontrolleres er også en situasjonsavhengig faktor av vesentlig betydning for nytteverdien av ITS-tiltaket. Jo mer trafikk som passerer kontrollpunktet, jo større blir den totale nytteverdien av spart reisetid for transportbransjen. I nyttekostnadsanalysen ble det lagt til grunn daglig klarering av 338 kjøretøy, mens følsomhetsanalysene viste at break-even-punktet for lønnsomhet på Sandmoen kontrollstasjon er på omtrent 180 klarerte kjøretøy per kontrolldag (Lervåg et al., 2015). Dette nivået vil variere avhengig av ventetid og kontrollomfanget på hver enkelt kontrollstasjon.

Relevans

Evalueringen har bekreftet antakelsen som lå til grunn for prosjektet om at NonStop-tiltaket har høy relevans for samfunnet. Spørreundersøkelsene blant transportører og kontrollører identifiserte et tydelig behov for å effektivisere kontrollvirksomheten og motvirke urettferdige konkurransevilkår i transportbransjen. Begge interessegruppene uttrykte en tydelig forventning om at innføring av NonStop-tjenesten vil gi et viktig bidrag i dette arbeidet. Nyttetekostnadsanalysen indikerer at det også er god samfunnsøkonomisk lønnsomhet for innføring av ITS-tiltaket.

Brukeraksept

De ulike interessegruppene (kontrollmyndighet og transportbransjen) har samsvarende interesser og et sterkt felles ønske om at NonStop-konseptet videreføres og tas i bruk i større skala. Dette er et forholdsvis sterkt signal, da man kunne forventet at tiltak som øker overvåkning og kontroll ville møtt motstand i transportbransjen, og tilsvarende at kontrollørene ville hatt en viss vegring mot teknologisystemer som endrer deres arbeidshverdag. Til tross for at selve NonStop-tjenesten i dette tilfellet var teknologisk umoden, ble det likevel dokumentert høy brukeraksept for *konseptet* NonStop klarering av tunge kjøretøy.

Funksjonalitet

Evalueringen har avdekket et avvik mellom funksjonaliteten som ble demonstrert i NonStop-prosjektet og den funksjonaliteten som etterspørres av kontrollører og transportaktører. I tillegg til indikasjoner på vekt og begjæringer/forsikringer, ønskes også indikasjoner på mer kompliserte forhold som kjøre-/hviletid, kabotasje, dekkutrustning og bremsetilstand. Tungbilførerne uttrykte høy aksept for elektronisk deling av relevante data fra kjøretøyet, noe som kan erstatte den delen av NonStop-systemet som viste seg sårbart med hensyn til teknisk kvalitet (kamera/sensorer langs vegen). Dersom det først etableres infrastruktur for elektronisk deling av kjøretøydata, er det imidlertid mulig å utvide funksjonaliteten i systemet etter hvert.

Kunnskapsgap

Prosjektet har avdekket hull i kunnskapsgrunnlaget, som ikke er adressert i denne evalueringen. Dette gjelder først og fremst målinger av konkrete effekter for trafikksikkerhet og miljø. Selv om man har etablert logiske årsaksvirkningskjeder for sammenhengen mellom for eksempel kjøretøy med overlast og trafikksikkerhet, finnes det lite data om ulykkesrisiko og skadegrad i forbindelse med ulykker og hendelser som involverer denne trafikkgruppen. Når det gjelder transportbransjens behov og holdninger i forbindelse med kontrollvirksomheten, har denne evalueringen etablert et godt og landsdekkende kunnskapsgrunnlag som dokumenterer dagens situasjon og forventninger til et fremtidig NonStop-tiltak. Dette kan benyttes som før-data i en senere komparativ studie av holdningsendringer og brukeraksept etter innføring av tiltaket.

6.9 NY KUNNSKAP OM EVALUERINGSMETODIKK

Case-studiet av *NonStop klarering av tunge kjøretøy* har gitt nyttige erfaringer med evaluering av ITS-tiltak. Underveis i prosjektet ble det foretatt en kursendring fra et planlagt eksperimentelt evalueringsdesign til en teoribasert evalueringsprosess. Dette viste seg å være et fornuftig grep – da

evalueringen bidro til nyttig og anvendbar kunnskap om ITS-tiltaket, til tross for at man ikke lyktes i å utvikle en teknologisk moden NonStop-tjeneste i løpet av prosjektperioden.

Kunnskapsbygging når teknologien er umoden

At utvikling og iverksetting av ITS-tiltaket viser seg å være mer tid- og ressurskrevende enn forutsatt, er ikke unikt for dette prosjektet. Flere tidligere studier har også erfart at det er krevende å få de ulike komponentene til å fungere godt sammen, og kalibrering av sensorer for å oppnå tilstrekkelig presisjon er en kostbar øvelse (Clough Harbour & Associates, 2012; Nova Scotia Transportation and Infrastructure Renewal, 2008; Stanczyk & Klein, 2012; Zhang et al., 2008).

Innenfor ITS-faget jobber man med avanserte og komplekse løsninger, som i realiteten består av en rekke komponenter (subsystemer) og virkemidler som til sammen utgjør det endelige tiltaket. Når man også beveger seg i front av utviklingen, har man begrenset erfaringsgrunnlag, og utviklingen bærer ofte preg av *learning by doing*. Dette viser at det er behov for en evalueringsmetodikk som kan håndtere ulike grader av teknologisk modenhet. Dette synes i liten grad å være drøftet i tidligere studier av ITS-tjenester for klarering av tunge kjøretøy, men de metodiske utfordringene kan være en forklaring på at det foreligger lite publisert forskning på virkninger av tiltaket. Når kravene til teknisk funksjonalitet ikke blir oppfylt, mangler også grunnlaget for å oppnå målbare effekter i trafikksystemet. Evalueringen av NonStop-tiltaket lykkes likevel å frembringe kunnskap om potensielle effekter og virkninger, samt om suksesskriterier og barrierer som må tas hensyn til i en fremtidig implementering.

Bruk av programteori

Utarbeidelse av programteorien som en visuell modell av konseptet for ITS-tiltaket, sørget for en ryddig struktur og avgrensning for evalueringsarbeidet. Selve modellen bidro til å synliggjøre relevante endringsmekanismer som ligger til grunn for effektene av tiltaket, i tillegg til å identifisere eksterne og situasjonsspesifikke faktorer av betydning. Dette gir mulighet til å forklare hva som fungerer (eller ikke) og hvilke faktorer som kan forsterke eller hindre en ønsket effekt.

Betydningen av eksterne faktorer

Evalueringen viste at effektene av ITS-tiltaket er svært avhengig av kontekst; det handler ikke bare om hva slags tiltak man iverksetter, men også om *hvor* og *hvordan man implementerer* tiltaket.

Innføring av den samme NonStop-tjenesten på ulike kontrollstasjoner vil trolig medføre ulike effekter i hvert tilfelle – og innføring av ulike varianter av NonStop-tjenesten på samme kontrollstasjon vil også kunne gi ulike utslag. Overføringsverdien av resultatene mellom ulike case er begrenset, hvis man ikke samtidig har kunnskap om hvordan ulike forhold og rammebetingelser

påvirker effektene. Dette nevnes også av Zhang et al. (2008), som trekker frem at kostnadene til kalibrering av WIM er svært avhengig av lokale klimatiske sesongvariasjoner. NonStop-studien har avdekket at stedsspesifikke faktorer som *trafikkvolum som klareres* og *fysisk utforming av kontrollstasjonen* vil ha stor betydning for nytteverdien, og en ex-ante vurdering av disse forholdene kan gi gode indikasjoner på effektene av en fremtidig innføring av NonStop-tiltaket.

Høy innovasjonsgrad og komplekse sammenhenger

Kunnskapsstatusen (presentert i kapittel 6.2) viste at de fleste studier av ITS-tiltak for klarering av tunge kjøretøy er basert på en kvasieksperimentell tilnærming, med før- og etterundersøkelser av målbare effekter. En slik metodikk forutsetter at sammenhengen mellom årsak (tiltaket) og virkning er entydig og kjent, og den søker å bevise en allmenngyldig og kontekstuavhengig effekt. Dette danner eksempelvis grunnlaget for effektkataloger som angir størrelsesorden på nytteverdier knyttet til spesifikke tiltak. Det finnes imidlertid lite historisk kunnskap om nytteeffekter av ITS-tiltak for klarering av tunge kjøretøy, og dagens empiri er i hovedsak begrenset til deteksjonsrater (Jones, 2008; Nova Scotia Transportation and Infrastructure Renewal, 2008; Stanczyk et al., 2008; Stanczyk & Klein, 2012). Foreliggende litteratur viser også at ITS-tiltakene som utvikles og tas i bruk er preget av høy innovasjonsgrad og komplekse årsakvirkningsforhold. Dette innebærer at ITS-tiltak for klarering av tunge kjøretøy i liten grad er standardisert, men derimot sammensatt av et utvalg komponenter og sensorer som medfører varierende funksjonalitet i de forskjellige installasjonene (Brown et al., 2007; Haugen et al., 2016). Dessuten har ITS-tiltaket en rekke ulike bruksområder (eksempelvis planlegging av kontrollvirksomhet, forhåndsutvelgelse av kjøretøy eller direkte håndheving) som også påvirker hvilke effekter som kan realiseres (FHWA, 2007; Jones, 2008). Forutsetningen om en allmenngyldig og entydig sammenheng mellom tiltak og effekt er dermed ikke til stede.

Gjennomførbarhet

Den eksperimentelle fremgangsmåten bygger på noen fundamentale metodiske premisser, blant annet i form av randomisering, bruk av kontrollgrupper og etablering av nullalternativ (baseline) for sammenligning. Disse premissene er i liten grad oppfylt i tidligere evalueringsstudier. Tvert imot, kan det synes som tiltak ofte iverksettes på steder med særskilte problemer og ofte uten bruk av kontrollgrupper (Nova Scotia Transportation and Infrastructure Renewal, 2008; Stephens et al., 2003). De metodiske utfordringene er i liten grad diskutert i tidligere publikasjoner, med unntak av Finne et al. (2012) som pekte på at det var vanskelig å finne indikatorer som kunne belyse effekter og virkninger på en enkel og entydig måte, og som også kommenterte at datagrunnlaget ikke var tilstrekkelig for å kunne etablere et nullalternativ og gjennomføre kontrafaktiske analyser. For øvrig

synes tidligere forskning å være begrenset til enkle sammenligninger av deteksjonsrater før og etter igangsetting av pilotprosjekter – uten videre drøfting av resultatenes gyldighet og overføringsverdi.

Den teoribaserte evalueringen viste seg å være gjennomførbar, til tross for et noe begrenset datagrunnlag. Muligheten til å kombinere flere metoder i evalueringsdesignet bidro til at datagrunnlaget ble utnyttet hensiktsmessig, og det resulterte i ny og anvendbar kunnskap på flere felt. Dette viser at evalueringer er nyttig også i tilfeller hvor ITS-tiltaket er teknologisk umodent og åpenbart prematurt for eksperimentelle før- og etterstudier. Når forutsetningene og premisene for bruk av den eksperimentelle metoden ikke er til stede, vil etablering av en velfundert programteori øke resultatenes teoretiske validitet.

I NonStop-prosjektet foregikk evalueringsarbeidet parallelt med utviklingen av selve ITS-tiltaket. Dette ga mulighet for kontinuerlig tilbakemelding og justering av NonStop-tjenesten underveis, og bidro dermed til at tiltaket ble utformet mest mulig relevant og effektivt med hensyn til intensjonene i prosjektet. På denne måten oppdages også barrierene (f.eks. teknologiske begrensninger og fysisk utforming av kontrollstasjon) i en tidlig fase, og man har mulighet til å justere tiltaket eller forkaste konseptet før det påløper store utviklingskostnader.

Evalueringsformål

Erfaringene fra dette case-studiet har vist at det er viktig å utforme evalueringsdesign og velge analysemetoder med utgangspunkt i evalueringens formål og de forskningsspørsmålene man ønsker å besvare. Dette står i kontrast til den tradisjonelle metodestyrte evalueringsprosessen innenfor transportfaget, hvor den eksperimentelle modellen nesten uvilkårlig legges til grunn for analysene. I et slikt design er forskningsspørsmålet allerede gitt på forhånd: *Virker tiltaket i sin nåværende form?* I tilfellet med NonStop-tjenesten hadde man på det tidspunktet evalueringen foregikk et tiltak som ikke oppfylte de tekniske kravene til funksjonalitet og kvalitet, og spørsmålet over kan besvares uten en kostbar datainnsamling og tidkrevende analyser: *Nei, (det umodne) tiltaket virker ikke.* Dette bidrar imidlertid i liten grad til nyttig kunnskap. ITS-tiltak vil nesten alltid være gjenstand for kontinuerlig teknologisk utvikling og forbedring, og man trenger kunnskap om hvilke tiltak man bør satse på og hvordan disse bør utvikles videre. I evalueringen av NonStop-tjenesten var det derfor mer hensiktsmessig å spørre: *Hva er de potensielle effektene av å innføre NonStop-konseptet, og hvordan bør et slikt system utformes for å gi optimale virkninger?* Den teoribaserte evalueringen viste seg å kunne belyse denne problemstillingen også når NonStop-tjenesten var teknologisk umoden, og evalueringen bidro med resultater og ny kunnskap som både kontrollmyndigheter og teknologiutviklere kan anvende i fremtidige beslutninger om hvordan et kontrollsystem bør utformes og implementeres. Ved å utvide evidensbegrepet til å inkludere

erfaringsbasert kunnskap og brukervurderinger, har man oppnådd relevant og anvendbar kunnskapsproduksjon om det aktuelle ITS-tiltaket.

6.10 OPPSUMMERING AV CASE A: NONSTOP

Evaluering av ITS-tiltaket *NonStop klarering av tunge kjøretøy* har frembragt ny kunnskap om selve tiltaket, samt metodisk innsikt i evaluering av ITS-tiltak generelt. Ved oppstarten av prosjektet var det planlagt en tradisjonell eksperimentell evalueringsprosess med måling av effekter før og etter implementering av NonStop-tjenesten. Utviklingen av den tekniske løsningen viste seg etterhvert å være mer tid- og ressurskrevende enn forutsatt, noe som fikk konsekvenser både for tiltaket og evalueringsstudien. NonStop-tjenesten tilfredstilte ikke kravene til teknisk funksjonalitet og modenhet i løpet av prosjektperioden. Dette medførte at ressurser ble allokert fra implementering og evaluering til teknologiutvikling, noe som resulterte i nedskalering av demonstrator og datainnsamling. Som en konsekvens av at tiltaket var umodent og evalueringsressursene begrenset, ble det ikke funnet hensiktsmessig å gå videre med det opprinnelige evalueringsdesignet. Isteden ble det utviklet en teoribasert evalueringsprosess. Denne tilnærmingen skilte seg fra den opprinnelige evalueringsplanen på to viktige områder: Først og fremst ble metodevalget i evalueringsdesignet styrt av de relevante forskningsspørsmålene, og evalueringen evnet således å frembringe ny og anvendbar kunnskap om *konseptet NonStop*, til tross for at det konkrete *tiltaket* ikke fungerte optimalt. I tillegg ble situasjonsavhengige faktorer (kontekst) inkludert i evalueringen, noe som ga nyttig kunnskap om hvilke eksterne faktorer som bidrar til å forsterke eller forhindre at man realiserer de effektene man ønsker å oppnå. Den teoribaserte evalueringen viste seg å være både anvendelig og i stand til å frembringe ny og relevant kunnskap om ITS-tiltaket:

- Innføring av en NonStop-tjeneste har potensiale til å øke produktiviteten i transportbransjen. Beregninger viser at innføring av tiltaket på Sandmoen kontrollstasjon kan redusere forsinkelser med 2,33 min per passering for klarerte kjøretøy.
- NonStop-tiltaket gir god samfunnsøkonomisk lønnsomhet. Beregninger viser at nytteverdien for transportbransjen, i form av reduserte reisetider alene, er dobbelt så høy som kostnadene ved innføring og drift av tiltaket.
- NonStop-tiltaket har potensiale til å effektivisere kontrollvirksomheten vesentlig. Estimer viser at innføring av tiltaket på Sandmoen kontrollstasjon vil medføre økt produktivitet tilsvarende seks hele kontroll dager per år (45 timer), gitt dagens kontrollnivå.
- NonStop-tiltaket oppnår høy brukeraksept både hos kontrollører og aktører i transportbransjen.

- Implementeringsstrategien med hensyn til varsling av kontrollkandidater og brukergrensesnitt mot kontrollører har stor betydning for hvilke effekter det er mulig å oppnå.
- Rammebetingelser og situasjonsspesifikke faktorer, for eksempel fysisk utforming av kontrollstasjonen og trafikkvolum (antall klarerte kjøretøy), har stor betydning for hvilke og hvor store effekter det er mulig å oppnå.
- NonStop-tiltaket gir grunnlag for bedre planlegging av kontrollvirksomheten gjennom økt datainnsamling og statistikk om godstransport på veg. På Sandmoen kontrollstasjon er kjøretøy med indikasjoner på feil/mangler overrepresentert på kveld og natt.

Styrken i den teoribaserte tilnærmingen har i dette casestudiet vært knyttet til følgende forhold:

- Den teoribaserte fremgangsmåten kan benyttes for tiltak som er i en kontinuerlig utviklingsprosess. Det har lyktes å frembringe kunnskap om det fremtidige potensialet, selv om NonStop-tiltaket fortsatt var teknologisk umodent og i en tidlig fase av innovasjonsprosessen.
- Den teoribaserte fremgangsmåten kan håndtere tiltak med høy grad av kompleksitet, og gir økt forståelse for hvorfor et tiltak virker eller ikke. Etablering av NonStop-tiltakets programteori og underliggende logikk bidro til kartlegging og bevisstgjøring av relevante årsakvirkningsforhold.
- Den teoribaserte fremgangsmåten finner kombinasjonen av innsats og kontekst som kan utløse den effekten man ønsker å realisere. For NonStop-tiltaket ga det liten mening å prøve å finne en generell allmenngyldig effekt, all den tid omgivelsene og eksterne faktorer hadde avgjørende betydning for hvilke virkninger som er mulig å oppnå i den enkelte installasjon.
- Den teoribaserte fremgangsmåten har vist seg å være praktisk gjennomførbar, også i en situasjon med begrenset datagrunnlag og hvor premisene for en tradisjonell eksperimentell evaluering ikke var oppfylt.

Alle forhold tatt i betraktning, har evalueringsstudien vist at en overgang fra den tradisjonelle eksperimentelle modellen til en teoribasert tilnærming ikke vil begrense kunnskapsproduksjonen. Tvert imot vil den i mange tilfeller bidra til økt forståelse omkring rammebetingelser og eksterne faktorer som må hensyntas for å oppnå en vellykket implementering. NonStop-caset er også et eksempel på at en teoribasert evalueringsmetodikk kan gi input til tradisjonelle beslutningsprosesser med bruk av nyttekostnadsanalyser.

7 CASE B: SAMKJØRING

*Dette kapittelet presenterer en evaluering av ITS-tiltaket **Samkjøring**, basert på et pilotprosjekt med pågående samkjøringsaktiviteter i Bergen. Studiet inngår som case i foreliggende doktorgradsarbeid. Evalueringen er utført i regi av SINTEF Teknologi og Samfunn på oppdrag fra Statens vegvesen Vegdirektoratet, med Seniorrådgiver Solveig Meland som prosjektleder og PhD-kandidat Lone-Eirin Lervåg som sentral medarbeider. Ved evalueringsprosjektets oppstart var det lagt til grunn et samarbeid mellom prosjektet og PhD-studiet, hvor prosjektet bidro med data og resultater til case-studiet, og hvor doktorgradsarbeidet bidro med utvikling av evalueringsmetodikk og innspill til evalueringsarbeidet. PhD-kandidaten har hatt hovedansvaret for gjennomføring av følgende arbeidsoppgaver i prosjektet:*

- *Utvikling av overordnet evalueringsmetodikk*
- *Litteraturstudie*
- *Analyser av data om holdninger, motivasjonsfaktorer og brukeraksept fra spørreundersøkelse blant ansatte i bedrifter som deltok i pilotprosjektet*
- *Analyser av data fra intervju med bedriftskontakter*

Prosjektet er rapportert i Meland, Lervåg og Roche-Cerasi (2015).

Innledningsvis presenteres rammene for evalueringen, med beskrivelse av samkjøringstiltaket og evalueringsprosjektet. Videre gjøres det rede for eksisterende kunnskap om ITS-tiltak for samkjøring, basert på tidligere pilotprosjekter og publiserte forskningsresultater innenfor fagfeltet. Samkjøring eksisterer i mange former, men i dette arbeidet er samkjøringstiltak avgrenset til organiserte, teknologibaserte løsninger, hvor trafikanter og turer kobles automatisk ved hjelp av en applikasjon på smarttelefon. Samkjøringstiltak krever lite investering i infrastruktur, da hensikten er å utnytte ledig kapasitet i biler som allerede er på veien. Det har derfor vært en del interesse for tiltaket, men det er få studier som kan vise til faktiske effekter. Kunnskapsstatusen inkluderer en gjennomgang av tidligere forskning med utgangspunkt i anvendte evalueringsmetoder og metodiske utfordringer. Basert på kunnskap om evalueringsteori og erfaringer fra litteraturstudien, ble det utviklet en teoribasert evalueringsmetodikk. Det er gjort rede for den trinnvise evalueringsprosessen frem mot analyser og resultater, etterfulgt av konklusjonene fra evalueringen. Til slutt drøftes hvordan case-studiet har bidratt til ny kunnskap om dette ITS-tiltaket spesielt og om evaluering av ITS-tiltak generelt.

7.1 BAKGRUNN

Transportvekst og økende køproblematikk i byområder har gitt grunnlag for å se på tiltak som kan effektivisere persontransporten, gjennom bedre utnyttelse av eksisterende infrastruktur.

Samkjøringstjenester har som formål å utnytte ledig kapasitet i privatbiler, ved å stimulere til at trafikanter som skal samme veg deler kjøretøy, istedenfor å benytte hver sin bil. Gjennom å øke personbelegget (antall passasjerer) i bilene, siktes det mot en reduksjon i antall biler på vegen, og dermed mindre kapasitetsproblemer på sikt. ITS-baserte samkjøringstiltak er blitt aktualisert gjennom nyere tids teknologiutvikling. I dag har de fleste norske trafikanter tilgang til en smarttelefon som gjør det mulig å bestille og administrere reisen underveis, GPS-navigasjon som lokaliserer kjøretøy/passasjer og optimaliserer kjøreruten, samt sosiale nettverk som kan bidra til å etablere tillit mellom fører og passasjer. Samkjøringen koordineres via kommersielle applikasjoner som kobler fører og passasjer, samt sørger for administrasjon av blant annet betaling mellom trafikantene. Carma, GoMore og SammeVei er eksempler på slike samkjøringsapplikasjoner tilgjengelig i Norge.

Statens vegvesen har gjennom flere år gjennomført en rekke forsknings- og utviklingsaktiviteter knyttet til samkjøringstiltak i Bergensområdet, med uttesting av ulike samkjøringsløsninger basert på mobiltelefonapplikasjoner (ITS Norge, 2011, 2013; Statens vegvesen, 2009). Aktivitetene har vært en del av prosjektet *Spontan samkjøring*, som blant annet har omfattet pilotprosjektet *Bergenspiloten* i samarbeid mellom Statens vegvesen, ITS Norge, Triona AS og ni bedrifter³⁹ med til sammen 7000 arbeidstakere i Kokstad/Sandsli-området sør for Bergen. Gjennom aktivt promoteringsarbeid og tilgang til den kommersielle samkjøringsapplikasjonen Carma⁴⁰, skulle arbeidstakerne i området stimuleres til å velge samkjøring på sine arbeidsreiser. Målsettingen med piloten var å oppnå 25 % reduksjon i antall arbeidsreiser med bil over en treårsperiode (2011-2014), og øke personbelegget i private biler til 1,4 personer/kjøretøy innen 2030 (ITS Norge, 2011; Statens vegvesen, 2009).

Kokstad/Sandsli ligger omtrent 15 km sør for Bergen sentrum, og består hovedsakelig av kontorbygg og næringsvirksomhet. Bilandelen for arbeidsreiser i dette området ligger på ca 73 % (Meland & Nordtømme, 2014), og i rushperiodene er det store utfordringer med kø og forsinkelser på vegen, i tillegg til begrenset kapasitet på parkeringsplasser i tilknytning til arbeidsplassene i området. På hovedvegen mellom Bergen sentrum og Kokstad/Sandsli er det etablert en 3 km lang

³⁹ Aibel, Aker Solutions, Avinor, BKK, Hansa Borg, Odfjell drilling, Owec Tower AS, Statoil og Reinertsen.

⁴⁰ Utviklet og tilbudt prosjektet av den irske teknologileverandøren Carma Technology Corporation.

strekning med sambruksfelt i begge retninger, tilgjengelig for kjøretøy med minst to passasjerer (inkludert fører).

7.1.1 KORT BESKRIVELSE AV ITS-TILTAKET SAMKJØRING

Samkjøringstiltaket som ble iverksatt gjennom Bergenspiloten var basert på den kommersielle ITS-tjenesten Carma, som ble gjort tilgjengelig for deltakerne både gjennom en nettversjon og applikasjon for smarttelefon. Ved oppstarten av pilotprosjektet fantes få mobilbaserte samkjøringsapplikasjoner, og Carma fremsto som en teknologisk moden applikasjon som allerede var tilgjengelig i det kommersielle markedet og brukt i stort omfang blant annet i San Francisco. En ny versjon av samkjøringsapplikasjonen ble lansert høsten 2012, og den tekniske løsningen ble videreutviklet og oppdatert gjennom hele prosjektperioden, med stadig forbedring av funksjonalitet og brukervennlighet. Deltakerne registrerte seg med en brukerprofil, og så ble applikasjonen brukt for videre organisering og gjennomføring av samkjøringsturer. Ved å poste forespørsler relatert til tidspunkt, sted (strekning) og hvorvidt man er fører eller passasjer, ble brukerne koblet sammen ved hjelp av algoritmer. Samkjøringsturen ble deretter registrert i mobilapplikasjonen ved gjennomføring, og det ble automatisk beregnet en kilometerbasert betaling/godtgjørelse for reisen. Avslutningsvis hadde både fører og passasjer mulighet til å gi hverandre en karakter (1-5 stjerner), som også ble gjort synlig for andre brukere av samkjøringstjenesten. I Bergenspiloten ble reisekostnadene dekt av pilotprosjektet og teknologileverandøren, slik at passasjerene i praksis reiste gratis, mens bilførerne fikk en beskjedne godtgjørelse for å ta med passasjer på turen.

ITS-tiltaket i Bergenspiloten omfattet flere virkemidler enn bare samkjøringsapplikasjonen på mobiltelefon. I tillegg til den teknologiske løsningen, ble det brukt vesentlige ressurser på promoteringstiltak og forankring i bedriftene. En prosjektleder var ansatt på timesbasis for å gjennomføre rekrutteringskampanjer, samt besørge tilrettelegging og kontinuerlig informasjonsformidling til deltakerne gjennom nettsider og sosiale media. Det ble gjennom hele pilotperioden arrangert månedskampanjer og delt ut premier (bl.a. iPhone, iPad) basert på reiseaktivitet og bruk av samkjøringsløsningen. Teknologileverandøren bidro med opplæring og veiledning til deltakerne, samt til dekning av passasjerenes bidrag til sjåførgodtgjørelse. Bedriftene utnevnte egne samkjøringsambassadører med spesielt ansvar for rekruttering og brukerstøtte til øvrige ansatte i egen bedrift. Bergenspiloten hadde også en svært aktiv prosjektgruppe som blant annet besørget en god del omtale i lokale aviser og andre media.

7.1.2 KORT BESKRIVELSE AV PROSJEKTET EVALUERING AV SAMKJØRING

Dette case-studiet er gjennomført som del av forskningsprosjektet *Evaluering av samkjøring*, på oppdrag fra Statens vegvesen Vegdirektoratet. Formålet med forskningsprosjektet var å avdekke hvorvidt samkjøring har potensiale som etterspørselsstyrende og trafikkdempende tiltak i Norge. Evalueringen skulle avdekke faktiske effekter med utgangspunkt i pågående aktiviteter tilknyttet Bergenspiloten, i tillegg til å vurdere potensielle virkninger av å iverksette samkjøringstiltak i større skala. Prosjektet var således ikke en del av Bergenspiloten, men kom inn i pilotprosjektets avsluttende fase. Evalueringen hadde et vesentlig element av læring og kunnskapsbygging, da en viktig del av oppdraget var å identifisere kritiske barrierer og suksessfaktorer for innføring og bruk av samkjøringstjenester i et fremtidig transportsystem.

Prosjektet var strukturert og ble gjennomført med utgangspunkt i følgende hovedaktiviteter:

1. **Litteraturstudie:** Kartlegging av kunnskapsstatus for samkjøringstiltak og tjenester knyttet til slike aktiviteter.
2. **Spørreundersøkelse blant arbeidstakere:** Kartlegging av holdninger, reiseatferd og erfaringer med samkjøring
3. **Driftsstatistikk fra samkjøringstjenesten:** Data om antall brukere og gjennomførte samkjøringsturer i pilotperioden
4. **Driftsstatistikk fra prosjektgruppa:** Informasjon om ressursbruk og kartlegging av insentiver og markedsføringstiltak
5. **Intervju og samtaler med bedriftsledelse og samkjøringsambassadører:** Kartlegging av brukererfaring (aksept)

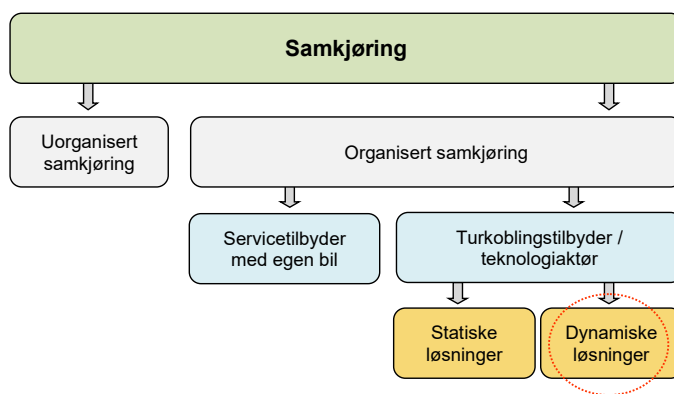
Prosjektet og aktivitetene er nærmere beskrevet i Meland et al. (2015).

7.2 EKSISTERENDE KUNNSKAP OM SAMKJØRINGSTILTAK

Det eksisterer mange ordninger for samkjøring, og konkrete samkjøringsløsninger er utviklet og testet som transportregulerende tiltak gjennom flere tiår. Chan og Shaheen (2012) trekker historien helt tilbake til andre verdenskrig, hvor begrenset tilgang til kjøretøy gjorde det nødvendig å organisere felles transportløsninger, etterfulgt av energikrisen på 1970-tallet, da samkjøring ble en løsning på drivstoffmangel og restriksjoner på bilbruk. I nyere tid har samkjøring ofte blitt drøftet som tiltak for å redusere kapasitetsproblemer og miljøutfordringer som følge av mye biltrafikk, spesielt i forbindelse med arbeidsreiser i rushtid. I Norge foregår arbeidsreiser hovedsakelig med personbil (65 %), og med et beskjedent personbelegg på i snitt 1,15 personer per kjøretøy (Hjorthol,

Engebretsen, & Uteng, 2014). Dette innebærer en god del ledig setekapasitet som ikke blir utnyttet, hvilket innebærer et uforløst potensial for samkjøring som trafikkreduserende tiltak.

Begrepet samkjøring har ingen entydig definisjon, og i praksis brukes flere begreper for å beskrive ulike varianter av samkjøringsløsninger⁴¹. Furuata et al. (2013) skiller mellom uorganisert og organisert samkjøring, se Figur 18.



Figur 18: Kategorisering av samkjøringsløsninger, basert på Furuata (2013). Bergenspilotten benytter seg av en dynamisk samkjøringsløsning (Carma-applikasjonen).

Den uorganiserte samkjøringen omfatter de tilfeldige avtaler om skyss som gjøres mellom familie, venner eller kollegaer som skal samme vei på omtrent samme tidspunkt. Denne formen for uorganisert samkjøring omfatter størsteparten av samkjøringen (bilpassasjerer) som foregår i transportsystemet i dag. Benevnelsen *slugging* brukes også om en tilfeldig og uorganisert form for samkjøring som forekommer flere steder i USA (bl.a. Washington DC, Houston og San Fransisco) (Chan & Shaheen, 2012). Der har det oppstått en kultur hvor bilførere plukker opp passasjerer ved bussholdeplasser og innfartsparkeringer, for å få tilgang til tidsbesparende sambruksfelt. Den organiserte samkjøringen involverer en tredje aktør som kobler førere og passasjerer som skal samme veg (Amey, 2010; Furuata et al., 2013). Her skiller litteraturen mellom såkalte servicetilbydere med egne kjøretøy (typisk Uber og konvensjonelle drosjeordninger) og den andre varianten som tilbyr turkobling (matching av folk og turer), men hvor det er trafikantene selv som stiller med kjøretøyet. Det er spesielt den siste formen for samkjøring som er aktuell som trafikkregulerende tiltak. Videre skiller det også mellom statiske løsninger hvor man poster forespørsler på en nettside eller via sosiale medier (f.eks. facebook-grupper), og deretter gjør

⁴¹ På norsk brukes både *samkjøring*, *spontan samkjøring*, *dynamisk samkjøring*, *tilfeldig samkjøring*, *kompiskjøring* og *bildeling*, mens det i engelskspråklig litteratur finnes et en rekke begreper rundt benevnelsen *ridesharing* og *carpooling* som i hovedsak beskriver det samme fenomenet.

personlige avtaler om hentested, tidspunkt og betaling – og mer dynamiske løsninger som kjennetegnes ved at det foregår en automatisk kobling av fører og passasjer, for eksempel med en samkjøringsapplikasjon på mobiltelefon. Det finnes flere eksempler på såkalt statiske samkjøringsløsninger⁴². Typisk for flere av disse er at de er startet med utgangspunkt i universiteter og organiserer litt lengre turer fra studiestedet i skoleferier, til og fra festivaler og idrettsarrangement. De dynamiske løsningene⁴³ legger til rette for at avtaler kan gjøres spontant og i sanntid, og inkluderer gjerne et etablert system for deling av reisekostnader (Agatz, Erera, Savelsbergh, & Wang, 2012). Samkjøringsløsningene gir i utgangspunktet ingen økonomisk profit, men sørger for at føreren som stiller med kjøretøyet får dekt en liten andel av reisekostnadene. Samkjøringstiltaket i Bergenspiloten er et eksempel på en dynamisk samkjøringsløsning basert på turkoblingsalgoritmer i en mobilapplikasjon. Videre bruk av begrepet *samkjøring* i dette kapittelet er knyttet opp mot de organiserte, teknologibaserte samkjøringstjenestene som kobler førere og passasjerer som skal reise tilnærmet samme strekning.

Felles for de fleste pilotene og samkjøringsprosjektene synes å være at de krever stor innsats for å fungere, og det har vist seg vanskelig å få opp bærekraftige løsninger som lykkes over tid (Agatz et al., 2012; Amey, Attanucci, & Mishalani, 2011; Chan & Shaheen, 2012; Deakin, Frick, & Shively, 2010). I Norge har samkjøringstjenester tidligere vært testet ut både i Bergen og på Lysaker/Fornebu, og selv om mange i utgangspunktet stiller seg positive til samkjøringskonseptet, har man kun oppnådd begrenset bruk av tiltaket (Gjerstad & Bayer, 2012). Dette er også i tråd med internasjonale erfaringer som viser at det å oppnå en kritisk brukermasse er den aller største barrieren for å lykkes med tiltaket. Dersom samkjøring skal være et reelt og konkurransedyktig transporttilbud, må det være tilstrekkelig mange personer som reiser samme vei, slik at man finner koblinger når det postes turforespørsler. De senere års raske teknologiutvikling har således avstedkommet positive forventninger hos samkjøringsentusiaster – fordi tilknytning til sosiale nettverk (f.eks. facebook) og automatiserte koblingsalgoritmer har effektivisert og forenklet organiseringen av samkjøringsaktivitetene (Chan & Shaheen, 2012; Dailey, Loseff, & Meyers, 1999). Etablering av en felles teknologiplattform som muliggjør kobling av brukere fra forskjellige samkjøringsløsninger (mobilapper) kan også bidra til at man oppnår en kritisk brukermasse for tjenestene. Erfaringer så langt tyder imidlertid på at det er tungt å endre transportvaner, og det er behov for å kombinere tiltaket med andre insentiver for å få folk til å velge å samkjøre som erstatning for private bilreiser.

⁴² Eksempelvis Zimride i USA, PickupPal i Canada og Liftshare i Storbritannia

⁴³ Eksempelvis skandinaviske GoMore og irske Carma

7.2.1 POTENSIELLE VIRKNINGER, SUKSESSKRITERIER OG BARRIERER

Hva er målsettingen med samkjøringstiltaket?

Når myndigheter eller lokale initiativtakere etablerer en samkjøringsordning, er hensikten som regel å løse flere utfordringer i transportsektoren (Agatz et al., 2012; Amey, 2010; Furuata et al., 2013). Det er i hovedsak tre argumenter som går igjen i alle samkjøringsprosjekter; tiltaket forventes å *redusere miljølempen, gi bedre kapasitetsutnyttelse av eksisterende infrastruktur, samt øke trafikantenes mobilitet*. Som miljøtiltak skal samkjøring bidra til å redusere energiforbruk (drivstoff), klimagassutslipp og luftforurensning. Vågane (2009) har beregnet potensielle miljøeffekter basert på et tenkt scenario hvor samkjøringstiltaket bidrar til å øke personbelegget i kjøretøy med 5 % per kjørte kilometer, fra 1,7 til 1,78 på landsbasis. Dette tilsvarer en årlig reduksjon på 234.000 tonn CO₂, 498.000 tonn NO_x og 27.900 tonn PM₁₀. Videre skal økt kapasitetsutnyttelse redusere biltrafikk og køproblematikk, samt gi mindre behov for arealbruk til blant annet parkering. Det er imidlertid lite tilgjengelig forskning som kan bekrefte at samkjøringstiltaket realiserer disse målsettingene i praksis. De fleste studier er begrenset til å beregne potensielle virkninger basert på en forutsetning om at man oppnår en gitt effekt på reisemiddelfordelingen. Det er imidlertid en utfordring for samkjøring som miljø- og trafikkreduserende tiltak at de fleste deltakerne i samkjøringsaktiviteter ser ut til å rekrutteres fra andre miljøvennlige transportformer som kollektivtransport, sykling og fotgjengere (Amey, 2010; Deakin et al., 2010; Gjerstad & Bayer, 2012). Til sist anses samkjøring også som et mobilitetstiltak som kan dekke geografiske områder med manglende offentlig kollektivtilbud, fungere som tilbringertjeneste til eksisterende kollektivknutepunkt eller også forbedre trafikantenes generelle transporttilbud gjennom eksempelvis økt komfort eller reduserte transportkostnader. Det er imidlertid sparsomt med forskningsresultater som fokuserer på samkjøringstjenester som supplement til øvrig kollektivtrafikk (Murray & Chase, 2012). I senere tid er det publisert beregninger hvor samkjøring knyttes sammen med bruk av selvkjørende kjøretøy i simuleringsmodeller, som viser hvordan introduksjon av selvkjørende og delt mobilitet kan dekke dagens transportbehov med en vesentlig redusert kjøretøyflåte (Arbib & Seba, 2017; International Transport Forum, 2017a, 2017b).

Suksesskriterier

Studier av forsøk med samkjøring peker på en rekke potensielle gevinster for ulike aktører (Agatz et al., 2012; Amey, 2010; Chan & Shaheen, 2012; Furuata et al., 2013; Griffin et al., 2015). Involvering av bedrifter og fokus på arbeidsreiser synes å være gunstig med hensyn til å rekruttere brukere av samkjøringstjenester (Griffin et al., 2015). Bedrifter som deltar i samkjøringsprosjekter er gjerne motivert av kostnadsbesparelser tilknyttet parkeringsanlegg, økt produktivitet blant ansatte

og miljøprofilering (Amey, 2010). Den individuelle nytten for trafikanter som benytter seg av samkjøringstjenesten er i hovedsak knyttet til kostnadsbesparelse gjennom deling av reisekostnader, økt mobilitet (flere reisemiddelvalg) og det idealistiske perspektivet knyttet til en miljøvennlig atferd. En del trafikanter vil også oppleve redusert reisetid, for eksempel passasjerer som ellers hadde reist kollektivt eller sjåfører som får tilgang til sambruksfelt og gunstige parkeringsplasser. I tillegg kommer gevinsten av eventuelle insentivordninger. En amerikansk studie av markedspotensialet for samkjøring, indikerte at dersom brukerne fikk tilbud om et (selvvalgt) insentiv som oppleves meningsfylt, kan den potensielle deltakelsen øke fra 20 % til 50 %, basert på selvrapporterte intensjoner (Deakin et al., 2010). De mest foretrukne insentivene var gratis parkering eller tilgang til attraktive parkeringsplasser, og mulighet for gratis retur med taxi i de tilfeller hvor man ikke fikk treff på hjemreise med samkjøring. Den samme studien avdekket imidlertid at samkjøringsordninger appellerte mest til de som allerede samkjører uorganisert (bilførere med passasjerer og bilpassasjerer), og viktigste motivasjonsfaktorer for deltakerne var redusert reisetid og kostnadsbesparelser. Norske erfaringer fra pilotprosjekter med samkjøring tyder på at deltakerne lar seg motivere av å unngå kø (tilgang til sambruksfelt), miljøhensyn, samt interesse for å være med å prøve ut ny teknologi (Gjerstad & Bayer, 2012).

Erfaringer så langt tyder på at de individuelle fordelene overskygges av ulempene ved samkjøring, og det er vesentlige barrierer som må overvinnes for å rekruttere tilstrekkelig mange brukere (Deakin et al., 2010). Alt tatt i betraktning synes tiltaket best egnet der samkjøring blir iverksatt i samarbeid med arbeidsgivere for å løse et spesifikt problem som parkeringsmangel eller store køproblemer, i kombinasjon med begrenset biltilgang (for eksempel som følge av økonomiske konjunkturer eller drivstoffrestriksjoner) (Amey, 2010; Amundsen, 2011). Samkjøringsløsninger som fungerer godt, er kjennetegnet av at de inngår i en helhetlig satsing med vedvarende promotering og markedsføring. Dette inkluderer blant annet informasjonskampanjer, aktiv rekruttering og premiering, og holdnings- og kulturskapende arbeid (Amey, 2010). Dette støttes også av en studie gjennomført av Griffin et al. (2015) som viser at det er sterk samvariasjon mellom markedsføringsaktiviteter, oppslag i lokale media og nye registrerte brukere, samt antall gjennomførte samkjøringsturer. Videre er tilgang til sambruksfelt en viktig suksessfaktor, fordi det gir umiddelbare fordeler i form av spart reisetid. Manglende parkeringskapasitet eller svært kostbar parkering synes også å være gunstig for rekruttering av deltakere, spesielt når dette kombineres med særlige fordeler (reduerte takster eller dedikerte plasser) for de som samkjører (Deakin et al., 2010). Et annet fellestrekk ved de pilotprosjektene som fungerer godt, er tilstedeværelsen av ekstern sponning. Dette innebærer at brukerne har opplevd reelle økonomiske fordeler med å samkjøre, eksempelvis med rabatterte priser ved bomplasseringer (Griffin et al., 2015). Alle faktorer tatt i

betraktning, kan det synes som om trafikantene må oppleve gevinster i form av redusert reisetid og/eller reduserte reisekostnader for at samkjøringstiltaket skal ha mulighet til å lykkes over tid.

Barrierer

Den største utfordringen med samkjøringstiltak er knyttet til rekruttering av et tilstrekkelig antall deltakere til at man oppnår en kritisk brukermasse for turkobling - slik at samkjøring blir et reelt transporttilbud for brukerne. En amerikansk studie (Griffin et al., 2015) dokumenterte resultatene fra et pilotprosjekt i Texas, hvor utgangspunktet for rekruttering av deltakere var forventet å være spesielt gunstig. Pilotprosjektet ble gjennomført med en teknologisk moden applikasjon (Carma), og i forbindelse med åpningen av et lokalt bomsystem – hvor deltakerne i piloten fikk rabatterte priser i bomstasjonen. Deler av rekrutteringsprosessen ble gjennomført i samarbeid med en teknologibedrift med 3000 ansatte, men selv etter omfattende promoteringsarbeid lykkes man kun å rekruttere 51 deltakere, hvorav 31 brukere foretok totalt 154 samkjøringsturer i løpet av en registreringsperiode på et par måneder. Videre viste analysene at de fem mest ivrige brukerne sto for hele 40 % av samkjøringsturene. Vanskelighetene med å rekruttere brukere underbygges også av resultatene fra en studie ved Berkeley Universitet i California (Deakin et al., 2010). Beregninger av markedspotensialet for samkjøring i tilknytning til universitetet, viste at kun 1 % av utgangspunktet på 70 000 ansatte og studenter var potensielle kandidater for å endre reiseatferd fra privatbil til samkjøring. Det metodiske grunnlaget for denne studien er nærmere presentert i kapittel 7.2.2.

Barrierer for å lykkes med samkjøring inkluderer brukerrelaterte, økonomiske, institusjonelle og teknologiske utfordringer (Amey, 2010; Amey et al., 2011; Correia & Viegas, 2011; Deakin et al., 2010; Furuhashi et al., 2013). Brukerrelaterte barrierer er blant annet knyttet til manglende fleksibilitet i reisen, pålitelighet og mulighet for å oppnå returreise, trafikantens holdninger og opplevelse av komfort, sikkerhet og trygghet, samt personvern. Mange trafikanter anser egentid i bil som en kvalitet i seg selv og opplever redusert komfort og et visst ubehag ved den sosiale situasjonen å skulle samkjøre med fremmede (Amey et al., 2011; Chan & Shaheen, 2012). Det å måtte vente og være avhengig av andre, ha begrenset mulighet til å endre avtaler underveis, samt være usikker på mulighet for returreise har også vist seg å være innskrenkende faktorer for viljen til å samkjøre (Deakin et al., 2010; Gjerstad & Bayer, 2012). De økonomiske barrierene er i hovedsak knyttet til manglende insentiver for å redusere bilhold; i dag bor 90 % av den norske befolkningen i husholdninger med bil, og 45 % av befolkningen har tilgang til minst to biler (Hjorthol et al., 2014). Betalingsmodellen for bilhold er også slik at enkeltturer fremstår som rimelige når man først har anskaffet kjøretøyet. Agatz et al. (2012) peker på at subsidiering av samkjøringsaktiviteter og premiering av deltakere kan være et rimelig og hensiktsmessig insentiv sett i forhold til for

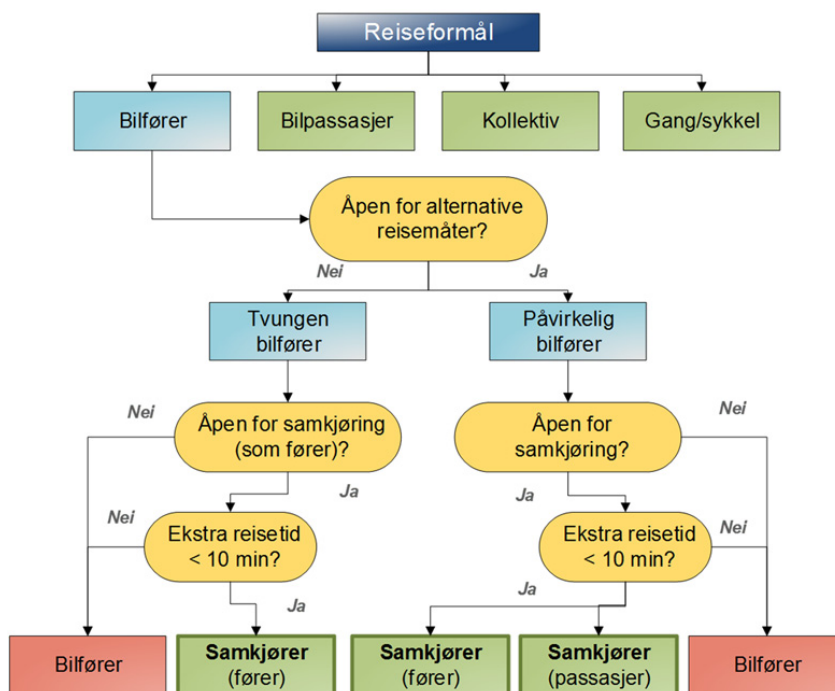
eksempel kapasitetsutbygging av vegnettet. Institusjonelle barrierer er knyttet til rolleavklaring mellom privat og offentlig sektor, forretningsmodeller og lover og reguleringer. Eksempelvis foreslår Amey (2010) at myndighetene bør innføre parkeringsrestriksjoner ved arbeidsplasser for å redusere antall arbeidsreiser med bil. Teknologiske barrierer er i stor grad knyttet til effektiv håndtering av dynamikken i turkoblingsapplikasjonen, eksempelvis akseptabel responstid, ruteoptimalisering og brukergrensesnitt (Agatz et al., 2012).

7.2.2 HVILKE EVALUERINGSMETODER BENYTTES?

Det finnes lite empiri og systematisk dokumentasjon på konkrete effekter av samkjøring, all den tid det er få eksempler på samkjøringstiltak som fungerer tilfredsstillende over tid. Chan og Shaheen (2012) refererer til samkjøring som et "usynlig transportmodus" fordi det er vanskelig å observere, studere eller samle kvantitative data om effektene av samkjøringstiltak. Av denne grunn peker forfatterne på et behov for forskning som bidrar til å forstå årsakssammenhenger og potensielle virkninger på infrastruktur, køutfordringer og miljøutslipp.

I den tidligere nevnte Berkeley-studien i California (Deakin et al., 2010) ble det gjennomført en mulighetsstudie ved å beregne markedspotensialet for samkjøring i tilknytning til universitetet. Evalueringemetodikken inkluderte en geografisk analyse av influensområde og arbeidsreiser, statistiske beregninger, fokusgruppeintervju med arbeidsreisende og spørreundersøkelse blant studenter og ansatte om nåværende reisevalg og preferanser, interesse for samkjøring, vurdering av et planlagt samkjøringsprogram med valgmuligheter og insentiver, samt selvrapportert potensiell bruk av samkjøringstjenesten. Markedspotensialet ble beregnet med utgangspunkt i 70.000 personer med arbeidsreiser til/fra Berkeley (30.000 ansatte og 40.000 studenter). Samkjøringstiltaket rettet seg inn mot bilførere som kjører alene på regulær basis, og denne målgruppen utgjorde 20 % av utgangspunktet. Basert på den geografiske analysen og resultatene fra spørreundersøkelsen, ble det satt som betingelse at de bodde i et egnet område (hvor flere reiser i samme retning) og at de uttrykte positive holdninger til samkjøring. Dette innskrenket målgruppen til 1850 personer, altså 3 % av utgangspunktet. Videre ble det gjennomført en simulering basert på ønsket reisetidspunkt, for å avdekke sannsynligheten for at deltakerne ville finne noen å samkjøre med. Det ble satt som betingelse at deltakerne på en typisk ukedag skulle ha minst 60 % sannsynlighet for å få kobling med en potensiell samkjører, og reisen skulle maksimalt medføre 10 minutter ekstra reisetid. Dette reduserte målgruppen til 2 % potensielle deltakere. Til sist ble det valgt å se bort fra personer som bor innenfor typisk gang- og sykkelavstand eller i et område med svært godt kollektivtilbud til universitetsområdet, da disse var utenfor målgruppen man ønsket å rekruttere. Etter det siste kriteriet, satt man igjen med 700 potensielle deltakere, altså 1 % av det totale utgangspunktet på

70.000 personer. I Berkeley-eksempelet ville man potensielt kunne oppnådd en maksimal bilreduksjon på omtrent 500 kjøretøy per dag, gitt et forholdsvis optimistisk anslag om at hver sjåfør plukket opp to passasjerer daglig. Dette viser at et stort geografisk nedslagsfelt med mange personer og relativt faste reisemønstre, ikke nødvendigvis gir et stort potensial for samkjøring som bilreducerende tiltak likevel. Det er derfor viktig at evalueringsstudier tar hensyn til kontekstavhengige faktorer som påvirker resultatene, og ikke kun baserer evalueringen på reisemiddelfordelingen alene. Beslutningsmodellen som ble lagt til grunn for vurdering av markedspotensialet er sammenstilt og illustrert i Figur 19.



Figur 19: Beslutningsmodell for vurdering av markedsgrunnlag for samkjøring, basert på erfaringer fra Deakin et al. (2010).

Tidligere samkjøringsaktiviteter i Norge er evaluert og rapportert i Gjerstad og Beyer (2012). Denne evalueringen var basert på utprøving av teknologibaserte samkjøringsløsninger⁴⁴ i en tidlig testfase, i tilknytning til arbeidsplasser på Kokstad/Sandsli i Bergen og Lysaker/Fornebu i Oslo. Studien ble gjennomført med en spørreundersøkelse blant ansatte hos medlemsbedriftene (N=754), samt intervju med 7 brukere og 4 ikke-brukere av tjenesten. De to pilotene rekrutterte til sammen 127 påmeldte deltakere, og spørreundersøkelsen la vekt på å avdekke reisevaner, holdninger til

⁴⁴ Mobilapplikasjonene Avego og Sharepool og den internetbaserte løsningen HentMeg (utviklet av Triona).

samkjøring og erfaringer fra piloten. De metodiske utfordringene knyttet til evalueringsstudien ble presentert og drøftet på en workshop i regi av Statens vegvesen Vegdirektoratet i etterkant av prosjektet (Gjerstad & Bayer, 2012; Lervåg, 2012). Erfaringene viste at studien var egnet til å løfte frem kompleksiteten i ITS-tiltaket, men det var vanskelig å formulere entydige konklusjoner og vurdere overførbarheten til en normal driftssituasjon. Dette skyldtes blant annet at det var få reelle brukere av tjenesten, og forutsetningen om en kritisk brukermasse var langt fra oppfylt. Det manglet grunnlag for beregning av effekter, blant annet fordi man ikke greide å identifisere en entydig årsakssammenheng. Manglende innsikt i betydningen av eksterne faktorer og insentiver gjorde det vanskelig å isolere effekten av selve tiltaket. Det ble også pekt på nødvendigheten av å registrere bakgrunnsinformasjon om reisevaner, og ikke kun vurdere effekten ut ifra brukere av samkjøringsløsningen, all den tid det var grunnlag for å tro at tiltaket rekrutterte sine brukere delvis fra kollektivtransport. Til sist ble det konkludert med at resultatene fra pilotprosjektet hadde begrenset gyldighet for en normal driftssituasjon, blant annet fordi den tekniske løsningen fremsto som et uferdig produkt med ustabil funksjonalitet og lite tilfredsstillende brukergrensesnitt.

Griffin et al. (2015) rapporterer resultater fra et amerikansk pilotprosjekt med samkjøring i Texas, hvor brukere av en spesifikk samkjøringsapplikasjon⁴⁵ fikk rabatterte priser ved passering av bomstasjoner. Evalueringen baserte seg på driftsstatistikk fra samkjøringsløsningen og en spørreundersøkelse blant de som hadde registrert seg med brukerprofil i systemet. Totalt ble det i løpet av pilotperioden (januar - desember 2014) registrert 322 deltakere, og med stigende aktivitet gjennom hele perioden ble det i den siste måneden av piloten registrert 515 samkjøringsturer (desember 2014). Spørreundersøkelsen ble gjennomført med 25 respondenter som besvarte spørsmål om siste ukes reisevaner i to omganger; første gang ved påmelding til pilotprosjektet (N=313 turer) og andre gang etter at applikasjonen var lastet ned og tatt i bruk (N=219 turer). Det ble ikke registrert signifikant endring i reisemiddelfordeling før og etter samkjøringsapplikasjonen ble tatt i bruk, og dette forklares med for lav deltakelse i samkjøringsaktivitetene generelt og at datagrunnlaget (N=25 respondenter) ikke var tilstrekkelig til å avdekke signifikante effekter. Felles for en del av evalueringsstudiene som ikke er i stand til å konkludere på effekter av samkjøringstiltaket, synes å være at de isteden forsøker å avdekke sekundære virkninger. I dette pilotprosjektet ble det konkludert med at samkjøringsteknologien var tilstrekkelig moden til kunne å benyttes som verktøy for differensiert vegprising basert på personbelegg i kjøretøy.

Amey (2010) har gjennomført en omfattende studie av samkjøringsaktiviteter i USA, basert på en litteraturgjennomgang og en rekke intervjuer med eksperter og interessegrupper. Hensikten var å

⁴⁵ Den kommersielle samkjøringsapplikasjonen Carma, samme løsning som ble anvendt i Bergenspiloten

avdekke og forstå betydningen av ulike barrierer for rekruttering av deltakere. Forskningen avdekket et vesentlig kunnskapsgap med hensyn til å forstå markedspotensialet for samkjøring, og studien ble fulgt opp med en simulering av samkjøringspotensialet for 22.000 ansatte og studenter ved Massachusetts Institute of Technology (MIT), basert på data fra reisevaneundersøkelser og de rent fysiske betingelsene for arbeidsreisen (kjøretøytilgjengelighet, reises startpunkt og destinasjon⁴⁶, avreise-/ankomsttidspunkt). Resultatene indikerte at på en ideell dag, kunne hypotetisk 65 % av alenaturene med bil vært erstattet av samkjøringsturer, med en reduksjon på 19 % i trafikkarbeidet (vehiclemiles) for arbeidsreisene. Differansen mellom disse resultatene, som er beregnet utelukkende basert på de fysiske egenskapene ved reisen, og tidligere studier som indikerer en potensiell reduksjon i trafikkarbeid i størrelsesorden 0,4-2 % (Amey, 2010), representerer dermed betydningen av ulike holdnings- og atferdsrelaterte barrierer hos trafikantene.

Kunnskapsgjennomgangen viser at ITS-tiltaket samkjøring i liten grad har evnet å dokumentere effekter på transportterspørsel og reisemiddelfordeling. De fleste pilotprosjekter og samkjøringsordninger har store problemer med å rekruttere tilstrekkelig antall deltakere, og det er et åpenbart behov for å etablere kunnskap om hvem som er potensielle brukere av samkjøringstjenester, og hva som kan motivere til økt deltakelse. Begrenset rekruttering medfører også metodiske utfordringer for evalueringsstudier som må baseres på små utvalg og liten mulighet for beregning av statistisk signifikans (lav statistisk validitet). Publiserte forskningsresultater tyder på at det ikke finnes en allmenngyldig effekt og et entydig forhold mellom samkjøringstiltaket og hvilke virkninger det avstedkommer. I hvilken grad man lykkes med å rekruttere brukere, og dermed kan realisere effekter, avhenger av en rekke situasjonsavhengige faktorer knyttet til innføringen av tiltaket. Det synes klart at den individuelle nytteverdien av samkjøring i de fleste tilfeller overskygges av ulempene for den enkelte bruker, og skal man lykkes med tiltaket må det følges av en rekke insentiver som kan bidra til å overvinne barrierene.

Suksesskriterier for samkjøring synes å være knyttet til arbeidsreiser og samarbeid med arbeidsgiver, at tiltaket adresserer et opplevd transportproblem, en aktiv prosjektorganisasjon som bidrar med kontinuerlig markedsføring og rekruttering, kostnadsbesparelser og/eller redusert reisetid for den enkelte bruker, samt bruk av insentiver som for eksempel tilgang til sambruksfelt, parkeringsplasser eller prisreduksjon i bomstasjoner.

⁴⁶ Analysene tok ikke hensyn til eventuelle reisekjeder, men la til grunn at alle reisene gikk direkte fra hjem til arbeid eller motsatt.

Til tross for manglende historisk suksess, er det stadig interesse for samkjøring som transporttiltak, og det er knyttet en forventning til at bedre teknologi og endrede holdninger til sosiale nettverk og delingsøkonomi skal bidra til å aktualisere samkjøring. Som mange andre ITS-tiltak, er samkjøringsløsninger gjenstand for en pågående teknologiutvikling, og dette gjør at evalueringresultater i liten grad er gyldige over tid. Mens tidlige pilotprosjekter var preget av lite effektive koblingsrutiner og umoden teknologi, fremstår nyere samkjøringsløsninger som teknisk stabile og med et tilfredsstillende brukergrensesnitt. Teknologiske oppdateringer forbedrer imidlertid stadig funksjonaliteten for brukerne, og evalueringstudier må ta hensyn til at tiltaket er i kontinuerlig utvikling.

Erfaringer så langt har vist at markedspotensialet for samkjøring er svært begrenset. For mange er det ikke realistisk å endre reiseatferd til samkjøring, enten på grunn av ytre omstendigheter (f.eks. henting i barnehage eller at det ikke finnes noen å samkjøre med på gitt sted og tid) eller på grunn av egne holdninger og overbevisninger om reisemiddelvalg. Selv om det finnes en del studier som kartlegger insentiver og motivasjonsfaktorer generelt, er det i liten grad funnet studier som fokuserer på de som faktisk samkjører. Fremtidige evalueringstudier bør kanskje legge vekt på å identifisere hvilke rammebetingelser og motivasjonsfaktorer som har størst betydning for den delen av målgruppen som faktisk kan påvirkes til å endre reiseatferd?

7.3 UTVIKLING AV EN TEORIBASERT EVALUERINGSMETODIKK

Med utgangspunkt i erfaringene fra litteraturstudien og evalueringens målsettinger, ble det besluttet å anvende en teoribasert tilnærming for evaluering av ITS-tiltaket samkjøring. Denne avgjørelsen var basert på tre sentrale argumenter:

Høy kompleksitet med mange involverte aktører og aktiviteter

Samkjøringstiltaket hadde høy kompleksitet med mange involverte aktører og parallelle aktiviteter. Selve samkjøringsløsningen var gjenstand for kontinuerlig teknologiutvikling, hvilket innebar at man ikke kunne forvente at den eksperimentelle metodens forutsetning om et stabilt tiltak ville være oppfylt. I tillegg ble det igangsatt en rekke aktiviteter fra ulike aktører, slik at ITS-tiltaket i praksis besto av en samling ulike virkemidler som kunne påvirke det endelige resultatet. Eksempelvis ble det lagt stor innsats i promoterings- og markedsføringsaktiviteter. Samtidig hadde litteraturstudien vist at nytten av tiltaket er avhengig av en viss mengde deltakere (kritisk masse), slik at man har tilstrekkelig grunnlag for kobling av sjåfører og passasjerer. Tidspunktet for gjennomføring av evalueringen ville dermed få stor betydning for resultatene, og det var behov for en

evalueringsmetodikk som også kunne fremskaffe kunnskap om tiltakets fremtidige potensiale dersom den aktuelle installasjonen ikke lykkes å rekruttere en kritisk brukermasse.

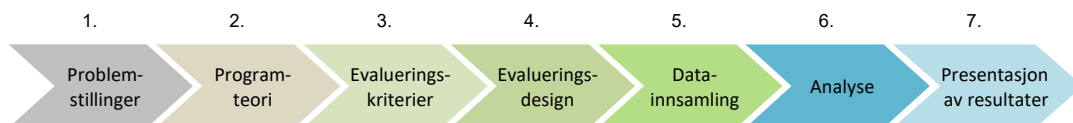
Kontekstavhengige effekter

Det var forventet at en rekke situasjonsspesifikke faktorer ville kunne påvirke resultatet av Bergenspiloten. Med utgangspunkt i kunnskapsstatusen, syntes Bergenspiloten å ha forholdsvis gode forutsetninger for å lykkes, med en aktiv prosjektorganisasjon, et etablert samarbeid med arbeidsgivere i området, tilgang til sambruksfelt, begrenset parkeringskapasitet med prioritet til samkjøring og økonomisk godtgjørelse for deltakelse. Det var behov for å inkludere disse faktorene i evalueringen, slik at man kan studere årsakssammenhenger og etablere kunnskap om hvilke insentiver som har betydning for å kunne påvirke reisemiddelvalg og realisere ønskede effekter i en eventuell fremtidig satsing på samkjøringsløsninger. Kunnskapsstatusen hadde også vist at resultater basert utelukkende på observerte effekter har begrenset overføringsverdi i seg selv, og det er viktig å beskrive årsaksvirkingskjeder for å forstå den fullstendige sammenhengen mellom innsats og gevinst.

Læringsperspektiv

Evalueringprosjektet hadde fokus på læring og kunnskapsbygging, i større grad enn å kun bedømme effektene av samkjøringstiltaket i Bergenspiloten. Det var viktigere å generere kunnskap om hvem som er potensielle samkjøringskandidater, forstå hvilke faktorer som kan bidra til ønsket effekt, samt påvise et realistisk markedspotensial i en fremtidig implementering - enn å avstedkomme vitenskapelige bevis for eventuelle observerte effekter i Bergen. Det var også knyttet usikkerhet til om man ville være i stand til å etablere et tilstrekkelig datagrunnlag for eksperimentelle analyser av reisemiddelvalg blant deltakerne.

Det ble utarbeidet en teoribasert evalueringsprosess, som illustrert i Figur 20.



Figur 20: Evalueringsprosessen i Samkjøringsprosjektet

Gjennomføringen av de syv trinnene i evalueringsprosessen er beskrevet i påfølgende kapitler.

7.3.1 PROBLEMSTILLINGER

ITS-tiltaket samkjøring adresserer flere utfordringer knyttet til køproblematikk, miljø og mobilitet. De relevante problemstillingene for evalueringsarbeidet ble definert og avgrenset med utgangspunkt i møter med oppdragsgiver, prosjektgruppen i Bergenspiloten og gjennomgang av dokumentene i prosjektet Spontan samkjøring (Gjerstad & Bayer, 2012; ITS Norge, 2011; Statens vegvesen, 2009). Hensikten med denne aktiviteten, var å få frem hvilke endringer prosjektet tok sikte på å oppnå, hvilke effekter som skal måles, samt identifisere hvilke årsakssammenhenger og prosesser som bør være i søkelyset i en evaluering.

Utgangspunktet for samkjøringspiloten var et ønske om løse konkrete rushtidsutfordringer på Kokstad/Sandsli, blant annet gjennom å påvirke brukeraksept for samkjøringstiltaket. De lokale initiativtakerne hadde i prosjektplanen formulert en målsetting om samkjøringsaktivitetene som trafikkdempende tiltak, ved at det skulle *bidra til å øke personbelegget i bil fra dagens nivå på 1,2 til 1,4 i 2030, hvorav minst halvparten av økningen skal være et resultat av færre bilister* (Statens vegvesen, 2009). Målet ble senere spesifisert i prosjektrapporten (ITS Norge, 2011): *Hovedmålet med prosjektet er å få flere til å samkjøre slik at trafikken og utslipp fra transportsektoren reduseres. Målet er færre biler på vegen. I dag foregår ca 12 000 arbeidsreiser med bil i området Sandsli-Kokstad-Flesland i Bergen. Målet er at samkjøringsprosjektet skal få redusert dette antallet til 9000 i løpet av en treårig testperiode* (ITS Norge, 2011). På Bergenspilotens nettside⁴⁷, ble det også understreket at prosjektets hovedmål var å redusere rushtrafikk og luftforurensning ved å utnytte kapasiteten i ledige bilseter. I samtaler med prosjektgruppen ble det imidlertid også lagt vekt på mobilitetsaspektet, ved at man hadde stort fokus på antall brukere av samkjøringsapplikasjonen – uavhengig av om disse var potensielle bilister eller syklist, fotgjengere og kollektivtrafikanter. Vegdirektoratets forventninger, som oppdragsgiver for evalueringsstudien, var i større grad knyttet til vurdering av ITS-tiltaket samkjøring på generell basis, med ønske om å avdekke tiltakets potensiale som trafikkreduserende tiltak i Norge. Dette innebar en vektlegging av rammebetingelser, suksesskriterier og eventuelle barrierer for vellykket innføring av tiltaket.

Tabell 23 oppsummerer evalueringsgrunnlaget som ble utarbeidet basert på problemstillinger og eksisterende kunnskap om samkjøringstiltaket.

⁴⁷ Ikke lenger tilgjengelig etter at prosjektet ble avsluttet

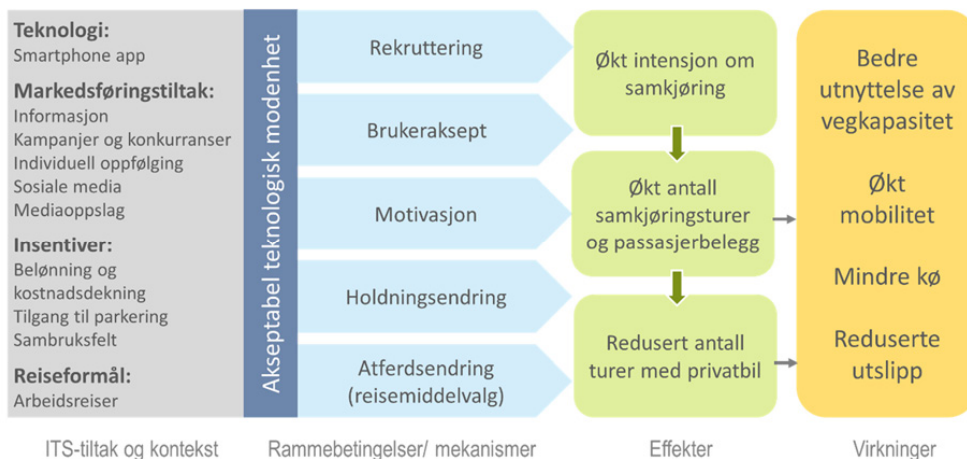
Tabell 23: Evalueringsgrunnlag for ITS-tiltaket samkjøring

Virkemidler	Forventet endring på kort sikt (effekt)	Ønsket konsekvens på lang sikt (virkning)	Mulige barrierer
Teknisk løsning: Samkjøringsapplikasjon Markedsføring: Informasjon Kampanjer og konkurranser Individuell oppfølging Sosiale media Mediaoppslag Insentiver: Økonomisk godtgjørelse Parkering Sambruksfelt	Positiv holdning, ønske om å benytte seg av samkjøring Endret reisemiddelvalg: <ul style="list-style-type: none"> • Mer samkjøring • Redusert bilbruk 	Økt kapasitetsutnyttelse/ personbelegg i bil Mindre kø Redusert lokal luftforurensning Redusert CO ₂ -utslipp Bedre transporttilbud/ økt mobilitet	Fleksibilitet Pålitelighet Komfort Sikkerhet Image/holdning Marginal økonomi Kritisk brukermasse Forretningsmodeller Personvern

Innholdet i tabellen beskriver tiltaket i form av de ulike virkemidler og insentiver som inngikk i Bergenspiloten, samt prosjektgruppens forventning om effekter og resultater, og mulige barrierer basert på erfaringer fra andre samkjøringsprosjekter. Denne oversikten danner grunnlaget for utvikling av tiltakets programteori i neste delkapittel.

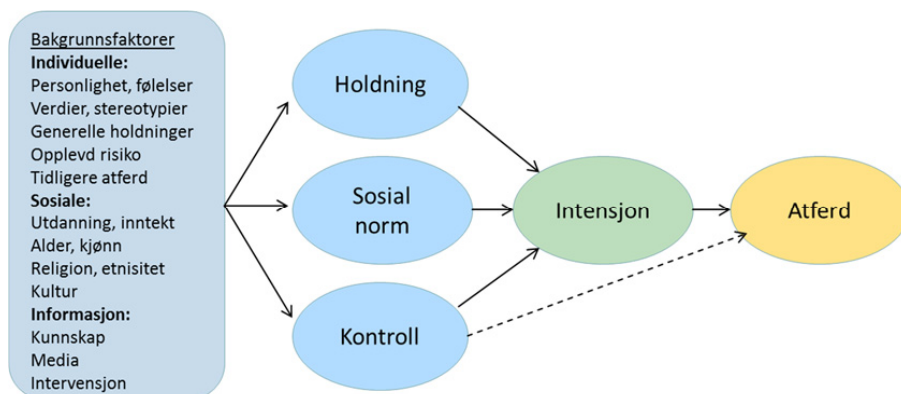
7.3.2 PROGRAMTEORI

ITS-tiltaket programteori ble utarbeidet som en konseptuell modell av sammenhengen mellom tiltakets virkemidler og forventede effekter. De relevante årsakssammenhengene ble identifisert med utgangspunkt i eksisterende kunnskapsgrunnlag (kapittel 7.2) og prosjektets problemstillinger (kapittel 7.3.1). Samkjøringstiltakets programteori er oppsummert og fremstilt i Figur 21.



Figur 21: Samkjøringstiltakets programteori

Programteorien er en sammenstilling av tiltakets endringsteori og implementeringsteori. Implementeringsteorien utgjør et logisk resonnement av hvilke betingelser som må være oppfylt for at tiltaket skal avstedkomme de endringene man ønsker å oppnå. Dette er illustrert ved de reaksjonene som skjer i boksene i figuren: Virkningene som oppnås på sikt er betinget av at tiltaket innføres i en gitt kontekst (grå boks), at teknologien oppnår akseptabel teknologisk modenhet, og at det realiseres en rekke påvisbare effekter (grønne bokser). Endringsteorien beskriver de underliggende mekanismer og kausale prosesser som sørger for at man oppnår ønsket effekter og virkninger. Dette er representert ved de blå pilene i figuren; rekruttering, brukeraksept, motivasjon, holdningsendring og atferdsendring. Samkjøringstiltakets endringsteori ble basert på Fishbein og Ajzens (2010) vitenskapsteoretiske atferdsmodell, som kan benyttes til å forklare og predikere sosial atferd, se Figur 22.



Figur 22: Forenklet versjon av the reasoned action model (Fishbein & Ajzen, 2010)

Denne modellen tar utgangspunkt i at atferd bestemmes av individets intensjoner (motivasjon), samt muligheten til å gjennomføre en spesifikk atferd (opplevd og faktisk selvkontroll). Intensjonene bestemmes igjen av individets holdninger, sosiale normer og opplevd selvkontroll. Disse faktorene påvirkes av individuelle og sosiale bakgrunnsfaktorer, samt individets kunnskap (informasjon) om positive og negative konsekvenser av en gitt atferd. Den relative betydningen av de ulike variablene varierer mellom situasjoner og ulike typer atferd, men erfaringer viser at faktoren *intensjon* har størst betydning, og *individets personlige holdning* har større betydning enn *den sosiale normen* (Ajzen, 1991). Denne vitenskapsteoretiske modellen er nyttig for å forstå i hvilken grad samkjøringstiltaket vil bidra til en eventuell atferdsendring, i dette tilfellet endring i reisemiddelvalg.

Programteorien med bakenforliggende endringsteori og implementeringsteori danner grunnlaget for etterfølgende empirisk testing av de antakelser og forventninger som ligger til grunn for tiltaket, og gir således en struktur for evalueringsstudien ved å identifisere sammenhenger og synliggjøre eventuelle hull i kunnskapsgrunnlaget.

7.3.3 EVALUERINGSKRITERIER

Evalueringskriteriene gjenspeiler formålet med evalueringen og angir hvordan man skal vurdere eventuelle nytteeffekter av tiltaket. I hovedsak kan evalueringskriteriene deles i tre kategorier:

- Operasjonelle kriterier: *Gjennomførte man det man skulle på en effektiv måte?*
- Taktiske kriterier: *Realiserte man de tiltenkte gevinstene for målgruppen?*
- Strategiske kriterier: *Var tiltaket relevant og nyttig i et samfunnsmessig perspektiv?*

I evalueringsstudien ble det lagt vekt på det taktiske perspektivet, og kriteriene ble valgt med utgangspunkt i forventninger om hva man skulle oppnå av resultater i Bergenspiloten. Dette er uttrykt i form av effekter og virkninger i tiltakets programteori (Figur 21). Tabell 24 gir en oversikt over definerte evalueringskriterier, med tilhørende hypoteser for samkjøringstiltaket.

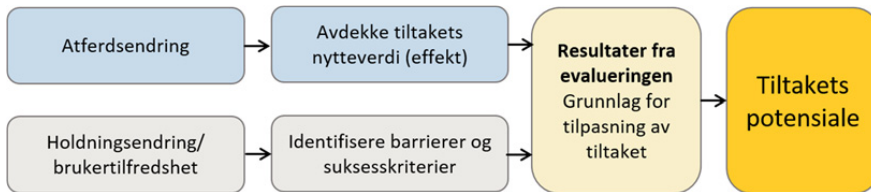
Tabell 24: Evalueringskriterier for samkjøringstiltaket

Evalueringskriterier	Grunnlag for hypoteser
<i>Bedre utnyttelse av vegkapasitet</i>	Økt personbelegg i bil - Mål om økning til 1,4 innen 2030
<i>Økt mobilitet</i>	Økt antall samkjøringsturer Opplevd forbedring i transporttilbudet (brukeraksept)
<i>Mindre kø</i>	Færre bilreiser, mål om 25 % reduksjon fra 12.000 til 9.000 ÅDT - Reisemiddelendring fra privatbil til samkjøring
<i>Reduserte utslipp</i>	Færre bilreiser, mål om 25 % reduksjon fra 12.000 til 9.000 ÅDT - Reisemiddelendring fra privatbil til samkjøring

Evalueringskriteriene og hypotesene er styrende for datainnsamlingen, og strukturerer i neste omgang evalueringens vurderinger og konklusjoner.

7.3.4 EVALUERINGSDESIGN

Evalueringsdesignet ble utformet med utgangspunkt i oppdragsgivers ønske om et kunnskapsgrunnlag for beslutning om hvorvidt (og eventuelt hvordan) det er formålstjenlig å utvikle og investere i samkjøring som trafikkdempende tiltak i det norske transportsystemet. Evalueringen inkluderte dermed to sentrale effekter; konkrete endringer i reisemiddelvalg og kartlegging av trafikantenes holdninger til ITS-tiltaket. Evalueringsformålet er illustrert i Figur 23.



Figur 23: Evalueringens formål for samkjøringstiltaket

Gjennom studier av samkjøringstiltakets faktiske effekter i Bergenspiloten, ønsket man å avdekke tiltakets nytteverdi med hensyn til de forhåndsdefinerte målsettingene. Supplerende kartlegging av holdninger og brukeraaksept hos deltakerne i pilotprosjektet skulle bidra til å identifisere relevante suksesskriterier og barrierer for vellykket implementering. Resultatene fra disse studiene utgjør et kunnskapsgrunnlag for tilpasning og optimalisering av en fremtidig implementering av ITS-tiltak for samkjøring.

Evalueringsdesignet for samkjøringstiltaket ble basert på en *realistisk evalueringsmetodikk* (Pawson, 2013; Pawson & Tilley, 1997) hvor programteoriens ulike bestanddeler testes med hensikt å finne de kombinasjoner av faktorer som har størst forklaringskraft (kontekst og endringsmekanismer). I praksis betyr dette at evalueringen hadde til hensikt å finne den kombinasjonen av virkemidler og situasjonsbestemte faktorer som er best egnet for å lykkes med en fremtidig implementering av samkjøringstiltak.

Samkjøringstiltakets evalueringsdesign er presentert i Tabell 25.

Tabell 25: Evalueringsdesign for samkjøringstiltaket

Virkning: Bedre utnyttelse av vegkapasitet		Metode
Endring/resultat:	Økt personbelegg (antall passasjerer) i bil	Logging av turdagbok (selvrapportert reiseatferd)
Betingelser:	Kritisk brukermasse (rekruttering)	Driftsstatistikk fra teknologileverandør og prosjektorganisasjon
	Brukeraksept, holdninger og motivasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
	Akseptabel teknologisk modenhet i applikasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
Virkning: Økt mobilitet		Metode
Endring/resultat:	Økt antall samkjøringsturer	Driftsstatistikk Logging av turdagbok (selvrapportert reiseatferd)
	Opplevd forbedring i transporttilbud (brukeraksept)	Kvantitativ spørreundersøkelse
Betingelser:	Kritisk brukermasse (rekruttering)	Driftsstatistikk fra teknologileverandør og prosjektorganisasjon
	Brukeraksept, holdninger og motivasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
	Akseptabel teknologisk modenhet i applikasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
Virkning: Mindre kø		Metode
Endring/resultat:	Færre bilreiser (25 %)	Logging av turdagbok (selvrapportert reiseatferd)
Betingelser:	Kritisk brukermasse (rekruttering)	Driftsstatistikk fra teknologileverandør og prosjektorganisasjon
	Brukeraksept, holdninger og motivasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
	Akseptabel teknologisk modenhet i applikasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
Virkning: Reduserte miljøutslipp		Metode
Endring/resultat:	Færre bilreiser (25 %)	Logging av turdagbok (selvrapportert reiseatferd)
Betingelser:	Kritisk brukermasse (rekruttering)	Driftsstatistikk fra teknologileverandør og prosjektorganisasjon
	Brukeraksept, holdninger og motivasjon	Kvantitativ spørreundersøkelse Kvalitative intervju
	Akseptabel teknologisk modenhet i applikasjon	Kvantitativ spørreundersøkelse Kvalitative intervju

Den metodiske fremgangsmåten kombinerer eksperimentelle analyser av reisemiddelvalg basert på logging av turdagbok blant deltakerne i piloten, med teoribaserte analyser av kausale årsakssammenhenger basert på spørreundersøkelse i målgruppen. Detaljer og vurderinger som ligger til grunn for datainnsamlingen er nærmere beskrevet i neste delkapittel.

7.3.5 DATAINNSAMLING

Det ble gjennomført en omfattende kvantitativ og kvalitativ datainnsamling basert på spørreundersøkelse blant alle ansatte i bedriftene som deltok i Bergenspiloten (N=1170), driftsstatistikk fra samkjøringsapplikasjonen og prosjektorganisasjonen, samt intervju og møter med bedriftskontakter og samkjøringsambassadører. I tillegg ble data fra den lokale reisevaneundersøkelsen i Bergensområdet (Meland & Nordtømme, 2014) brukt som referanse eller sammenligningsgrunnlag der hvor dette var relevant. Tabell 26 viser en oversikt over datainnsamlingen som ble foretatt i evalueringsstudien.

Tabell 26: Datainnsamling i evalueringsstudien av Samkjøringstiltaket

Metode	Omfang	Formål
Spørreundersøkelse blant arbeidstakere i området	Internettbasert spørreundersøkelse blant ansatte i åtte deltakerbedrifter. Gjennomført i januar 2014 med N=1170. Antall spørsmål og detaljnivå var tilpasset svarene som ble gitt underveis, og tidsforbruk til utfylling av skjema varierte i størrelsesorden 6-20 min.	Selvrapporterte reisevaner, inkl logging av turdagbok i tre dager, kartlegging av holdninger, motivasjonsfaktorer, brukerksept og bakgrunnsinformasjon.
Driftsstatistikk fra samkjøringsapplikasjon	Statistikk om antall brukere og gjennomførte samkjøringsturer i perioden oktober 2012 - juni 2014.	Kunnskap om gjennomførte samkjøringsturer og hvordan tjenesten brukes blant deltakerne.
Driftsstatistikk fra pilotprosjektet	Kartlegging av rekruttering og markedsføringstiltak i perioden oktober 2012 t.o.m. februar 2014.	Analyser av ressursbruk til drift og koordinering av samkjøringstiltaket. Kartlegging av insentiver.
Intervju og samtaler med bedriftskontakter og samkjøringsambassadører	Innspill basert på bedriftenes erfaringer, gjennom intervjuer (N=3), og to møter med kontaktgruppen for alle bedriftene.	Brukeraksept og erfaringer fra deltakerne. Kartlegging av insentiver.

Spørreundersøkelsen ble gjennomført i åtte av deltakerbedriftene i Bergenspiloten i januar 2014. Undersøkelsen ble distribuert gjennom bedriftsinterne eposter og nettsider til anslagsvis 7000 ansatte, og studien mottok til sammen 1170 besvarelser. Antall spørsmål og detaljnivå ble tilpasset svarene som ble gitt underveis, slik at de med mest erfaring fra samkjøring også mottok flere og mer omfattende spørsmål. Besvarelsene ble gruppert etter grad av kjennskap til og involvering i samkjøringsaktivitetene, og disse grupperingene dannet senere grunnlag for sammenligning mellom ulike målgrupper i analysene; a) *de som ikke hadde fått (eller oppfattet) informasjon om samkjøringstjenesten*, b) *de som hadde fått informasjon, men ikke meldt seg på*, c) *de som hadde meldt seg på, men ikke benyttet seg av samkjøring*, d) *de som hadde prøvd samkjøring en eller to ganger, men sluttet*, e) *de som hadde prøvd samkjøring flere ganger, men sluttet*, og til sist f) *de som fortsatt samkjører*. Undersøkelsen omfattet følgende tema:

- Hvem benytter seg av samkjøringstjenesten? Bakgrunn og representativitet, problemforståelse og aksept for transportiltak, teknologierfaring og forhold til personvern
- Reiseatferd, reisemiddelvalg, tilgang til transportalternativer og eksterne bindinger
- Motivasjonsfaktorer og årsaker for å slutte med samkjøring
- Erfaring, brukeraksept, betalingsvilje
- Vurdering av insentiver

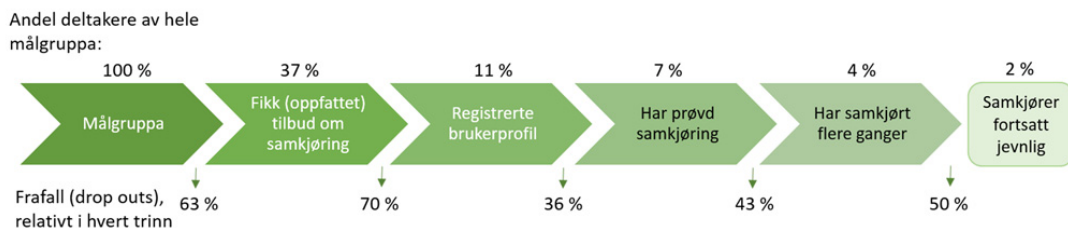
Driftsstatistikk fra samkjøringsapplikasjonen ble innhentet for perioden oktober 2012 – juni 2014. Starttidspunktet sammenfalt med lansering av en ny og oppdatert versjon av mobil-appen, samtidig som det ble satt i gang en høstkampanje for å øke rekruttering og samkjøring blant de ansatte på Kokstad/Sandsli. Datagrunnlaget inkluderte deltakernes rekrutteringstidspunkt, kjønn og bedriftstilknytning, mens samkjøringsturene er registrert med henholdsvis fører og passasjer, tidspunkt, sted, turens varighet og distanse, beregnet sjåførgodtgjørelse/passasjerkostnad, samt karaktervurdering fra brukerne. Analysegrunnlaget utgjorde etter kvalitetssikring totalt 7172 turer, hvorav turer med mer enn en passasjer telles flere ganger (for hver passasjer). Datamaterialet inneholdt ikke informasjon om forespørsler som ikke resulterte i en faktisk kobling mellom fører og passasjer. Driftsstatistikken fra pilotprosjektet inneholdt informasjon om ressursbruk til drift og koordinering av samkjøringsaktivitetene. Dette ga grunnlag for å se på sammenhengen mellom prosjektledelsens innsats i forbindelse med rekruttering og markedsføringstiltak, og utslag på antall deltakere og bruk av samkjøringstjenesten gjennom prosjektperioden. Data ble samlet inn fra en periode på 17 måneder, fra oktober 2012 til og med februar 2014. Det ble også gjennomført kvalitative undersøkelser av brukernes erfaringer og aksept, gjennom intervjuer med tre utvalgte bedriftsambassadører og samtaler med kontaktpersoner fra deltakerbedriftene gjennom prosjektperioden. Dette datagrunnlaget ble blant annet benyttet til tolkning og utfylling av resultater og indikasjoner fra de øvrige analysene.

7.3.6 ANALYSER

Rekruttering av deltakere og kritisk brukermasse

Dersom samkjøringstiltaket skal forbedre trafikantenes mobilitet, forutsetter det at man greier å rekruttere tilstrekkelig antall brukere til at man har en realistisk mulighet for kobling mellom sjåfører og passasjerer. Skal tiltaket også få betydning for kapasitetsutfordringene i vegnettet, må man i tillegg oppnå en merkbar reduksjon i antall biler på vegen. Analysene viste at det var vanskelig å oppnå en vesentlig brukermasse. Trafikantene som benytter samkjøringstjenesten må igjennom en rekrutteringsprosess, som inkluderer å få kjennskap til tilbudet, fatte beslutninger om å delta, samt å ta aktivt del i samkjøringen. Spørreundersøkelsen avdekket at hele 63 % av

respondentene ikke kunne huske å ha fått informasjon om eller tilbud om å delta i samkjøringsaktivitetene. Dette til tross for massive markedsføringstiltak både fra prosjektgruppen og innad i bedriftene. Figur 24 synliggjør i hvilken grad respondentene hadde erfaring med samkjøringstiltaket, og ved hvilke terskler de falt ut av rekrutteringsprosessen.



Figur 24: Rekrutteringsprosessen for samkjøringstiltaket fra deltakerne i Bergenspiloten.
N=1007 ansatte i bedrifter på Kokstad/Sandsli.

Mens 37 % oppga at de hadde fått informasjon om samkjøringsaktivitetene, var det kun 11 % som registrerte seg med brukerprofil for å delta i samkjøringsaktivitetene, og kun 7 % som kom så langt at de faktisk hadde samkjørt. Av disse sluttet nær halvparten med samkjøring etter et par turer, mens 4 % oppga at de hadde samkjørt flere ganger. Kun 2 % av deltakerne i spørreundersøkelsen svarte at de fortsatt samkjører jevnlig (ikke nødvendigvis daglig) – og basert på driftsstatistikken ser det ut som spørreundersøkelsen har fanget opp de som faktisk benytter seg av samkjøringstjenesten. Det relative bortfallet var størst ved de to første tersklene, med 63 % av den potensielle deltakermassen som ikke oppfattet tilbudet om samkjøring, mens en tilsvarende andel av de som fikk (og oppfattet) tilbudet lot være å melde seg på samkjøringsaktivitetene (70 %). Dersom man kun baserer analysene på de respondentene som oppga at de hadde fått informasjon og tilbud om å delta, er det fortsatt en forholdsvis beskjeden andel deltakere som samkjører på jevnlig basis (tilsvarer 6 %). Disse resultatene underbygges av bruksdata fra samkjøringsapplikasjonen og driftsstatistikken fra prosjektgruppen. Analysene av data fra samkjøringsapplikasjonen viste at i alt 276 samkjørere⁴⁸ var involvert i totalt 7172 samkjøringsturer, enten som fører eller passasjer, over en periode på vel halvannet år⁴⁹. 25 % av samkjørerne var registrert med kun 1-2 turer, mens gjennomsnittlig aktivitet var på 52 samkjøringsturer per person i løpet av perioden. Over halvparten av turene besto alltid av samme sjåfør og passasjer, og kan ansees som forholdsvis faste avtaler om samkjøring. Driftsdata fra Bergenspiloten viste at man totalt lyktes med å rekruttere 1259 deltakere i samkjøringsaktivitetene (antall som registrerte seg med brukerprofil) og at det ble utført 7034

⁴⁸ Dette utgjør anslagsvis 3,9 % av de ansatte i bedriftene som deltok i Bergenspiloten

⁴⁹ Bruksdata fra uke 41/2012 - uke 24/2014.

samkjøringsturer over en periode på 17 måneder⁵⁰, med et snitt på 414 turer per måned. I slutten av registreringsperioden lå samkjøringsaktiviteten forholdsvis stabilt på 0,4 samkjøringsturer i måneden per registrerte deltaker. Analysene ga indikasjoner på at samkjøringstjenesten var mest attraktiv for yngre aldersgrupper, kvinner og personer med lang utdanning, og at det var mindre aktuelt for småbarnsforeldre å delta i samkjøringsaktivitetene.

Reiseatferd og reisemiddelvalg

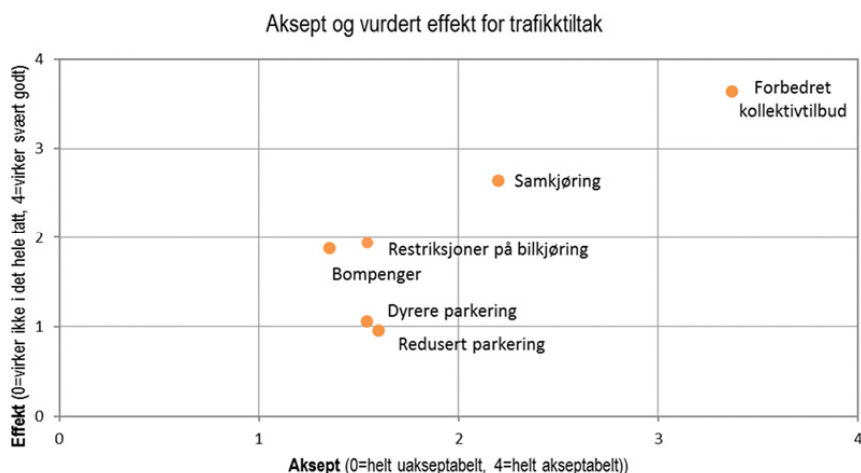
De ansatte i bedriftene i Bergenspiloten ble bedt om å beskrive reisemiddelbruk til og fra arbeid for tre påfølgende arbeidsdager, som del av spørreundersøkelsen. Dette ga informasjon om reisemiddelbruk for 5412 arbeidsreiser på Kokstad/Sandsli. Analysene viste at bil var dominerende transportform, og bilen ble brukt i tre av fire arbeidsreiser i området, enten som sjåfør (67 %) eller passasjer (7 %). Kollektivtransport ble benyttet i 12 % av arbeidsreisene, og ellers fordelte reiseaktiviteten seg forholdsvis likt på kategoriene sykkel (7 %) og fotgjenger (6 %). Dette samsvarte i hovedsak med resultatene fra reisevaneundersøkelsen i området (Meland & Nordtømme, 2014). Passasjerandelen var naturlig nok noe høyere hos respondentene som oppga at de aktivt benyttet seg av samkjøringstjenesten, og i denne gruppen hadde man et personbelegg per kjøretøy på 1,35 (mot 1,13 hos resten av utvalget). De aktive samkjørerne hadde også en høyere andel reiser som kollektivpassasjer. Samkjøringsapplikasjonen ble benyttet i forbindelse med 8 % av de rapporterte bilførerturene med passasjer, og på 8 % av bilpassasjerturene. På spørsmål om hva som ville vært alternativ reisemåte på samkjøringsturene, svarte samtlige av sjåførene at de ville fortsatt å kjøre bil, mens de fleste passasjerene ville reist kollektivt (58 %) eller fortsatt som passasjer ved å gjøre avtaler på annet vis (38 %). Videre rapporterte gruppen med aktive brukere av samkjøringstjenesten at tiltaket ikke hadde forårsaket endringer i hvor ofte de kjørte bil til arbeid. Dette kan tyde på at forskjellene i reisemiddelvalg skyldes bakenforliggende faktorer som er uavhengig av samkjøringstiltaket og/eller at deltakerne ble rekruttert fra andre transportmidler enn bil.

Holdninger

Analysen av datamaterialet fra spørreundersøkelsen ga indikasjoner på at deltakernes erfaringsgrunnlag, problemforståelse og holdninger påvirket hvordan de forholdt seg til samkjøringstiltaket. De som hadde brukerprofil i samkjøringsapplikasjonen hadde mer teknologierfaring og var mer entusiastisk overfor bruk av digitale tjenester og teknologiske løsninger, enn de som ikke meldte seg på samkjøringsaktivitetene. Den aktive gruppen var også minst bekymret for personvern og ga uttrykk for at de stolte på at innsamlede data ble forsvarlig

⁵⁰ Driftsstatistikk fra oktober 2012 – februar 2014.

behandlet. Videre ble luftforurensning vurdert som et mer alvorlig problem av de som deltok i samkjøringsaktivitetene, enn av øvrige respondenter. Samkjørerne hadde også større aksept for generelle tiltak mot trafikkrelaterte problemer. I Figur 25 illustreres respondentenes tiltro til og aksept for ulike tiltak for å redusere biltrafikken.



Figur 25: Vurdering av generelle trafikkreduserende tiltak (N=876)

Som figuren viser, hadde respondentene størst tiltro til effekten av *et forbedret kollektivtilbud*, og det var også dette tiltaket som oppnådde høyest brukeraksept. *Samkjøring* fikk nest høyeste score. Likevel var den rapporterte betalingsviljen for samkjøring lav, og ingen ville betale like mye for samkjøring som det koster å reise med øvrig kollektivtransport. De øvrige tiltakene som ble vurdert var *bompenger*, *reduisert antall parkeringsplasser*, *dyrere parkering* og *restriksjoner på bilkjøring ved bestemte værforhold*.

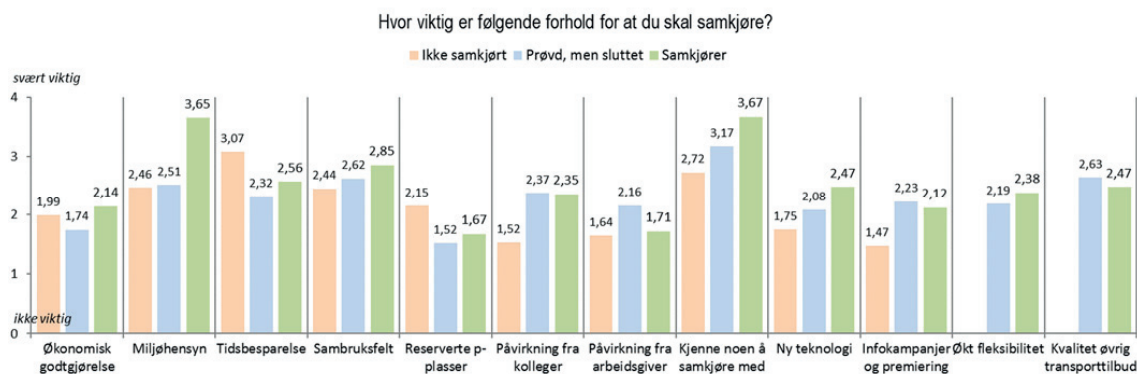
Valgmuligheter og selvkontroll

Forhold som påvirker de ansattes fleksibilitet med hensyn til arbeidstid og faste forpliktelser på arbeidsreisen, vil kunne påvirke reisemiddelvalg (Meland, 2012). Spørreundersøken avdekket at de fleste ansatte på Kokstad/Sandsli (77 %) hadde fleksibel arbeidstid, og nesten like mange var uten faste forpliktelser (67 %) knyttet til arbeidsreisen. Analysene bekreftet at frihet i arbeidstidsordning hadde en positiv betydning for de ansattes involvering i samkjøringsaktivitetene, og hele 90 % av de aktive samkjørerne hadde fleksibel arbeidstid. Daglige forpliktelser som medførte stopp underveis på arbeidsreisen (f.eks. henting i barnehage) viste seg ikke uventet å være vanskelig forenelig med samkjøring. Arbeidstakerne hadde i gjennomsnitt 18 km reiseavstand mellom hjem og arbeidsplass, og samkjøring ble oppfattet som minst aktuelt for de med lengst reisevei. Majoriteten av

respondentene hadde førerkort for bil (97 %), og disponerte i snitt 1,5 biler per husstand. Analysene viste at samkjøring ble vurdert som et konkurransedyktig transporttilbud av dem som hadde liten eller ingen mulighet til å kjøre bil selv, men var mindre attraktivt for dem med god tilgang på bil. Arbeidstakere som anså kollektivtransport som et reelt alternativ på arbeidsreisen, var også mer tilbøyelig til å velge samkjøring som reisemåte.

Motivasjonsfaktorer

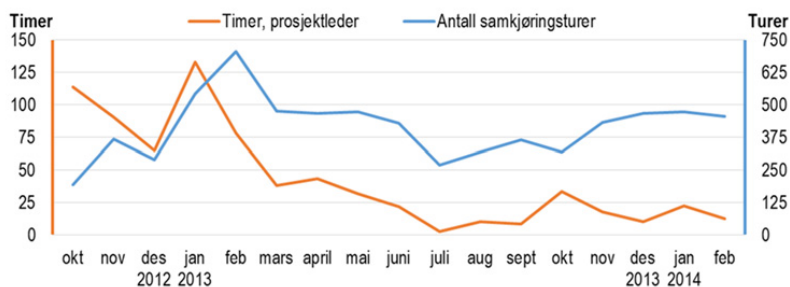
Deltakerne i spørreundersøkelsen fikk flere spørsmål knyttet til motivasjonsfaktorer og betydningen av en rekke insentiver i de ulike fasene av rekrutteringsprosessen. Analysene viste at de viktigste faktorene ved påmelding var en forventning om tidsbesparelse på arbeidsreisen og det at man allerede kjente noen å samkjøre med. En oversikt over motivasjonsfaktorer er presentert i Figur 26, som viser en sammenstilling av resultater for følgende grupperinger; i) de som ikke har samkjørt, ii) de som har prøvd samkjøring, men sluttet, og iii) de som fortsatt samkjører på jevnlig basis.



Figur 26: Motivasjonsfaktorer for deltakelse i samkjøringsprosjektet, rangert på en skala fra 0 (ikke viktig) til 4 (svært viktig). N=949. Spørsmål knyttet til alternativene "økt fleksibilitet" og "kvalitet på annet transporttilbud" ble kun stilt til de som hadde erfaring med samkjøring.

Tidsbesparelse ble utpekt som den viktigste faktoren for påmelding i den gruppen som ikke kom så langt at de faktisk prøvde samkjøring. De som deltok i samkjøringsaktivitetene, oppga det å *kjenne noen å samkjøre med* som viktigste motivasjon for påmelding. I denne gruppen var også det å få tilgang til sambruksfelt et viktig insentiv. Påvirkning fra kolleger, arbeidsgiver og kampanjer hadde større betydning for de som sluttet etter et par forsøk, enn for de som fortsatte å samkjøre. Miljøhensynet var også en vesentlig motivasjonsfaktor for påmelding blant de som fortsatte med samkjøring på jevnlig basis. Økonomisk godtgjørelse ser ut til å ha hatt liten betydning for deltakelsen i samkjøringsaktivitetene. Det kan skyldes at den økonomiske faktoren ikke var viktig for brukerne i Bergenspiloten, men det kan muligens også forklares med at den reelle godtgjørelsen

var for liten til å påvirke deltakernes atferdsmønster. Det ble imidlertid avdekket en klar sammenheng mellom prosjektorganisasjonens innsats i form av informasjonsarbeid og kampanjer, og rekruttering av brukere og utførte samkjøringsturer. Sammenhengen mellom timeinnsats fra prosjektleder og samkjøringsaktivitet gjennom pilotperioden er illustrert i Figur 27.



Figur 27: Ressursinnsats og aktivitetsnivå for samkjøring i Bergenspiloten (Meland et al., 2015).

Motivasjonen for å opprettholde samkjøringsaktivitetene over tid ble i stor grad begrunnet i *tilgang til sambruksfelt, tidsbesparelse på arbeidsreisen og kampanjer med premiering av deltakelse*. De som sluttet med samkjøring underveis i pilotperioden, oppga problemer knyttet til teknologien, vanskelig å få treff på forespørsler (finne samkjører) og manglende fleksibilitet i hverdagen som viktigste årsaker til at de sluttet med samkjøring. Dette er illustrert i Figur 28. Samtalene med bedriftsambassadørene ga også signal om at teknologien ble oppfattet som forholdsvis umoden, og at dette hadde stor betydning for frafall allerede før man kom i gang med samkjøringen.

Hvor viktig var følgende forhold for at du sluttet med samkjøring?



Figur 28: Årsaker til å slutte med samkjøring, rangert på en skala fra 0 (ikke viktig) til 4 (svært viktig). Utvalget består av alle som har prøvd samkjørings tjenesten, men sluttet etter et eller flere forsøk (N=48).

Evalueringen viste at de viktigste motivasjonsfaktorene for å delta i samkjøringsaktivitetene var å *kjenne noen å samkjøre med* og å *få tilgang til sambruksfelt (tidsbesparelse)*. Forutsetninger for å lykkes med samkjøringstiltaket inkluderer en teknologisk moden samkjøringsapplikasjon, aktiv prosjektorganisasjon og nok brukere til at løsningen blir et fleksibelt og reelt transporttilbud. Detaljerte analyser av flere motivasjonsfaktorer og insentiver for ulike grupper av deltakere er presentert i Meland et al. (2015) og Lervåg og Meland (2014).

Analysen av potensiale for atferdsendring

Med utgangspunkt i erfaringene fra Bergenspiloten, ble det gjort en analyse av hva som ville vært det maksimale potensialet for tiltaket, gitt at man nådde ut til alle ansatte i rekrutteringsprosessen. Estimaten er basert på informasjon fra spørreundersøkelsen med logging av reiseatferd for 5412 arbeidsreiser, og det er tatt hensyn til en bilførerandel på 67 %⁵¹, og at omtrent 1/3 av de som har oppfattet tilbudet om samkjøringstjenesten har valgt å delta i aktivitetene. Videre er det lagt til grunn at arbeidsreiser med faste daglige forpliktelser i hovedsak vil foregå som bilfører. I alt anses i overkant av 80 % av alle arbeidsreisene å være uaktuell for samkjøring, og dersom man legger reiseatferd/reisemiddelfordeling hos den aktive samkjøringsgruppen til grunn for de resterende arbeidsreisene, får man en reduksjon i antall bilførerturer med 57 turer i løpet av de tre aktuelle arbeidsdagene, noe som gir en relativ endring på 1,6 %. Det må bemerkes at utvalget som samkjører på regelmessig basis er forholdsvis lite (21 personer), og datagrunnlaget anses derfor som svært usikkert. Samtidig er dataene samlet inn i en situasjon med forholdsvis lav brukermasse, hvor det er begrenset sannsynlighet for å få treff på forespørsler om samkjøringsturer. Dersom man oppnådde en større brukermasse, ville brukerne sannsynligvis opplevd mer nytte av tiltaket, noe som kunne resultert i større trafikale effekter av samkjøringen. Hadde man hatt et større datagrunnlag ville det også vært hensiktsmessig å sett på alternativ bruk av det kjøretøyet som ikke ble benyttet på arbeidsreisen.

7.3.7 RESULTATER

Resultatene fra analysen tyder på at samkjøringstiltaket i liten grad har oppfylt målsettingene som ble lagt til grunn for aktivitetene i Bergenspiloten. Rekrutteringsprosessen har vært krevende, og man har ikke lykkes med forutsetningen om å tiltrekke seg en stor nok brukermasse. Tabell 27 gir en oversikt over resultatene fra evalueringsstudien.

⁵¹ Tallet er hentet fra selvrapporterte reisevaner i spørreundersøkelsen

Tabell 27: Oppsummering av evalueringsresultater

Forventet resultat/ endring	I hvilken grad ble effektene bekreftet?
<i>Økt personbelegg i privatbiler</i>	De aktive samkjørings sjåførene hadde i snitt tre ganger så mange passasjerer som øvrige bilførere, og oppnådde et personbelegg i sine kjøretøy på 1,35 mot 1,13 for øvrige bilførere. Dette er ikke så langt unna det definerte målet for 2030 (1,4 per kjøretøy), og indikerer at samkjøringsaktiviteter i prinsippet kan bidra til å nå målsettingen, vel og merke innenfor en avgrenset målgruppe. Datagrunnlaget er imidlertid nokså begrenset, da gruppen med aktive samkjørere kun utgjør 21 personer (2% av datagrunnlaget).
<i>Økt antall samkjøringsturer</i>	Rekruttering av deltakere ser ut til å ha vært en krevende prosess, og til tross for vesentlig innsats på drift og markedsføring av tiltaket, har man kun nådd ut med budskapet til drøyt 1/3 av målgruppen. I løpet av pilotperioden har man oppnådd 1259 registrerte deltakere, men langt de fleste har ikke kommet lengre enn å registrere en brukerprofil i samkjøringsapplikasjonen. Samkjøringsturene er gjennomført av 276 unike brukere, hvorav ¼ kun har registrert 1-2 turer, og omtrent halvparten synes å ha faste avtaler, hvor turene gjennomføres jevnlig med samme sjåfør og passasjer. Totalt har man registrert i overkant av 7000 samkjøringsturer i løpet av Bergenspiloten, hvilket tilsvarer ca 400 turer per måned, 20 turer per arbeidsdag, og 0,4 turer i måneden per påmeldte deltaker. Tiltaket synes ikke å ha nådd en kritisk brukermasse, da brukerne rapporterer om for dårlig respons (treffrate) på forespørsler om samkjøring.
<i>Opplevd forbedring i transporttilbudet</i>	Det faktum at samkjøringstiltaket kun har hatt marginal effekt på transportatferden, tyder på at innsatsen i liten grad har påvirket deltakernes mobilitet. En avgrenset målgruppe bestående av aktive samkjørere rapporterer imidlertid om positive brukeropplevelser, og det kan synes som om tiltaket bidrar til et forbedret transporttilbud for de som har liten tilgang til egen bil i arbeidsreisen. Betalingsviljen for samkjøring synes imidlertid å være lavere enn kostnadene forbundet med å reise med annen kollektivtransport.
<i>Færre bilreiser</i>	Samkjøringstiltaket har ikke bidratt til påvisbar reduksjon i antall bilreiser i området. Basert på driftsstatistikken til samkjøringsapplikasjonen, har man i løpet av prosjektperioden oppnådd et snitt på 20 samkjøringsturer per arbeidsdag. Gitt en bilførerandel på 67 % blant de ansatte på Kokstad/Sandsli, utgjør dette kun 0,2 % av arbeidsreisene med bil i området ⁵² . Resultatene fra spørreundersøkelsen gir i tillegg grunn til å tro at de fleste samkjøringsturene i realiteten erstatter kollektivreiser. Flertallet av de som aktivt benytter samkjøringstjenesten, oppgir også selv at dette ikke har medført en endring i hvor ofte de kjører bil til arbeid. Analysene viser et maksimalt potensial på 1-2 % reduksjon i biltrafikken, gitt dagens nivå på reiseatferd i området. Dette er langt ifra den definerte målsettingen om 25 % reduksjon i løpet av pilotperioden.

Evalueringen har identifisert en rekke suksesskriterier og barrierer som har betydning for i hvilken grad man vil kunne lykkes med fremtidige implementeringer av samkjøringstiltak i transportsystemet. De viktigste motivasjonsfaktorene for rekruttering av brukere er å legge til rette for at deltakerne kjenner noen å samkjøre med⁵³. Dette har vesentlig betydning for de som fortsetter

⁵² Utgangspunkt i 7000 ansatte, 2 arbeidsreiser per dag (til/fra jobb) og 67 % bilførerandel i henhold til turdagbok i spørreundersøkelsen.

⁵³ I en av bedriftene som deltok i Bergenspiloten ble dette gjort ved å utarbeide beboelseskart for de ansatte, der kollegaer som bodde i samme område kunne få utdelt kontaktinformasjon til hverandre etter samtykke.

med samkjøring over tid, noe som også underbygges av Amey et al. (2011) som viste at kun 3 % av samkjøringen foregår mellom fremmede personer. I rekrutteringsprosessen fremstår tidsbesparelse på arbeidsreisen og miljøhensyn som de viktigste salgsgargumentene for å overbevise folk til å prøve samkjøring. Resultatene tyder på at de som deltar i samkjøringsaktivitetene i dag, til en viss grad motiveres av egen samvittighet (idealisme) og miljøargumentet. Det viktigste insentivet fra omgivelsene, synes å være tilgang til sambruksfelt. Erfaringene fra Bergenspiloten tyder på at forutsigbarhet er viktigere enn spontanitet. De fleste samkjøringsturer blir avtalt på forhånd, før appen deretter brukes til å administrere gjennomføringen av turen. Dette påvirkes sannsynligvis også av at man ikke har lyktes å opprette en kritisk brukermasse, så sannsynligheten for å få treff på spontane forespørsler om samkjøring oppleves å være lav. Bare 8 % av arbeidsreisene med passasjer til/fra bedriftene i Bergenspiloten involverer bruk av samkjøringsapplikasjonen, noe som indikerer at de fleste som samkjører allerede har en relasjon og de har ikke nødvendigvis behov for den teknologiske løsningen for å samkjøre. De største barrierene mot samkjøring er, i tillegg til manglende brukermasse, knyttet til manglende fleksibilitet i form av at samkjøring ikke er forenelig med daglige forpliktelser knyttet til arbeidsreisen. Videre blir umoden teknologi (applikasjonen fungerer for dårlig) oppgitt å være til hindring for rekruttering og bruk av samkjøringstjenesten. Evalueringen har fremskaffet erfaringer med og kunnskap om en rekke av forventningene og betingelsene som ble lagt til grunn i utviklingen av samkjøringstiltakets programteori i kapittel 7.3.2. De viktigste suksesskriteriene for bruk av samkjøring er oppsummert sammen med vurderingen av potensielle virkninger i Tabell 28.

Tabell 28: Suksesskriterier for innføring av tiltaket

Kontekst	Mekanismer	Virkning
<p>Bakgrunnsfaktorer: Yngre aldersgrupper, kvinner, ikke småbarnsforeldre, høyere utdanning, teknologierfaring</p> <p>Insentiver: Tilgang til sambruksfelt</p> <p>Rammebetingelser: Moden teknologisk løsning Dekning av reisekostnader Fleksibilitet i arbeidstid</p>	<p>Rekrutteringsprosessen: Aktiv prosjektorganisasjon</p> <p>Brukeraksept, holdninger og motivasjon for deltakelse: Kjenne noen å samkjøre med Tidsbesparelse Miljøhensyn Tillit til personvern Teknologioptimisme</p> <p>Endring i reiseatferd: Begrenset biltilgang Ikke daglige forpliktelser</p>	<p>Potensielle virkninger i et optimalt scenario: Passasjerbelegg på 1,35 Redusert biltrafikk på 1-2 %.</p>

7.4 WHAT IF? POTENSIELLE RESULTATER FRA EN EKSPERIMENTELL EVALUERING

Dersom man i evalueringsstudien hadde valgt et rent eksperimentelt evalueringsdesign, hadde datainnsamling og analyser blitt konsentrert rundt måling og beregning av atferdsendringer i form av redusert biltrafikk, og man ville hatt et sammenligningsgrunnlag i den lokale reisevaneundersøkelsen som ble gjennomført i 2013 (Meland & Nordtømme, 2014). Dette ville gitt klare og valide resultater; en reduksjon av antall turer som bilfører i størrelsesorden 0,2 % er langt fra det definerte målet om 25 % nedgang i biltrafikk som følge av samkjøringsaktivitetene i Bergenspiloten. Det er åpenbart at tiltaket ikke har realisert ønskede effekter, og det er trolig langt igjen til samkjøring vil ha noen merkbar effekt som trafikkdempende tiltak.

Litteraturgjennomgangen viste at samkjøringstiltak ikke kan forventes å gi et tilstrekkelig datagrunnlag for statistiske beregninger i eksperimentelle analyser, da deltakelsen generelt synes å være svært beskjeden. Slik var tilfellet også i dette case-studiet, hvor man med en målgruppe på 7000 personer, kun oppnår 21 personer som fortsetter å samkjøre over tid. Den eksperimentelle modellen har liten forklaringskraft når det gjelder årsakssammenhenger, og den ville ikke gitt kunnskap om hvilke bakgrunnsfaktorer, rammebetingelser, insentiver og motivasjonsfaktorer som påvirker effektene av tiltaket.

Den teoribaserte tilnærmingen har bidratt til økt forståelse for hvorfor tiltaket ikke utløser de ønskede effekter, og den har identifisert suksesskriterier og barrierer som må overvinnes dersom man ønsker å videreutvikle og ta i bruk tiltaket i fremtiden. Videre har bruk av teoribaserte (realistiske) evalueringsmetoder i analysene bidratt til å identifisere hvilken kombinasjon av målgruppe og situasjonsspesifikke faktorer som er best egnet til å lykkes med tiltaket – samt gi en indikasjon på tiltakets potensiale under optimale forutsetninger. Dette er relevant kunnskap for beslutninger om fremtidig satsing på samkjøringstiltak. Til sist har også utviklingen av programteorien som visuell modell utgjort en felles referanseramme i kommunikasjonen med interessenter og involverte aktører, noe som var nyttig i avklaringer rundt hvorvidt samkjøringstiltaket var et rent trafikkdempende tiltak eller også et mobilitetstiltak.

7.5 KONKLUSJONER FRA EVALUERINGEN

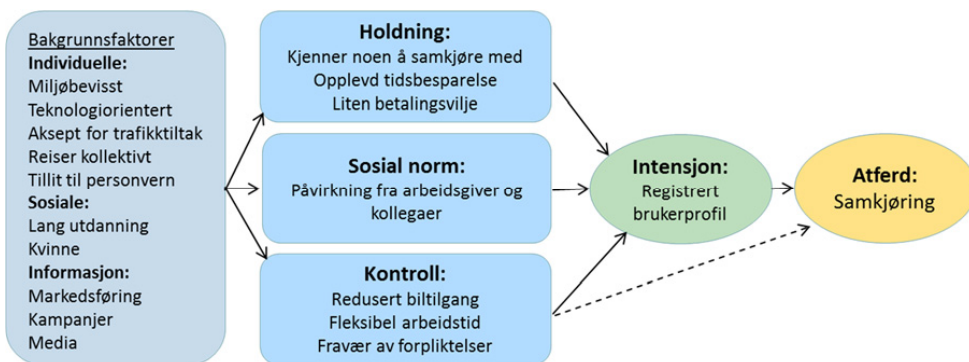
Evalueringen har avdekket at teknologibasert og spontan samkjøring i liten grad synes å ha potensiale som etterspørselsstyrende og trafikkdempende tiltak i Norge per i dag. Det er en stor utfordring å rekruttere tilstrekkelig mange brukere til at tjenesten blir fleksibel og forutsigbar nok til å fungere som et reelt transporttilbud. Videre er det mye som tyder på at tjenesten i hovedsak rekrutterer sine brukere fra kollektivtransport, og således har marginal betydning for biltrafikken.

Tabell 29 gir en oversikt over konklusjonene fra evalueringsarbeidet, relatert til evalueringskriteriene som ble definert i kapittel 7.3.3.

Tabell 29: Evalueringskriterier og konklusjoner fra evalueringen av samkjøring

Evalueringskriterier og konklusjoner	
Bedre utnyttelse av vegkapasiteten	<i>Evalueringen gir liten grunn til å tro at samkjøringstiltaket vil ha noen merkbar effekt i transportsystemet generelt, men analysene viste at tiltaket potensielt kan ha en viss effekt på personbelegget i privatbiler - i en avgrenset målgruppe. De som deltar aktivt i samkjøringsaktivitetene har i snitt tre ganger så mange passasjerer som øvrige bilførere, med et personbelegg i Bergenspiloten på 1,35 for samkjørings sjåfører.</i>
Økt mobilitet	<i>Evalueringen viser at samkjøringstiltaket kun i begrenset grad bidrar til økt mobilitet. Svært få deltakere har endret sin transportatferd som følge av tiltaket, og betalingsviljen for denne type tjenester er lav. Det kan synes som om tiltaket bidrar til et noe forbedret transporttilbud for de 2 % arbeidstakerne i området som aktivt tar del i samkjøringsaktivitetene.</i>
Mindre kø	<i>Evalueringen gir ingen grunn til å anta at man vil oppnå en nedgang i køproblemer som følge av samkjøringstiltaket. I Bergenspiloten ble det gjennomført i snitt 20 samkjøringsturer per arbeidsdag, med utgangspunkt i 7000 arbeidstakere. Dersom alle samkjøringsturene erstattet en bilreise, ville det medført en beskjeden nedgang i biltrafikken i størrelsesorden 0,2 %. Evalueringen viste imidlertid at samkjøringstiltaket rekrutterer sine brukere i hovedsak fra kollektivtransport, og i realiteten påvirkes ikke biltrafikken nevneverdig. Det maksimale markedspotensialet for samkjøring ble estimert å ligge i størrelsesorden 1-2 % reduksjon i bilreiser, med utgangspunkt i datamaterialet fra Bergenspiloten.</i>
Reduserte utslipp	<i>Evalueringen gir ingen grunn til å anta at man vil oppnå reduserte utslipp av luftforurensning og klimagasser. Denne faktoren er knyttet til en reell nedgang i antall bilturer og/eller lengden på disse, hvilket ikke synes oppfylt i Bergenspiloten. Effekten av en eventuell endring i reisemiddelvalg vil også modifieres dersom det medfører alternativ bruk av kjøretøyet, det vil si at bilen som ikke lenger brukes til arbeidsreisen kan generere nye turer hos andre i husstanden.</i>

Evalueringsstudien avdekket suksesskriterier og barrierer som kan bidra til å forklare i hvilken grad samkjøringstiltaket kan forventes å realisere en ønsket atferdsendring hos trafikantene. Med utgangspunkt i den vitenskapsteoretiske atferdsmodellen som ble presentert i kapittel 7.3.2 (Fishbein & Ajzen, 2010), har analysene identifisert en rekke bakgrunnsfaktorer, holdninger, sosiale normer, grad av selvkontroll og intensjoner som gir grunnlag for å predikere endringer i reisemiddelvalg. Figur 29 viser de mest relevante faktorene, basert på resultatene som er presentert i kapittel 7.3.7. Det tas forbehold om et forholdsvis beskjedent antall faktiske brukere av samkjøringstjenesten, som ikke har gitt grunnlag for statistisk testing av de ulike forklaringsfaktorenes bidrag eller modellens teoretiske gyldighet.



Figur 29: Oppdatert atferdsmodell med resultater fra evalueringsstudien

7.6 NY KUNNSKAP OM ITS-TILTAKET

Evaluering av samkjøringsaktivitetene i Bergenspiloten har avstedkommet ny kunnskap om ITS-tiltaket samkjøring. Selv om effektstudien av reisevaner bekreftet tidligere forskning som indikerer at tiltaket har liten effekt på reisemiddelfordeling og trafikkarbeid, har evalueringen bidratt til økt forståelse av markedspotensialet, hvem som kan rekrutteres til samkjøring, viktige motivasjonsfaktorer – og hvordan bruk av spesifikke insentiver kan bidra til å overvinne iboende barrierer mot samkjøring. I avsnittene som følger presenteres kunnskapsstatus basert på evalueringsarbeidet.

Effekter

Samkjøringstiltaket har i praksis dokumentert liten eller ingen effekt som etterspørselsstyrende og trafikkreduserende tiltak. Dette er i tråd med tidligere forskning, som viser at svært få samkjøringstjenester lykkes å rekruttere tilstrekkelig antall deltakere til å nå en kritisk brukermasse. En av de største overraskelsene i evalueringsstudien, var at etter flere år med samkjøringsaktiviteter og aktiv deltakelse fra bedriftene i området, var det så mange som 67 % av de ansatte som oppga at de ikke kjente til eller hadde fått tilbud om å benytte samkjøringstjenesten. Dette til tross for massivt og vedvarende promoteringsarbeid, med lokale arrangementer, kampanjer og informasjonsarbeid i bedriftene. Først og fremst viser dette at det er krevende å fange oppmerksomheten og nå igjennom med budskapet til målgruppen. Samtidig kan det også tyde på et uforløst potensial som man ikke har fanget opp i Bergenspiloten. Videre viser analysene at det er vanskelig å beholde deltakerne over tid; kun 18 %⁵⁴ av de som har registrert seg med brukerprofil i samkjøringsløsningen står igjen som aktive brukere ved slutten av pilotprosjektet. Resultatene fra

⁵⁴ Data fra spørreundersøkelsen: Av 115 påmeldte brukere, er det kun 21 som oppgir at de har fortsatt å samkjøre på jevnlig basis

evalueringstudien viser at tiltaket har et øvre potensiale for endring i konkurranseflaten mot bil i størrelsesorden 1-2 %. Dette er også i tråd med tidligere forskning på effekter av samkjøringstiltak (Amey, 2010; Deakin et al., 2010). Analysene viser imidlertid at passasjerene som deltar i samkjøringsordningen i hovedsak rekrutteres fra kollektivtransport, og det er store barrierer som må overvinnnes for å endre reiseatferd bort fra bilkjøring. Når tiltaket ikke utløser en nedgang i antall bilreiser, gir det heller ikke grunnlag for miljøeffekter i form av reduserte klimagassutslipp eller lokal forurensning. Samkjøring gir til en viss grad økt mobilitet for en liten målgruppe som kjennetegnes av liten tilgang til bil og stor fleksibilitet i forhold til sin egen arbeidsreise.

Rammebetingelser og kontekst

Evalueringen har lyktes i å identifisere en rekke rammebetingelser og situasjonsavhengige faktorer som har betydning for rekruttering av brukere og tiltakets evne til å utløse effekter og realisere virkninger. Først og fremst må tiltaket kombineres med insentiver som gjør at brukerne opplever umiddelbare gevinster i form av redusert reisetid og/eller reduserte reisekostnader. Tilgang til sambruksfelt er et konkret eksempel på virkemiddel som motiverer til økt deltakelse. Videre er det at man kjenner noen å samkjøre med en vesentlig årsak til at folk benytter seg av samkjøringstjenestene. Dette tyder på at man ikke skal undervurdere virkemidler som bidrar til å fremme også uorganisert samkjøring. En viktig rammebetingelse for bruk av samkjøringstjenesten over tid er at teknologien oppleves som moden og brukervennlig. Tekniske problemer knyttet til bruk av samkjøringsapplikasjonen var viktige årsaker til at deltakerne sluttet med samkjøring underveis i piloten, sammen med manglende respons på forespørsler. Dette viser viktigheten av at ITS-tiltaket har god nok funksjonalitet og stabilitet før det implementeres i stor skala. Det kan være tungt å rekruttere brukere på nytt dersom de allerede har prøvd tjenesten og fått en negativ oppfatning, selv om den teknologiske utviklingen tilsier at løsningene stadig blir bedre og mer effektiv. Det er også avdekket en samvariasjon mellom prosjektorganisasjonens innsats i form av markedsføringsvirksomhet og effekter i form av rekrutterte brukere og antall samkjøringsturer. De deltakerne som fortsetter å samkjøre over tid uttrykker at de lar seg motivere av samkjøringskampanjer og premiering. Det er imidlertid en forutsetning for den enkeltes deltakelse at de har en viss kontroll over eget reisemiddelvalg. Fleksibilitet i arbeidstid og fravær av forpliktelser som medfører at man trenger bil i arbeidstiden eller for eksempel til henting av barn på arbeidsreisen er faktorer som fremmer muligheten for samkjøring – mens tilgang til egen bil har negativ effekt på rekrutteringen.

Brukeraksept

Den typiske samkjøringskandidaten synes å være kvinne, fra yngre aldersgrupper, uten barn, med lang utdannelse, og liten tilgang på egen bil. Det er en tydelig sammenheng mellom trafikantenes holdninger, problemforståelse og aksept for samkjøring. Personer med større teknologioptimisme og mindre bekymring for personvern, samt personer som vurderer miljøutfordringer som mer problematiske, har generelt større aksept for både samkjøring og andre trafikkreduserende tiltak. Påvirkning av disse faktorene (for eksempel gjennom holdningskampanjer) kan gi et positivt bidrag til rekrutteringen, men denne kunnskapen kan også brukes for å identifisere et realistisk markedspotensial før en eventuell satsing på og innføring av samkjøringstiltak. Eksempelvis kan man se for seg at den raske teknologiutviklingen i samfunnet, økt miljøbevissthet og trender knyttet til delingsøkonomi påvirker generell sosial aksept for samkjøringstiltak. Ved å observere hvordan disse holdningene endrer seg i befolkningen, kan man få indikasjoner på når tiden eventuelt er moden for å satse videre på samkjøringstiltak.

Kunnskapsgap

Evalueringstudien har bidratt til ny kunnskap om hva som er suksessfaktorer og hvilke barrierer som må overvinnes hvis man skal lykkes med samkjøringsløsninger i Norge. Relevante rammebetingelser og motivasjonsfaktorer er avdekket for ulike faser av rekrutteringsprosessen; fra man henvender seg til målgruppen, via påmelding, første gangs utprøving og til man har blitt en fast bruker av tjenesten. Dette er kunnskap som er relevant og anvendbar ved utforming og implementering av fremtidige samkjøringstiltak. Det finnes imidlertid ennå ikke erfaringer med hvordan en målrettet tilpasning av samkjøringstiltaket i kombinasjon med bestemte insentiver vil kunne påvirke markedspotensialet og effektbildet. Siden man i liten grad har lyktes med samkjøringstiltak tidligere, foreligger det heller ikke grunnlag for å vurdere hva som er kritisk brukermasse for at tjenesten skal fungere formålstjenlig. Siden man synes å være forholdsvis avhengig av ekstern sponing (som regel offentlige tilskudd) i en oppstartsfasen inntil tjenesten blir bærekraftig, er det også behov for gjennomføring av nyttekostnadsanalyser av tiltaket. Til sist har denne evalueringstudien vært avgrenset til samkjøring som et eget transporttilbud på arbeidsreiser, men det er også mulig å tenke seg samkjøring som del av et mer fleksibelt og etterspørselsbasert kollektivtilbud, der målsettingen først og fremst vil være å styrke kollektivtrafikken i konkurranseflaten mot personbiler.

7.7 NY KUNNSKAP OM EVALUERINGSMETODIKK

Case-studiet av samkjøring har gitt nyttige erfaringer med evaluering av ITS-tiltak.

Evalueringsstudien ble planlagt med en teoribasert tilnærming som kombinerte bruk av eksperimentelle og realistiske metoder i analysene. De eksperimentelle analysene var knyttet til beregning av konkrete effekter på transportarbeid og reisemiddelfordeling, mens de realistiske metodene avdekket suksesskriterier og barrierer for vellykket innføring av samkjøringstiltak. Evalueringen var både praktisk gjennomførbar og relevant med hensyn til å frembringe ny og anvendbar kunnskap. Studien bekrefter tidligere forskning som viser at effektene av samkjøringstiltak foreløpig er nokså begrenset, men at rekruttering av deltakere er svært situasjonsavhengig. Tiltakets kontekst, i form av en rekke rammebetingelser og situasjonsavhengige variabler, er inkludert i den realistiske metodikken, og er dermed blitt en viktig del av evalueringen. Resultatene av tiltaket er helt avhengig av *hvor* det innføres og *hvordan* det implementeres og gjennomføres – og evalueringsmetodikken har bidratt til å avdekke gunstige betingelser som muliggjør målrettet utforming og bruk av samkjøringstiltaket i fremtiden.

Evalueringen har gitt grunnlag for noen metodiske betraktninger med relevans for ITS-tiltak generelt. Samkjøringsapplikasjonen som ble benyttet i Bergenspiloten var relativt moden, i den forstand at det var en kommersiell løsning som var tilgjengelig på markedet og allerede i bruk i andre samkjøringstiltak, blant annet i USA. Likevel opplevde brukerne problemer med funksjonalitet og brukergrensesnitt, og applikasjonens (manglende) teknologiske modenhet var en vesentlig årsak til bortfall av deltakere underveis i pilotperioden. Allerede året etterpå, ble imidlertid den samme samkjøringsapplikasjonen brukt i et pilotprosjekt i Texas, som konkluderte med at løsningen hadde god teknisk funksjonalitet og var svært brukervennlig (Griffin et al., 2015). Typisk for slike applikasjoner er at de gjennomgår en rask og kontinuerlig teknologisk utvikling og forbedring. Resultatene fra eksperimentelle effektmålinger er dermed ikke gyldige over tid - fordi de er knyttet til et spesifikt nivå av teknologisk funksjonalitet, og overføringsverdien til andre implementeringer blir begrenset. Dette innebærer at man ved valg av evalueringsmetodikk bør ta hensyn til muligheten for å forstå tiltakets potensiale over tid, og ikke begrense seg til et øyeblikksbilde av påvisbare effekter i en gitt situasjon.

Videre har evalueringen avdekket betydningen av nettverkseffekter for ITS-tiltak. For samkjøringstiltaket handler dette om muligheten til å oppnå en kritisk brukermasse for at tjenesten skal ha tilfredsstillende funksjon. Tjenestens verdi for en potensiell bruker avhenger av antall andre brukere som også benytter seg av tjenesten. Jo flere deltakere, desto større sjans for å få treff på forespørsler om samkjøringsturer. Tilsvarende effekter kjenner vi fra andre sosiale nettverk, som for

eksempel facebook, hvor man når et punkt der tjenesten blir selvrekutterende og nærmest oppnår en monopoltilstand. I et evalueringsperspektiv er det relevant å forstå at tjenesten har liten nytte med få brukere, mens den potensielle nytten øker drastisk så snart tjenesten når en viss brukermasse.

En annen viktig erfaring er knyttet til betydningen av kontekst og årsaksvirkningsskjeder. Som allerede nevnt, viser evalueringen at rammebetingelser og bruk av insentiver har stor påvirkning på hvilke resultater som er mulig å oppnå med samkjøringstiltak. Det finnes fortsatt begrenset empiri på hvordan ulike faktorer bidrar til å fremme eller hindre effekter, og det er behov for en evalueringsmetodikk som gir kunnskap om og økt forståelse for slike årsakssammenhenger. Etablering av programteorien basert på en kombinasjon av tiltakets underliggende logikk (implementeringsteori) og vitenskapelige atferdsmodeller (endringsteori) styrker validiteten i dette arbeidet.

Til sist ble det også gjort en subjektiv betraktning i forhold til utvikling og bruk av programteori som visuelt verktøy i evalueringsprosessen: I samtaler med prosjektgruppen og andre involverte aktører var dette et nyttig og kommunikativt instrument som ga en felles referanseramme og synliggjorde viktige sammenhenger og avgrensninger i gjennomføringen av evalueringsstudien.

Kunnskapsstatusen viser at en god del studier av samkjøringstiltak drøfter potensielle virkninger og argumenterer for innføring av tiltaket, uten at det er basert på faktiske målinger eller observasjoner. Det som finnes av effektstudier har i stor grad en kvasieksperimentell oppbygning, med analyser av effekter på personbiltrafikk – enten basert på selvrapportert reiseatferd eller kun med utgangspunkt i registrerte samkjøringsturer. Det siste tilfellet er en dårlig indikator på samkjøring som trafikkreduserende eller miljøfremmende tiltak, da det ikke er en entydig sammenheng mellom samkjøring og nedgang i biltrafikk. Premissene for bruk av en kvasieksperimentell metodikk kan til en viss grad oppfylles, med bruk av kontrollgrupper og mulighet for å basere et nullalternativ (baseline) på tilgjengelige reisevaneundersøkelser. Likevel synes dette i liten grad å være inkludert i foreliggende effektstudier. Studiene er heller ikke randomiserte, da tiltakene stort sett iverksettes på steder med spesifikke trafikale problemer eller spesielt gunstige rammebetingelser – noe som er fornuftig i et samfunnsmessig perspektiv.

Felles for de eksperimentelle analysene er at de er basert på et begrenset datagrunnlag. Få aktive brukere av samkjøringstjenestene gir små utvalg i analysene og dårlig grunnlag for videre statistiske analyser (lav statistisk validitet). Et av hovedproblemene ved en tradisjonell eksperimentell fremgangsmåte er håndtering av kontekst og eksterne faktorer. Det er åpenbart at tiltakets utforming

og omgivelser, bakgrunnsfaktorer hos målgruppen og bruk av insentiver har stor betydning for den effekten som oppnås i hvert enkelt tilfelle. Mens noen studier behandler dette som en svakhet i datagrunnlaget som det er vanskelig å kontrollere (Gjerstad & Bayer, 2012; Griffin et al., 2015), prøver de fleste studier til en viss grad å håndtere disse faktorene med kvalitative analyser (Amei, 2010; Deakin et al., 2010). De metodiske utfordringene med evaluering av samkjøringstiltak er i liten grad diskutert, utover problemene knyttet til rekruttering av en kritisk brukermasse og små utvalg i effektanalyser (Gjerstad & Bayer, 2012; Griffin et al., 2015).

Den teoribaserte fremgangsmåten har vært fleksibel og nyttig i form av muligheten til å kombinere ulike metoder og analyseteknikker. Bruk av programteori som utgangspunkt for etterfølgende analyser har bidratt til identifisering og økt forståelse for relevante årsaksvirkningskjeder, eksperimentelle analyser har gitt grunnlag for vurdering av observerbare effekter, mens den realistiske metodikken har bidratt til å identifisere den kombinasjonen av innsats og kontekst som gir best effekt. Det er ikke funnet eksempler på tidligere evalueringsstudier som har benyttet et realistisk metodegrunnlag for systematisk behandling av motivasjonsfaktorer - med hensikt å finne den sammensetningen som gir det beste grunnlaget for å lykkes med innføring av samkjøringstiltak. Således har denne evalueringsstudien hatt et direkte anvendt perspektiv, med svært relevant kunnskapsbygging både for tjenesteutviklere og beslutningstakere.

7.8 OPPSUMMERING AV CASE B: SAMKJØRING

Evaluering av ITS-tiltaket samkjøring har frembragt ny kunnskap om selve tiltaket, samt metodisk innsikt i evaluering av ITS-tiltak generelt. Formålet med evalueringsstudien var å avdekke hvorvidt samkjøring har potensiale som etterspørselsstyrende og trafikkdempende tiltak i Norge, samt å identifisere relevante suksesskriterier og barrierer for innføring av tiltaket. Evalueringen tok utgangspunkt i et pågående pilotprosjekt i Bergen, og benyttet en teoribasert fremgangsmåte som inkluderte syv etterfølgende trinn; a) identifisere problemstillinger, b) etablere tiltakets programteori, c) definere evalueringskriterier og forskningsspørsmål, d) valg av evalueringsdesign, e) datainnsamling, f) analyser og g) presentasjon av resultater. Innledningsvis ble det gjennomført en litteraturstudie som grunnlag for å identifisere prosjektets problemstillinger og utvikle tiltakets programteori. Evalueringsdesignet ble utformet med en kombinasjon av kvasiekperimentelle og realistiske analysemetoder, og datainnsamlingen omfattet en spørreundersøkelse blant potensielle deltakere i pilotprosjektet, driftsstatistikk fra samkjøringsapplikasjonen og prosjektgruppen, samt intervju og samtaler med personer i bedrifter som deltok i samkjøringsordningen. Den metodiske fremgangsmåten skilte seg fra tidligere forskning, ved at den i større grad la vekt på å identifisere rammebetingelser og motivasjonsfaktorer for å lykkes med samkjøringstiltaket.

Den teoribaserte fremgangsmåten viste seg å være både anvendelig og egnet til å frembringe ny og relevant kunnskap om samkjøringstiltaket:

- Evalueringen gir liten grunn til å tro at samkjøringstiltaket vil ha noen merkbar effekt på utnyttelse av vegkapasitet eller køproblematikk. Selv om samkjøringssjåfører oppnår høyere personbelegg i sine kjøretøy enn øvrige bilførere (henholdsvis 1,35 mot 1,13), utgjør denne gruppa en svært liten del av det totale trafikkbildet. Dersom alle samkjøringsturene i Bergenspiloten hadde erstattet en bilreise, ville dette utgjort en beskjeden nedgang på 0,2 %. Dessuten viser analysene at samkjøringstiltaket i hovedsak rekrutterer sine brukere fra kollektivtransport, så i realiteten påvirkes biltrafikken ubetydelig. Det maksimale markedspotensialet for samkjøring i Bergenspiloten ble estimert å ligge i størrelsesorden 1-2 % reduksjon i bilreiser, noe som samsvarer med tidligere forskning på området. Samkjøringstiltaket forventes per i dag ikke å bidra til en reell miljøeffekt, da dette i hovedsak vil være knyttet til en nedgang i antall bilreiser.
- Evalueringen viser at samkjøringstiltaket kun i begrenset grad bidrar til økt mobilitet, da det i liten grad utløser endring i transportatferd, og den rapporterte betalingsviljen er lav. Tiltaket synes å gi et noe forbedret transporttilbud for de 2 % av arbeidstakerne i området som benytter seg av samkjøringstjenesten.
- Den største utfordringen med etablering av samkjøringstiltak er knyttet til rekrutteringsprosessen og det å oppnå en kritisk brukermasse for at tiltaket skal fremstå som funksjonelt og fleksibelt for brukerne. Viktige barrierer for deltakelse omfatter blant annet samkjøringsapplikasjonens brukervennlighet og teknologiske modenhet, liten fleksibilitet i arbeidstid eller forpliktelser på arbeidsreisen (f.eks. henting av barn). I tillegg er god tilgang på egen bil en vesentlig årsak til at man ikke ønsker å delta i samkjøringsaktivitetene. En aktiv prosjektorganisasjon med vedvarende markedsføringsaktiviteter synes å være en forutsetning for rekrutteringen av deltakere.
- De som benytter seg av samkjøringstjenester er i stor grad kvinner, yngre aldersgrupper, de som ikke har barn, de med høyere utdanning, og de med større teknologierfaring. Det er også en positiv sammenheng mellom miljøbevissthet og aksept for samkjøringstiltak. Viktigste motivasjonsfaktor for deltakelse er å kjenne noen som man kan samkjøre med, samt utsiktene til å spare tid på arbeidsreisen. Det viktigste insentivet synes å være tilgang til sambruksfelt.

Styrken i den teoribaserte evalueringstilnærmingen har i dette casestudiet vært knyttet til følgende forhold:

- Den teoribaserte fremgangsmåten legger til rette for en systematisk evaluering med mulighet til å kombinere flere metoder innenfor et teoretisk rammeverk. Dette styrker resultatenes validitet, særlig med tanke på at samkjøringstiltaket har forholdsvis begrenset deltakelse og evalueringen må baseres på små datautvalg. De ulike metodene adresserer forskjellige forskningsspørsmål; der bruk av programteori er egnet for å identifisere årsakssammenhenger og etablere struktur for det videre evalueringsarbeidet, de eksperimentelle analysene avdekker hvorvidt tiltaket har målbare effekter i en gitt prosjektperiode, mens den realistiske evalueringen kartlegger suksesskriterier og barrierer for å finne den kombinasjonen av innsats og kontekst som gir optimale virkninger. Spesielt den realistiske metodikken har bidratt til et anvendt evalueringsperspektiv som gir grunnlag for målrettet innsats og vurdering av tiltakets fremtidige potensiale.
- Den teoribaserte evalueringen inkluderer tiltakets kontekst og årsaksvirkningskjeder i analysene. For samkjøringstiltaket er betydningen av rammebetingelser og bruk av insentiver helt vesentlig for muligheten til å realisere effekter, og det synes helt nødvendig å trekke dette inn i evalueringsarbeidet. Dette er en motsetning til eksperimentelle evalueringer som behandler eksterne faktorer som et forstyrrende element som må kontrolleres for og holdes utenfor analysene.
- Den teoribaserte evalueringen er i stand til å håndtere komplekse tiltak og tiltak i stadig utvikling. Dette har vært relevant for evaluering av samkjøring, hvor samkjøringsapplikasjonen fortsatt fremsto som teknologisk umoden for brukerne av tjenesten. Evalueringen har vært i stand til å identifisere betydningen av denne faktoren.
- Den teoribaserte evalueringen har bidratt til et skifte i fokus fra selve teknologiløsningen til brukerne av tjenesten. Selv om mange ITS-prosjekter har sitt utspring i ønsket om å ta i bruk en spesifikk teknologi, har evalueringen vært i stand til å løfte brukernes behov, gjennom fokus på reisemiddelvalg og motivasjonsfaktorer.

Den teoribaserte fremgangsmåten har vist seg å være praktisk gjennomførbar og svært effektiv med hensyn til å frembringe ny og anvendbar kunnskap om ITS-tiltakets potensielle virkninger, suksesskriterier og barrierer for vellykket innføring.

8 SAMMENFATNING OG DRØFTING AV RESULTATER

I dette kapitlet drøftes erfaringer og resultater fra teoridelen som er presentert i kapittel 2-5, samt casestudiene som er dokumentert i kapittel 6 (NonStop) og 7 (Samkjøring). Litteraturstudiene har gitt grunnlag for et analytisk blikk på teorigrunnet og praksis innenfor dagens transportsystem, og ledet frem til valget av en teoribasert tilnærming for evaluering av ITS-tiltak. Casestudiene har gitt mulighet for utprøving av en teoribasert evalueringsmetodikk i praksis. To konkrete evalueringsstudier med svært forskjellige tiltak har vist at metodikken kan være anvendbar og hensiktsmessig for innovative tiltak i transportsystemet. Resultatene fra dette arbeidet har gitt innsikt i hvordan en evalueringsmetodikk for gjennomføring av fremtidige ITS-evalueringer kan utformes og anvendes.

Hensikten med avhandlingen er ikke å kritisere tradisjonelle metoder eller gjeldende evalueringspraksis innen fagfeltet. Det er isteden å synliggjøre erfaringer med det teoribaserte evalueringsperspektivet, samt å argumentere for styrker og begrensninger med denne tilnærmingen. I tillegg tilbyr avhandlingen et analytisk verktøy for utforming av en teoribasert evalueringsmetodikk, slik at andre som ønsker å anvende denne tilnærmingen kan bygge videre på erfaringer fra dette arbeidet.

8.1 ERFARINGER OG RESULTATER FRA TEORIDELE

Evalueringsbegrepet er vidt og dekker mange ulike formål, deriblant forskning som skal gi kunnskap om hvordan et tiltak virker og påvirker sine omgivelser. Den metodiske verktøykista for evalueringsarbeid er også stor og mangfoldig, og ulike fagdisipliner har ulike tradisjoner og standarder som legges til grunn for evalueringer. Innenfor transportsektoren benyttes i hovedsak et begrenset utvalg av tilgjengelige metodeverktøy. Evalueringer av veg- og trafikktiltak er tradisjonelt svært metodestyrt, med eksperimentelle sammenligningsstudier⁵⁵ som gullstandard for effektevalueringer, samt påfølgende nyttekostnadsanalyser for vurdering av samfunnsøkonomisk lønnsomhet (Böhm et al., 2016; EasyWay EEG, 2012; FOT-NET, 2014; Odeck & Welde, 2010; Statens vegvesen, 2011a). Gitt at premissene for bruk av metodegrunnet er oppfylt, gir denne fremgangsmåten vitenskapelig evidens for tiltakets observerbare og allmenngyldige effekter og konsekvenser.

Samtidig er transportsystemet i ferd med å endres radikalt som følge av mulighetene som ligger i anvendelsen av teknologiske nyvinninger. Konsekvensene er todelt: På den ene siden er det behov

⁵⁵ Kontrafaktiske analyser med kontrollerte (og helst randomiserte) utvalg

for regulering og kontroll av teknologiutviklingen, og på den andre siden er det behov for kunnskap om hvordan teknologien kan bidra til å løse de store samfunnsutfordringene (Downes, 2009; Genus & Stirling, 2018; Kelly, 2009). Eksisterende empiri er naturlig nok begrenset, mens usikkerheten og kunnskapsbehovet er stort. Evaluering av intelligente transportsystemer er derfor helt avgjørende i arbeidet med å utvikle og oppnå gode og bærekraftige transportløsninger for fremtiden.

Et analytisk blikk på den tradisjonelle evalueringsmetodikken og premissene som legges til grunn for anvendelsen, viser at det er dårlig samsvar mellom teorigrunnet og virkeligheten i det moderne transportsystemet (Bristow et al., 1997; Martens & Jenssen, 2012; Mitsakis et al., 2016; Newman-Askins et al., 2003). Dels skyldes dette at innføringen av nye teknologiske virkemidler løfter helt nye forskningsspørsmål som ikke kan besvares med klassiske effektstudier. Dels skyldes dette også at ITS har vesentlig forskjellige egenskaper og konsekvenser sammenlignet med tradisjonelle veg- og trafikktiltak. Den eksperimentelle modellen er rettet inn mot å bevise hvorvidt et tiltak gir en forhåndsdefinert og allmenngyldig effekt, uavhengig av kontekst og eksterne faktorer. Kunnskapen om hvordan innovative ITS-tiltak påvirker sine omgivelser er i liten grad etablert, og i mange tilfeller finnes det ikke tilstrekkelig grunnlag for å sette opp gode hypoteser (Mitsakis et al., 2016; Odeck & Welde, 2010). Det blir dermed mer relevant å utforske hvordan tiltaket virker, hvilke effekter man kan forvente å oppnå og hvilke rammebetingelser som fremmer eller hindrer eventuell måloppnåelse.

Det vitenskapelige eksperimentet er tilpasset situasjoner med a) etablert empiri om tiltaket, b) et modent og stabilt tiltak, c) en kontekstuavhengig og allmenngyldig effekt, og d) entydig sammenheng mellom årsak og virkning. Videre bruk av nyttekostnadsanalyser forutsetter også at det foreligger e) et begrenset antall nyttevariabler som er kjent og kan prissettes. Disse fem premissene er i liten grad oppfylt for ITS-tiltak. ITS-tiltakene kjennetegnes tvert imot av manglende empiri, høy innovasjonsgrad, teknologi i stadig utvikling, et effektbilde som er avhengig av både innsats og kontekst, komplekse forklaringsfaktorer, samt et stort antall kvalitative nyttevariabler (Mitsakis et al., 2016; Newman-Askins et al., 2003). Dette stiller store krav til forskere og andre aktører innen ITS-faget, som må utvise betydelig kreativitet og evne til kompromisser i utforming og gjennomføring av sine studier.

Forskning på ITS er i stor grad problemdrevet og anvendt, med fokus på å løse foreliggende samfunnsproblemer. I mange tilfeller kan det være vanskelig å oppfylle den eksperimentelle metodikkens grunnleggende krav om randomisering og bruk av kontrollgrupper. I det virkelige liv er det gjerne viktigere å løse et konkret og praktisk problem, enn å oppfylle kriterier for bruk av evalueringsmetodikken. Tiltaket iverksettes dermed der utfordringene er størst, og potensialet for

endring er best. Dessuten er eksperimentene ofte basert på frivillig deltakelse (selvseleksjon) av forsøkspersoner og små utvalgsstørrelser (begrenset datagrunnlag). Slike utprøvinger gir resultater med lav statistisk validitet og begrenset gyldighet i generelle kunnskapsoversikter.

Teorigjennomgangen har vist at det er behov for å utvide, eller i det minste ta i bruk en større del av, den metodiske verktøykista. Transportfaget befinner seg i grenselandet mellom ulike fagdisipliner, og forskningsspørsmålene som reises krever gjerne en bred tverrfaglig tilnærming og helhetlig forståelse av problemstillingene. Et konkret eksempel på dette finner vi i Mitsakis et al. (2016) som etterspør en helhetlig metodikk som både kan forklare effektbildet, kvantifisere nyttevirkninger, gi grunnlag for investeringsbeslutninger og dertil også optimalisere utforming og drift av eksisterende systemer. ITS medfører altså en kompleksitet som vanskelig kan håndteres av perspektivene i ett fag eller en profesjon. Fraværet av et felles akseptert teorigrunnlag og manglende innsikt i virkemidlenes forklaringsfaktorer gjør det nødvendig å kombinere ulike innfallsvinkler og metoder. Mens kvalitative verktøy kan identifisere nye fenomener, egenskaper og sammenhenger som gir grunnlag for utvikling av hypoteser og teori, kan kvantitativ forskning bidra med testing av hypoteser og tallfesting av effekter.

Transportforskning er i praksis både naturvitenskap og samfunnsvitenskap. Fagfeltet har utviklet seg fra et ingeniørpreget søkelys på kjøretøy, teknikk og mekaniske prosesser, til et mer komplekst fagområde som omfatter et helhetlig samspill mellom trafikant, kjøretøy, infrastruktur og omgivelser. Dette synliggjøres spesielt i evalueringer av ITS, hvor forskningen ivaretar alt fra selve teknologiutviklingen til anvendelsesaspektet, med samhandling mellom menneske, teknologi og transportsystemet for øvrig. Trafikkatferd og transportmønster er i realiteten en rekke sosiale valg, i tråd med Ropohls (1999) beskrivelse av teknologiske innovasjoner. Således har ITS-evalueringer kanskje mer til felles med evalueringer av sosiale intervensjoner og offentlige programmer, enn med de rigide medisinske laboratorieeksperimentene som danner grunnlaget for den tradisjonelle evalueringsmetodikken og restriktive fortolkningen av evidensbegrepet i naturvitenskapen (se drøfting av evalueringsforskning som fagfelt i f.eks. Baklien, 1993; E. G. Guba & Lincoln, 1989; Halvorsen, 2013; Scriven, 1991; Shadish, 1998; Sverdrup, 2002)

En overgang fra metodestyrte til teoribaserte evalueringer, gir større fleksibilitet i metodevalget (Blamey & Mackenzie, 2007; Hills & Junge, 2010; Pawson & Tilley, 1997). I stedet for å låse evalueringsprosessen til én spesifikk metodikk, åpnes det for å kombinere de metodene som er best egnet til å besvare de relevante forskningsspørsmålene, og således oppfylle formålet med evalueringen. Videre synes den teoribaserte evalueringsprosessen godt egnet til å håndtere egenskapene ved ITS, da den som utgangspunkt a) har læring og kunnskapsbygging som formål, b)

tar hensyn til tiltakets teknologiske modenhet, c) finner kombinasjonen av innsats og virkemidler som gir ønsket effekt, d) bidrar til kunnskap om hvorfor et tiltak virker eller ikke, og e) identifiserer og vurderer alle konsekvenser, også de som er kvalitative eller utilsiktet (Blamey & Mackenzie, 2007; Bredgard et al., 2016; Chen, 2012; Funnel & Rogers, 2011; Government of Canada, 2012; Hills & Junge, 2010; Pawson & Tilley, 1997; Weiss, 1995). Anvendelse av en teoribasert fremgangsmåte i ITS-evalueringer åpner for en bredere fortolkning av evidensbegrepet, i tråd med Sackett et al. (1996). Dette innebærer at kunnskap fra metodeorienterte effektstudier kombineres med ekspertenes erfaringsbaserte kunnskap og trafikantenes ønsker og behov. Dette resulterer i et kunnskapsgrunnlag som i større grad har relevans også for spørsmål rundt implementeringen av det aktuelle ITS-tiltaket som vurderes. Metodiske begrensninger innenfor det teoribaserte perspektivet er i hovedsak knyttet til spørsmål rundt evalueringer med et kontrollformål (dokumentasjon av effektiviteten til et tiltak), fordi metodikken anses å ha svakere vitenskapelig evidens i kvantitative analyser og fordi det tette samspillet mellom evaluator og involverte aktører kan svekke evalueringens objektivitet. Den teoribaserte tilnærmingen synes derfor mest relevant når formålet med ITS-evalueringen er helhetlig kunnskapsproduksjon i et anvendt perspektiv. Etablering av programteorien sikrer involvering og kunnskapsutveksling mellom aktører i prosjektet, og den visuelle modellen utgjør et felles rammeverk for avklaring av forventninger og bevisstgjøring av målbildet. Videre har teoribaserte evalueringer en formativ karakter, som gjør at verdifull kunnskap kan etableres allerede før tiltaket implementeres i stor skala. Dette er nyttig i pilotstudier og evaluering av umodne ITS-tjenester.

Den teoretiske gjennomgangen har altså gitt grunnlag for å foreslå at evaluering av ITS-tiltak tar utgangspunkt i en teoribasert evalueringstilnærming. Denne fremgangsmåten tilrettelegger en teoretisk forankring og struktur for evalueringsarbeidet, som gir rom for en mer innholdsrik metodisk verktøykiste og instrumenter som håndterer et større spekter av relevante forskningsspørsmål. I neste delkapittel drøftes de praktiske erfaringene med utprøving av en teoribasert evalueringsprosess i to konkrete casestudier.

8.2 ERFARINGER OG RESULTATER FRA CASE-STUDIENE

En teoribasert evalueringsprosess ble utviklet og benyttet i to konkrete evalueringsprosjekter med vidt forskjellige ITS-tiltak. Casestudiene har både bekreftet en del tidligere dokumenterte utfordringer med ITS-evalueringer, samt gitt erfaring med den teoribaserte evalueringsprosessen som en hensiktsmessig tilnærming. En kort beskrivelse av de to aktuelle ITS-tiltakene er oppsummert i Tabell 30.

Tabell 30: ITS-tiltakenes ulike innretning i de to casestudiene

ITS-tiltak	NonStop	Samkjøring
ITS-applikasjon	NonStop beslutningsstøtte på kontrollstasjon	Samkjøringsapplikasjon på mobil (Carma-appen)
ITS-tjeneste ⁵⁶	Non-stop klarering	Styring av delte reiser
ITS tjenestegruppe	Forhåndsklarering av kommersielle kjøretøy	Etterspørselsbasert kollektivtransport
ITS tjenesteområde	Godstransport	Kollektivtransport
IKT infrastruktur	Nettbrett, sensorikk langs veg, samt dataregistre og beslutningsstøttesystem	Mobiltelefon og koblingsalgoritmer i et sentralsystem
ITS-bruker	Myndighetene (trafikk-kontrollør)	Privatpersoner (arbeidsreisende)
Innføringsløp	Obligatorisk	Frivillig
Teknologisk modenhet	Tidlig fase, teknologisk umoden	Kommersiell løsning, antatt teknologisk moden
Markedspenetrasjon	Effekt fra første bruker	Forutsetter kritisk brukermasse
Antatt nyttevariabel	Redusert reisetid for næringstransport, økt deteksjonsrate kontrollstasjon	Endring i reiseatferd, reduksjon i antall bilreiser
Situasjonsavhengige faktorer	Fysisk utforming kontrollstasjon, bruksområder for tjenesten og trafikkvolum	Rammebetingelser, insentiver og bakgrunnsfaktorer i målgruppen

Tabellen underbygger tidligere utsagn om at det er stor variasjon i ulike ITS-tjenester, både med hensyn til teknologi, målgrupper og virkningene det tas sikte på å oppnå (Foss, 2017b). ITS er en felles betegnelse på svært heterogene virkemidler, noe som gjør det vanskelig å etablere en felles evalueringsmetodikk for ITS-tiltak. I tillegg påvirkes effektbildet av rammebetingelser og situasjonsavhengige faktorer rundt iverksettelsen av tiltaket. Det er en direkte sammenheng mellom ITS-tjenestenes innføringsløp i markedet og virkningene som oppnås i et samfunnsperspektiv, som blant annet beskrevet i Zlocki et al. (2016) i form av ulike scenarier for markedsintroduksjon. Det har eksempelvis stor betydning hvorvidt ITS-tjenesten er obligatorisk, som i tilfellet med NonStop-tiltaket (all næringstrafikk som passerer kontrollpunktet blir påvirket), kontra Samkjøringstiltaket som er frivillig, og dermed i mye større grad avhengig av brukeraksept, betalingsvilje og eksterne insentiver. Til tross for at de to ITS-tiltakene har svært ulik innretning og egenskaper, har den teoribaserte evalueringsprosessen vært anvendbar og bidratt med nyttig kunnskap i begge tilfellene.

De to casestudiene har gitt en rekke sammenfallende metodiske erfaringer. Formålet med begge evalueringene var å fremme kunnskapsbygging og økt forståelse for hvordan tiltaket kan fungere i en fremtidig anvendelse. Felles for tiltakene, var at den innledende litteraturgjennomgangen avdekket utfordringer med kvantitativ innsamling og analyse av effektdata, og det forelå lite publisert forskning på faktiske virkninger. Evalueringspraksis synes likevel i stor grad å basere seg på klassiske effektstudier uten særlig forklaring av antakelsene som ligger til grunn for analysene. De fleste studiene manglet kontrollgrupper, og årsaksvariabelen var i realiteten en miks av ulike tiltak (ITS-applikasjon, holdningskampanjer, reguleringer og opplæring m.m.). Kvalitative analyser

⁵⁶ Kategorisering av ITS-tiltaket i tjeneste, tjenestegruppe og tjenesteområde i henhold til Statens vegvesens ITS-veileder (Foss, 2017b)

var i liten grad satt i system, og gjerne gjennomført sekundært når konklusjoner ble fattet på et sparsomt datagrunnlag.

Begge tiltakene har vært forholdsvis komplekse, enten i form av en omfattende IKT infrastruktur (som i NonStop-tiltaket) eller med en kombinasjon av flere tjenester og virkemidler (i begge tiltakene). På dette viset sammenfaller både evalueringsformålet og kompleksiteten med typiske ITS-evalueringer (Böhm et al., 2016; Mitsakis et al., 2016). I begge casene har den teoribaserte evalueringsmetodikken vist seg å være praktisk gjennomførbar, til tross for både randomiseringsproblem og små datautvalg. Metodikken har også vist stor fleksibilitet med hensyn til tidspunktet for evaluering, i tråd med tidligere diskusjoner i Chen (2005), Samset og Christensen (2013), Urban et al. (2014) og Bredgard et al. (2016). NonStop-tiltaket var fortsatt i en tidlig fase av innovasjonsprosessen, og dermed preget av lav teknologisk modenhet og ustabil pilotering. Samkjøringstiltaket befant seg i motsatt ende av skalaen, med en ferdig kommersiell løsning som allerede var på markedet. Likevel ble det erfart at begge applikasjonene var gjenstand for en kontinuerlig teknologiutvikling, som endret egenskaper og stabilitet underveis i evalueringsprosessen. Det er altså tydelig at ingen av tiltakene oppfylte den eksperimentelle effektstudiens antakelse om teknologisk modenhet og stabil drift over tid; forutsetninger som også legges til grunn for FESTA-metodikken (Barnard et al., 2016).

Disse erfaringene gir støtte til Urban et al. (2014) i argumentasjonen for et evolusjonsperspektiv som tar høyde for at både tiltaket og omgivelsene er i stadig utvikling. Selv om forfatterne baserer sine vurderinger på evaluering av sosiale programtiltak, synes tankesettet å være relevant også for innføring av teknologiske løsninger i transportsystemet: I stedet for å forvente at ITS-tiltaket er ferdig utviklet og fungerer perfekt fra tidlig implementering, er det mer realistisk å ta utgangspunkt i en livssyklus med prøving, feiling og stadig forbedring av tiltaket. Dette skjer i samspill med ITS-brukerne og øvrige trafikanter som også gjennomgår en løpende læringsprosess og atferdstilpasning som respons på tiltaket (Martens & Jenssen, 2012). Dette perspektivet skiller seg fra FESTA-metodikken, som bygger på den eksperimentelle modellens krav om tiltaket som en konstant og uforanderlig enhet. Når evolusjonsperspektivet legges til grunn for evalueringen, blir konsekvensen at evalueringsmetodikken må kunne inkludere flere former for validitet og kausalitetsteori. Dette harmoniserer godt med den teoribaserte evalueringstilnærmingen som anerkjenner betydningen av metodisk pluralisme og inkrementell kunnskapsproduksjon over tid (Hills & Junge, 2010; Pawson & Tilley, 1997). Dette forutsetter imidlertid god kjennskap til de ulike metodenes styrker og svakheter, samt kompetanse til å vurdere resultatenes gyldighet og overførbarhet til andre situasjoner og implementeringer.

I begge case-studiene ble teknologisk funksjonalitet, kvalitet og modenhet behandlet som en rammebetingelse i forbindelse med etablering av programteorien. Temaet var imidlertid svært sentralt både med hensyn til valg av evalueringsdesign og realisering av tiltakets effekter. Det ble erfart at teknologien ga føringer for hvilke forskningsspørsmål som er relevant å stille, samt datainnsamlingsmetodene som ble benyttet i gjennomføringsfasen. Tiltakets teknologiske modenhet kunne således vært løftet frem som et eget punkt tidlig i evalueringsprosessen.

Det er verdt å merke seg at evalueringer i en tidlig fase av tiltakets innførlingsløp øker arbeidets formative kapasitet, gjennom å tidlig avdekke kritiske barrierer og nødvendige justeringer (Bredgard et al., 2016; Chen, 2005; Karlsen & Jentoft, 2013). I så måte kan teoribaserte evalueringer bidra til mer effektiv ressursbruk både for teknologitilbyder og transportmyndighet. Eksempelvis ble det i NonStop-evalueringen tidlig avdekket at den fysiske utformingen av kontrollområdet hadde stor betydning for hvorvidt nytteeffekten i form av redusert ventetid for næringstransporten kunne realiseres. Det ble dermed iverksatt fysiske tilpasninger av omgivelsene parallelt med teknologiutviklingen. I tillegg ble det i en tidlig fase av evalueringen avdekket at brukergrensesnittet ikke hadde oppnådd tilstrekkelig funksjonalitet. Dette la føringer for både gjennomføringen av piloten og evalueringsdesignet, slik at man unngikk feilslått bruk av evalueringsressursene. Fleksibiliteten i den teoribaserte evalueringsprosessen gjorde det dermed mulig å skille mellom det som Dahler-Larsen (2013) benevner som *teorifeil* (tiltaket virker ikke) versus *implementasjonsfeil* (ytre faktorer hindrer at tiltaket virker) i tiltakets effektvurdering. Erfaringene fra casestudiene gir altså støtte til litteraturen som beskriver den teoribaserte tilnærmingen som et godt virkemiddel for proaktiv kunnskapsproduksjon (Chen, 2005; Karlsen & Jentoft, 2013) og hensiktsmessig både i planleggingsfasen av tiltaket, som styringsredskap og som evalueringsmodell (Bredgard et al., 2016). Tilnærmingen kan dermed bidra til å motvirke Collingridges dilemma om sosial kontroll av teknologisk utvikling (Genus & Stirling, 2018), ved å frembringe kunnskap i en tidlig fase hvor det fortsatt er mulig å foreta kursendring eller justering.

Etablering av tiltakets programteori ble opplevd som en hensiktsmessig øvelse i begge casestudiene. Allerede tidlig i evalueringsprosessen, ble det mulig å avdekke logiske brister og mismatch mellom ulike aktørers forventninger (og behov) og hva som er realistisk å oppnå. I NonStop-caset hadde interessegruppene ulikt syn på hvilke lovbrudd som skulle inkluderes i beslutningsstøttesystemet, mens i Samkjørings-caset var det til dels urealistiske forventninger til miljøeffekten av tiltaket. Programteorien ble i begge tilfellene et nyttig verktøy for forventningsavklaring, og bidro til å skape en felles forståelse av det teoretiske grunnlaget for innføringen av ITS-tiltaket. Erfaringene samsvarer med tidligere teoretiske betraktninger som peker på den teoribaserte

evalueringsmodellens styrke som kommunikativt verktøy (Kaplan & Garrett, 2005; Marchal et al., 2013; van der Knaap, 2004). Fremgangsmåten kan på dette viset bidra til å bygge bro over det som i Böhm et al. (2016) beskrives som et kommunikasjonsgap mellom ITS-tilbyder og beslutningstaker. Det finnes ikke en generell oppskrift på hvilke elementer som skal inngå i en programteori for ITS-tiltak, og programteorien må utvikles og tilpasses det konkrete tiltaket i hvert tilfelle. Dette krever en viss innsats fra prosjektgruppa i en tidlig fase av evalueringsprosessen. En mulig utfordring som trekkes frem i metodelitteraturen, er at det kan være krevende å etablere programteorien med mange involverte aktører og motstridende interesser (Berger & Vrangbæk, 2011). Dette var ikke en fremtredende erfaring i de to aktuelle case-studiene. Tvert imot ble det opplevd som en styrke at evalueringsprosessen var grundig forankret i et felles teorigrunnlag.

Konteksten for iverksettelsen av tiltaket har hatt stor betydning i begge case-studiene. Det er allerede nevnt at den fysiske utformingen av kontrollstasjonen hadde stor innvirkning på nytteverdien av NonStop-tiltaket. Tilsvarende ble samkjøringstiltaket i sin tid iverksatt i omgivelser som var forventet å være spesielt gunstig med tanke på å rekruttere brukere, både med hensyn til problemopplevelse (store trafikkutfordringer) og insentiver (blant annet tilgang til sambruksfelt). Dette betyr at den eksperimentelle metodikkens forutsetninger om en allmenngyldig effekt og entydig sammenheng mellom årsak og virkning ikke var oppfylt i noen av casestudiene. Tidligere evalueringer av slike tiltak, har i stor grad behandlet kontekst og eksterne faktorer som en svakhet i datamaterialet som det er vanskelig å kontrollere (Gjerstad & Bayer, 2012; Griffin et al., 2015). I den teoribaserte tilnærmingen ble disse faktorene tatt inn som en integrert del av evalueringsstudien.

Den sterkeste kritikken mot teoribaserte metoder er knyttet til manglende evidensorientering; altså liten evne til å produsere kvantitative bevis med samme vitenskapelige styrke som eksperimentelle modeller (Blamey & Mackenzie, 2007; Hills & Junge, 2010; Weiss, 1998). Det må understrekes at den teoribaserte *evalueringsprosessen* ikke er avgrenset til bruk av teoribaserte *metoder*, og fremgangsmåten åpner for å kombinere flere hensiktsmessige metoder og innfallsvinkler. I evalueringen av samkjøringstiltaket ble det for eksempel inkludert en eksperimentell studie av reisemiddelvalg før og etter innføring av tiltaket. Erfaringen fra case-studiene var imidlertid at ITS-tiltakene i de aller fleste henseender ikke oppfylte premisene for bruk av eksperimentell metodikk, og det ville dermed ikke vært mulig å gjennomføre adekvate analyser som dokumentasjon på tiltakenes allmenngyldige effektivitet. Den teoribaserte evalueringsprosessen har likevel lyktes å frembringe nyttig kunnskap både om tiltakenes kapasitet som trafikkstyrende virkemiddel og om viktige rammebetingelser for å lykkes med tiltakene. Således har de teoribaserte metodene gitt økt innsikt og merverdi, selv i situasjoner der det ikke er mulig å bevise en umiddelbar effekt.

8.3 VURDERINGER AV VITENSKAPELIGHET

Utprøving og evaluering av ITS-tiltak har blitt stadig mer aktuelt, som et resultat av samfunnets behov for å løse utfordringer knyttet til transportsektoren. Forskningsfeltet fremstår som tverrfaglig, med behov for stor grad av samarbeid mellom ulike fagprofesjoner for å fremskaffe kunnskap om komplekse problemstillinger i et anvendt perspektiv. Transportforskningen har imidlertid ikke lyktes å oppnå konsensus om et enhetlig metodesett eller felles teoretisk grunnlag for beskrivelse av ITS-tiltak, og ulike metodiske tilnærminger må kombineres for å løse sentrale forskningsspørsmål. I et vitenskapsteoretisk perspektiv, kjennetegnes transportfaget således av flere karakteristikker som ifølge Nowotny et al. (2003) medfører endringer i den vitenskapelige kunnskapsproduksjonen. Fagfeltet føyer seg dermed inn i rekken av nye tverrfaglige, problemløsningsbaserte vitenskaper som erstatter det tradisjonelle, disiplinbaserte, akademiske regimet (se drøfting av modus 1- og modus 2-teorier i kapittel 3.5.1).

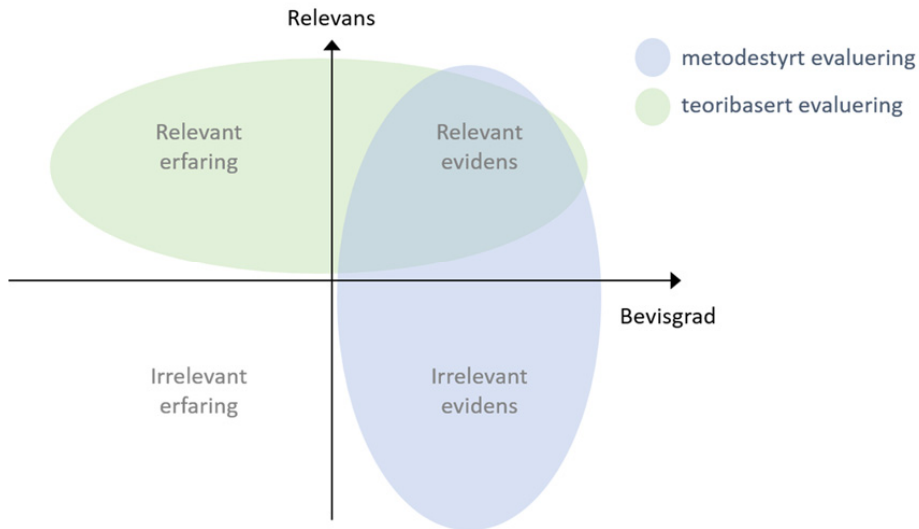
Den teoribaserte tilnærmingen har vist seg å være både anvendbar og hensiktsmessig for ITS-evalueringer, med kapasitet til å fremskaffe helhetlig forståelse og relevant kunnskap om virkemidlene i det moderne transportsystemet. Det må imidlertid understrekes at heller ikke den teoribaserte evalueringprosessen er et trylleformular som kan gi svar på alle spørsmål omkring ITS-tiltakenes virkninger og konsekvenser. Fremgangsmåten adresserer riktignok flere (og kanskje mer relevante) forskningsspørsmål enn den tradisjonelle metodestyrt tilnærmingen, og bidrar således til en bedre forståelse av tiltakenes potensial i et fremtidig transportsystem. Dette i motsetning til den eksperimentelle modellen som gir evidens til et øyeblikksbilde for en gitt effekt. Den teoribaserte evalueringen er også pragmatisk med hensyn til å identifisere de forskningsspørsmålene som lar seg besvare i de ulike fasene av innovasjonsprosessen eller tiltakets innføringsløp, og tilrettelegger således for en mer effektiv og riktig bruk av evalueringsressurser underveis. I mange tilfeller betyr dette at vi må erkjenne at fagområdet fortsatt er ungt og umodent, og vi faktisk må *finne* og *forstå* de relevante effektene, før vi kan sette inn store ressurser på å *bevise* dem. Innenfor transportsektoren finnes det knapt eksempler på bruk av teoribaserte metoder i evalueringsstudier, og en utvidelse av det metodiske grunnlaget som anvendes, kan absolutt bidra til å skape et mer helhetlig kunnskapsgrunnlag og en bedre utnyttelse av evalueringsressursene.

Et grunnleggende og relevant spørsmål, er om overgangen fra metodestyrt til teoribaserte evalueringprosesser vil ivareta vitenskapeligheten i evalueringsarbeidet? Erfaringene fra dette doktorgradsarbeidet gir støtte til at så er tilfelle. Først og fremst er det viktig å huske at den teoribaserte tilnærmingen ikke forkaster bruk av eksperimentell metodikk, men åpner for andre fremgangsmåter i tillegg. Fleksibiliteten i metodevalg gjør at det fortsatt er mulig å basere

evalueringen helt eller delvis på eksperimentelle analyseteknikker, dersom betingelsene ligger til rette for dette. I slike tilfeller kan den teoribaserte tilnærmingen bidra med et rammeverk og teoretisk forankring som supplement til dagens evalueringsmetodikk, og dermed bidra til å motvirke det Osland et al. (2007) omtaler som et transparensproblem i dagens evalueringspraksis i transportsektoren. I tilfellene hvor grunnlaget for eksperimentelle analyser ikke er til stede, vil den teoribaserte tilnærmingen øke validiteten i et ikke-eksperimentelt design, gjennom å knytte forskningen opp mot teori (programteorien). Den samme fleksibiliteten gjør det imidlertid vanskeligere å sammenligne resultatene fra ulike ITS-evalueringer (ekstern validitet). Inkludering av kontekst og variasjon i metodebruk, gjør at det er lite hensiktsmessig å gjennomføre meta-analyser eller sette opp generelle allmenngyldige effektkataloger for ITS-tiltak. Med utgangspunkt i evolusjonsteorien i Urban et al. (2014) kan det riktignok argumenteres for at dette uansett er en konsekvens av ITS-tiltakenes kontinuerlige teknologiutvikling og iboende situasjonsavhengige egenskaper – uavhengig av metodebruk.

Case-studiene har vist at det er en sterk sammenheng mellom ITS-tiltakenes effekter og aspekter ved implementeringen av tiltaket. Eksperimentelle effektstudier kan således kun bringe frem deler av kunnskapsgrunnlaget vi trenger for å vurdere hvordan ITS-tiltak vil fungere i den virkelige verden. I realiteten foregår ITS i et komplekst samspill mellom ulike faktorer både i og utenfor tiltaket. Det eksperimentelle metodeoppsettet er et godt egnet forskningsdesign for å avdekke hvorvidt et eksisterende tiltak gir en ønsket effekt, i situasjoner hvor det finnes en entydig årsakssammenheng og hvor forsøkspersonene som inngår i eksperimentet er mange nok til å gi statistisk pålitelige svar. Men hva gjør man om disse metodiske premissene ikke lar seg oppfylle? Eller om tiltaket som skal vurderes ennå ikke er ferdig utviklet? Ifølge Sackett et al. (1996) innebærer en kunnskapsbasert praksis og politikktutforming å benytte den beste tilgjengelige kunnskapen om hvordan tiltaket vil fungere i sosialt samspill med brukerne og omgivelsene. Innenfor transportsektoren er dette blitt tolket ensbetydende med metodisk sterkeste evidens på hvorvidt tiltaket virker i kontrollerte omgivelser. Det er imidlertid grunn til å knytte evidensbegrepet sterkere opp mot problemstillingen som studeres. Erfaringene fra dette arbeidet tyder på at det er lite trolig at et ITS-tiltak som er skreddersydd til en bestemt implementasjon også skal passe like godt et annet sted. Det er derfor behov for en mer nyansert tilnærming som også åpner opp for erfaringsbasert kunnskap og en mer fleksibel kausalitetsforståelse. I beslutningsgrunnlaget for innføring av ITS-tiltak er det lite realistisk å jakte etter et endelig og sikkert effektbevis som er gyldig over tid. Kunnskapsproduksjonen må isteden betraktes som en dynamisk prosess, hvor man til enhver tid velger ut den mest pålitelige og relevante kunnskapen for det forskningsspørsmålet som skal besvares. I dette bildet må kunnskapens kvalitet vurderes både

med hensyn til resultatenes gyldighet og i hvilken grad de er relevante for den settingen som vurderes, som illustrert i Figur 30.



Figur 30: Kunnskapsproduksjon som produkt av resultatenes gyldighet og relevans

Som figuren illustrerer vil en overgang fra den metodestyrt eksperimentelle tilnærmingen til en teoribasert evalueringsprosess kunne åpne for en kunnskapsproduksjon som i større grad inkluderer relevant erfaring basert på ekspertvurderinger og brukerundersøkelser. Denne tilnærmingen til kunnskapsproduksjon utfordrer det tradisjonelle evidenshierarkiet, ved å knytte evidensbegrepet sterkere opp mot forskningsspørsmålet som skal besvares. Faren med et overdrevet metodefokus er at man i jakten på evidens går glipp av forskning som havner i kategorien relevant erfaring – eller ender opp med å bruke evalueringsressursene på sikker, men irrelevant kunnskap. Dette er for eksempel tilfelle dersom evalueringen benyttes til å bevise effektene av en teknologisk umoden ITS-tjeneste. Det er også en risiko for at man på søken etter evidens ender opp uten resultater i det hele tatt, eksempelvis dersom datagrunnlaget viser seg å ikke oppfylle premisene for forsøksdesignet.

Den teoribaserte fremgangsmåten innehar flere frihetsgrader enn den eksperimentelle evalueringemetodikken, og det stilles følgelig større krav både til den som gjennomfører evalueringen og den som skal bruke resultatene. Der den tradisjonelle metodikken gir en spesifikk oppskrift som kan følges rutinemessig av den som evaluerer, vil den teoribaserte evalueringsprosessen forutsette både en viss grad av metodekompetanse og kunnskap om tiltaket som evalueres (Baklien, 1993; E. G. Guba & Lincoln, 1989; Halvorsen, 2013; Sverdrup, 2002).

Dette bør riktignok være et betimelig krav i alle tilfeller. Som Pawson and Tilley (1997) påpeker, er det en fare for at den eksperimentelle oppskriften reduserer evalueringen til en teknisk øvelse som gjennomføres ukritisk. Det er kanskje grunn til å stille spørsmål om manglende teknologiforståelse eller svak metodekunnskap har fått lov til å prege en del ITS-evalueringer - hvor eksperimentelle analyser er oppskriftsmessig gjennomført, til tross for at de metodiske premissene ikke er oppfylt (Dziekan et al., 2013; Lervåg, 2012). Mitsakis et al. (2016) og Belin et al. (2010) berører den samme problemstillingen, når de understreker at evaluering av ITS skiller seg fra øvrige vegtiltak, fordi de krever omfattende kunnskap om både ITS-systemet i seg selv (teknologi) og hvordan det fungerer i samspillet med trafikanter og omgivelsene. Dette samsvarer med budskapet fra en rekke evalueringsforskere som understreker at evalueringer betinger mer omfattende kompetanse en ren forskning, fordi resultatene skal vurderes i en større samfunnsfaglig kontekst (Baklien, 1993; Balla et al., 2015; Scriven, 1991).

Innvendinger mot den teoribaserte fremgangsmåten er gjerne knyttet til mangel på objektivitet i forskningen (Bredgard et al., 2016; Fitzpatrick et al., 2012; O'Sullivan, 2012). For det første trekkes evalueringsforskeren tettere inn i tiltakets utviklingsløp, ved at resultatene fra evalueringen brukes formativt både i innovasjonsprosessen og implementeringen av tiltaket. Dette gjør at det kan stilles spørsmål ved om den som gjennomfører evalueringen har nødvendig avstand både til prosjekteier og produktet som evalueres. Dertil kommer også frihetsgradene i gjennomføring av evalueringen. Dette har dårlig gjenklang i den naturvitenskapelige tradisjonen, hvor objektivitet i stor grad defineres med henvisning til bruk av den eksperimentelle metodikken. Disse argumentene rommer imidlertid en implisitt forventning om at evalueringens formål er bedømmelse av tiltakets umiddelbare suksessopnåelse. Hvis man derimot legger til grunn at hensikten med ITS-evalueringer oftere er et tverrfaglig samarbeid om kunnskapsbygging, slik det påpekes i f.eks. Böhm et al. (2016), endres også synet på objektivitet. I dette perspektivet ivaretas objektiviteten i større grad gjennom åpenhet og transparens, ved at evalueringens utgangspunkt, metodiske valg og avgrensninger synliggjøres og dokumenteres. Dette samsvarer godt med den teoribaserte evalueringsprosessen, hvor premissene for forskningen fremkommer eksplisitt av tiltakets programteori.

Den eksperimentelle fremgangsmåten er forankret i naturvitenskapens krav om metodisk stringens, mens den teoribaserte evalueringsprosessen baserer sin gyldighet på teoretisk validitet. Dette innebærer en vesensforskjell med hensyn til vitenskapelig ståsted og forståelsen av kausalitetsbegrepet (se f.eks. Bhaskar, 2008; Pawson & Tilley, 1997; Tufte, 2013). Den eksperimentelle modellen tar for gitt at det er en automatikk i sammenhengen mellom tiltak og

effekt, hvor kausaliteten bevises gjennom kontrafaktiske analyser. Den teoribaserte fremgangsmåten anser derimot kausalitet som en latent egenskap som aktiveres under gitte omstendigheter, det vil si at tiltaket øker sannsynligheten for at en effekt inntreffer, men at det også er avhengig av rammebetingelser og kontekst. Kausaliteten bestemmes dermed gjennom studier av årsakvirkningskjeder. Selv om den naturvitenskapelige tradisjonen står sterkt innen ingeniørfaget, er det i utgangspunktet ikke grunnlag for å rangere det ene vitenskapssynet som bedre enn det andre. Metodikkens egnethet synes i større grad å være knyttet til evalueringsformålet, hvor den eksperimentelle fremgangsmåten har sin styrke i et bedømmelsesperspektiv, og den teoribaserte fremgangsmåten har sin styrke i et kunnskaps- og læringsperspektiv. Bildet kompliseres imidlertid noe av at den teoribaserte evalueringsprosessen også har kapasitet til å romme flere metodiske innfallsvinkler som delprosesser i evalueringsarbeidet. Som overordnet rammeverk, vil dermed den teoribaserte fremgangsmåten kunne ivareta både en eksplorativ utforskning av et forholdsvis ukjent fagfelt, og samtidig bidra til kvantifisering og verifikasjon av eventuelle effekter.

I henhold til det klassiske naturvitenskapelige tankesettet, er vitenskapens rolle å vise hvordan verden ser ut og forklare hvordan ting henger sammen. Den gir derimot ikke svar på hva som bør skje i fremtiden. Evalueringens oppgave er i større grad å bidra til kunnskapsbasert praksis og politikk i et fremtidsperspektiv. Store teknologiske innovasjoner medfører som regel nye bruksområder og atferdsendringer, som vi ikke umiddelbart ser eller kan tenke oss til. I begynnelsen av forrige århundre var det nok vanskelig å forutse hvordan privatbilens inntog skulle påvirke ikke bare vårt reisemønster, men også hvordan vi bor og lever; infrastrukturen, bosetningen, arbeidsmarkedet, byutviklingen, økonomiske levekår og klimaendringer blant andre forhold. For ti år siden kunne man knapt ane betydningen av smarttelefonen for vårt hverdagsliv og sosiale relasjoner. Innenfor transportfaget er det i dag et stort behov for vitenskapelige metoder som kan fremskaffe et kunnskapsgrunnlag for beslutninger som vil følge oss inn mot fremtidens transportsystem. En overgang fra metodestyrte til teoribaserte evalueringer vil øke evalueringens relevans på bekostning av metodisk rigiditet. Skal vi lykkes med kunnskapsdannelsen og oppnå gode beslutningsgrunnlag i det intelligente transportsystemet, må vi være i stand til å kombinere gode vitenskapsidealer med evnen til pragmatisme og kompromisser.

9 FORSLAG TIL RAMMEVERK FOR EVALUERING AV ITS-TILTAK

Teorigjennomgang og praktiske casestudier har vist at ITS et komplekst og mangfoldig virkemiddel, og det finnes ikke ett bestemt evalueringsdesign eller en felles metode som fungerer for alle ITS-tiltak. Evalueringsmetodikken må dermed skreddersys for hvert enkelt case, slik at den imøtekommer forskningsspørsmålet som skal besvares og tiltakets egenskaper ved hver enkelt implementering.

Dette kapittelet presenterer en teoribasert evalueringsprosess som hjelper forskeren å bygge opp en hensiktsmessig og tilpasset evalueringsmetodikk. Rammeverket gir en stegvis veiledning som støtter forskeren i prosessen med å definere relevante forskningsspørsmål, etablere en programteori og metodevalg, samt gjennomføre datainnsamling og analyser. Den teoribaserte evalueringsprosessen kan benyttes for evalueringer gjennom hele tiltakets livsløp (både ex ante og ex post-studier).

9.1 EN TEORIBASERT EVALUERINGSPROSESS

Erfaringene fra doktorgradsarbeidet har gitt grunnlag for å anbefale en teoribasert tilnærming for evaluering av ITS-tiltak. Det teoribaserte evalueringsperspektivet innebærer at effektspørsmålet først og fremst handler om hvordan tiltaket kan forstås i en kompleks sammenheng, ved å studere hvordan kombinasjonen av innsats og kontekst (årsak) kan forklare eventuelle forandringer (virkning). Som del av evalueringsprosessen utvikler man derfor en programteori som beskriver hvordan tiltaket logisk forventes å utløse ønskede effekter og virkninger i en gitt kontekst. Evalueringen søker altså å besvare både *hvordan tiltaket virker* og *hvorfor tiltaket virker*. Relevante eksterne faktorer som har betydning for hvorvidt man lykkes med tiltaket inkluderes i evalueringen. Evalueringsdesign og analysemetoder velges med utgangspunkt i årsaksvirkningskjedene som danner grunnlaget for programteorien. I praksis betyr dette at man ikke kan følge en detaljert metodisk oppskrift som er gyldig for alle typer evalueringer eller tiltak, men rammeverket tilbyr derimot en veiledet prosess mot valg av et hensiktsmessig evalueringsdesign med anvendbare metoder som er tilpasset evalueringens formål og de forskningsspørsmålene som søkes besvart.

Denne evalueringsprosessen imøtekommer de fem anbefalingene for ITS-evalueringer som ble definert i kapittel 4:

- 1) *Evalueringen bør ha et læringsperspektiv som bidrar til helhetlig kunnskapsbygging*
- 2) *Evalueringen bør kunne håndtere tiltak med ulik teknologisk modenhet*
- 3) *Evalueringen bør kunne identifisere de kontekstuelle faktorenes betydning*
- 4) *Evalueringen bør gi kunnskap om hvorfor et tiltak virker eller ikke*
- 5) *Evalueringen bør kunne identifisere og vurdere alle relevante konsekvenser, også kvalitative*

Rammeverket som foreslås i dette avsnittet gir en trinnvis veiledning for evaluering av ITS-tiltak, og fremskaffer en skreddersydd og formålstjenlig evalueringsmetodikk gjennom kunnskapsbaserte valg i hvert steg i evalueringsprosessen. Dette bidrar til effektiv og målrettet bruk av evalueringsressurser, samt en synliggjøring av hvilke sammenhenger som faktisk studeres – og hvilke kunnskapshull som eventuelt fortsatt eksisterer.

Et sammendrag av evalueringsprosessen er gjengitt i Figur 31 og nærmere beskrevet i avsnittene som følger.



Figur 31: Forslag til teoribasert evalueringsprosess for ITS-tiltak

Den foreslåtte evalueringsprosessen er egnet til å frembringe relevant og anvendbar kunnskap om en ITS-tjeneste gjennom hele tiltakets livssyklus, også i en tidlig fase mens tiltaket fortsatt er på planstadiet, eller etter kun kort tids utprøving. Hvert trinn i evalueringsprosessen bidrar til en systematisk og vitenskapelig oppbygging av en evalueringsmetodikk med evalueringdesign og analysemetoder som er egnet til å besvare de aktuelle forskningsspørsmålene som ligger til grunn for evalueringen.

Trinn 1: Teknologisk modenhet

Kontinuerlig teknologiutvikling og inkrementell forbedring av ITS-tjenester over tid gir vesentlige implikasjoner for evalueringsprosessen. De ulike fasene i teknologiens livsløp gir grunnlag for å besvare forskjellige forskningsspørsmål, og tiltakets teknologiske funksjonalitet og modenhet har dermed avgjørende betydning for senere valg av evalueringsmetode og analyser. Vurdering av teknologisk modenhet kan eksempelvis ta utgangspunkt i Technology Readiness Levels (TRL) som benyttes som felles grunnlag for vurdering av teknologistatus i EUs Horizon 2020-program. TRL-konseptet adresserer hele innovasjonsprosessen med en detaljert oversikt over ulike forsknings- og utviklingssteg fra tidlig konseptutvikling til ITS-tjenesten rulles ut på markedet, se Figur 32.

TRL 1	TRL 2	TRL 3	TRL 4	TRL 5	TRL 6	TRL 7	TRL 8	TRL 9
Basic principles observed	Technology concept formulated	Experimental proof of concept	Technology validated in lab	Technology validated in relevant environment	Technology demonstrated in relevant environment	System prototype demonstrated in operational environment	System complete and qualified	Actual system proven in operational environment
Concept development			Proof of principle			Proof of performance		

Figur 32: Technology Readiness Scale (TRL) for vurdering av teknologisk modenhet

Evaluering av ITS-tjenester i en tidlig fase av innovasjonsprosessen (typisk pilotstudier på TRL 3-5) kan ikke besvare de samme forskningsspørsmålene som evaluering av ITS-tjenester med en høyere grad av teknologisk modenhet (typisk implementeringsstudier på TRL 6-9).

Evalueringsdesignet (metodevalg og fremgangsmåte) må derfor tilpasses ITS-tjenestens modenhet og stadium i innovasjonskjeden, slik at evalueringen bidrar til å løfte tjenesten til neste nivå – eller i motsatt fall gir grunnlag for å forkaste videre utvikling. Tilgangen til hensiktsmessige evalueringsmetoder og datagrunnlag endrer seg etter nivå av teknologisk modenhet.

Evalueringsdesign på lavere nivå vil ofte være basert på ekspertvurderinger, modellering og simulering, mens feltforsøk og effektmåling krever et høyere nivå av teknologisk modenhet. Ex-post-evaluering basert på den eksperimentelle fremgangsmåten (inkludert før- og etteranalyser) er kun hensiktsmessig å gjennomføre når ITS-tjenesten har nådd en høy grad av teknologisk modenhet, når implementeringsfasen er fullført og tiltaket fungerer stabilt over en lengre periode.

Det er også verdt å merke seg at et ITS-tiltak kan være sammensatt av flere teknologier og tjenester som hver for seg innehar ulik grad av modenhet.

Trinn 2: Behovsanalyse

Som grunnlag for evalueringen må det gjøres en kartlegging av hva man søker å oppnå ved innføring av det aktuelle ITS-tiltaket; altså tiltakets bruksområde og målsettinger. Dette er særlig relevant når tiltak iverksettes med den teknologiske løsningen som utgangspunkt, uten at denne adresserer et definert behov eller samfunnsmessig utfordring (technology push). Behovsanalysen skal gi svar på hvilke underliggende problemer man ønsker å løse og/eller hvilke behov som skal dekkes, samt en spesifisering av hvilke målgrupper som berøres av tiltaket. Definisjon av tiltakets bruksområde (kontekst) og målsettinger kan blant annet skje på bakgrunn av møter med transportmyndigheter og interessegrupper, ekspertvurderinger (f.eks. Delphi-undersøkelser), intervjuer med brukere og spørreundersøkelser. Behovsanalysen gir grunnlag for utarbeidelse av programteori og evalueringskriterier senere i evalueringsprosessen, samt mulighet til å vurdere tiltakets relevans, for eksempel i et samfunnsperspektiv.

Trinn 3: Programteori

Utvikling av ITS-tiltakets *programteori* er en kritisk oppgave som danner grunnlaget for hele evalueringsprosessen. Programteorien er en konseptuell modell av sammenhengen mellom tiltakets innsats og virkninger, inkludert de underliggende mekanismene som utløser effekter på kort og lang sikt. Den beskriver altså *tiltakets underliggende logikk*; hvordan tiltaket er ment å operere og medføre endringer i en gitt kontekst. Dette representerer således prosjekteiers (ofte transportmyndighetenes) forventninger og rasjonelle grunnlag for innføring av ITS-tiltaket. Som regel inkluderer dette forventninger om effekter knyttet til trafiksikkerhet, trafikkavvikling, komfort, miljø, mobilitet og/eller økonomiske gevinster for bruker, næringsliv eller samfunnet for øvrig.

Programteorien utformes basert på resultater fra behovsanalysen og tilgjengelig kunnskap om tiltakets årsakvirkningskjeder, forventede effekter og virkninger på lang sikt. Den konseptuelle modellen kan inneha ulik grad av detaljering og kompleksitet, avhengig av ITS-tjenestens beskaffenhet, eksisterende kunnskapsgrunnlag og evalueringens formål. Med grunnlag i de to case-studiene er det laget et eksempel på en generell lineær programteori for ITS-tiltak, se Figur 33. Når tiltakets kompleksitet og evalueringens detaljfokus tiltar, kan denne utvikles til mer avanserte modeller som også kan fremstilles hierarkisk eller på matriseform.



Eksempler på relevante elementer i programteori for ITS-tiltak:



Figur 33: Skjematisk fremstilling av programteori for ITS-tiltak, med eksempler på relevante elementer.

Listen er ikke uttømmende, og forholdene kan være mer eller mindre relevant avhengig av tiltakets beskaffenhet og formålet med den enkelte evalueringen.

Figuren gir grunnlag for å utdype en rekke betingelser som må oppfylles for å lykkes med innføringen og intensjonen med tiltaket (boksene), samt hvilke underliggende mekanismer som bidrar til at betingelsene oppfylles i de ulike fasene av innføringsprosessen (pilene). Figuren viser også eksempler på forhold som kan inngå i en programteori for ITS-tiltak, med utgangspunkt i teoridiskusjonen og case-studiene i dette forskningsarbeidet.

Typiske rammebetingelser kan være knyttet til juridiske og organisasjonsmessige forhold, eksempelvis at ITS-tjenestene som inngår i tiltaket må oppfylle gjeldende krav til personvern. Avhengig av tiltakets kontekst, vil øvrige eksterne krav også legge premisser for innføring av tiltaket. Dersom tiltaket iverksettes i regi av transportmyndighetene, legges det typisk vekt på politisk aksept og transportpolitiske problemstillinger og retningslinjer. Private aktører som skal tjene penger på å tilby en ITS-tjeneste, vil i større grad være avhengig av sosial aksept, etterspørsel og forretningsmuligheter.

Når det gjelder selve tiltaket og anvendelsen av det, påvirkes utfallet både av teknologisk funksjonalitet og modenhet, teknisk ytelse og kvalitet, brukergrensesnittet, samt systemet det innføres i med nødvendig infrastruktur (f.eks. internetttilgang), støttende ITS-tjenester (f.eks. booking og betalingsløsninger) og øvrige virkemidler (f.eks. informasjonskampanjer, prismodeller og insentiver).

Når tiltaket er etablert, går man inn i en tidlig driftsfase som er kritisk med hensyn til å forstå hvordan brukerne responderer på tiltaket. Hvilke effekter man oppnår i driftsfasen vil være avhengig både av brukernes holdninger (motivasjonsfaktorer, tillit, om tjenesten dekker et uløst transportbehov), samt hvordan tiltaket utløser en faktisk atferdsendring (rekruttering, betalingsvilje og brukeraksept). I denne fasen inkluderes typisk erfaringer fra tidligere forskning på relevante ITS-tiltak.

Effektene man ønsker å oppnå med tiltaket, kan eksempelvis være endringer i reisetid, antall ulykker, reisemiddelvalg eller føreratferd – her vil detaljeringen av programteorien være avhengig av hva som finnes av eksisterende kunnskap og antakelser om tiltakets potensial.

Til slutt vil samfunnseffektene (virkninger) som oppnås i et lengre tidsperspektiv være avhengig av hvordan tiltaket implementeres og skaleres i transportsystemet. Tiltakets markedsutbredelse (spesielt når ITS-tiltaket er en frivillig tjeneste som kan velges (bort) av trafikantene), nettverkseffekter (kritisk brukermasse) og ringvirkninger for andre trafikanter vil ha stor betydning for utfallet av tiltaket på lengre sikt. ITS-tiltak innføres typisk for å oppnå gunstige virkninger på mobilitet, trafikkikkerhet, miljø, arealbruk eller økonomi.

Programteorien kan sees på som et dynamisk rammeverk som hele tiden justeres og oppdateres i løpet av ITS-tiltakets levetid. I et anvendt perspektiv, fungerer kunnskapsbyggingen optimalt, når resultater fra evalueringstudien tas inn i beslutninger og prioriteringer underveis (feedback loops). Slik kan tiltaket justeres for å optimalisere effektene i de ulike fasene, basert på kunnskapen som genereres om de relevante årsakssammenhengene i programteorien. For konkrete eksempler på programteorier, vises det til case-studiene i kapittel 6 og 7, og eksempler på anvendelse av evalueringsmetodikken i avsnitt 9.2.1 og 9.2.2.

Programteorien består av to sentrale dimensjoner som begge har betydning for ITS-tiltakets endelige utfall; *implementeringsteorien* som forklarer hvordan tiltaket skal implementeres i transportsystemet og *endringsteorien* som forklarer de underliggende mekanismene som utløser effekter (på kort sikt) og virkninger (på lengre sikt).

Implementeringsteorien – hvordan skal ITS-tiltaket implementeres?

Implementeringsteorien er en anbefalende (preskriptiv) fremstilling av hvilke betingelser som må være oppfylt for at ITS-tiltaket skal avstedkomme ønsket effekt og virkning. Denne beskrivelsen er knyttet opp mot selve implementeringen av tiltaket, og viser hvordan prosjektet er konstruert med et logisk resonnement for å aktivere endringene i neste punkt. I Figur 33 illustreres dette ved de

aktivitetene som skjer i selve boksene; rammebetingelser som må være oppfylt, tiltakets kapasitet (bruksområde, funksjonalitet og ytelse), umiddelbare effekter, samt de langsiktige virkningene man forventer å realisere. Et eksempel på en implementeringsbetingelse kan være at ITS-tjenesten forutsetter tilgang til en applikasjon på mobiltelefon, slik tilfellet var for samkjøringstiltaket i kapittel 7. Evalueringen kan da innebære en vurdering eller måling av om denne infrastrukturen er tilstrekkelig til stede for at tjenesten skal fungere optimalt (teknisk funksjonalitet og kvalitet).

Endringsteorien – hvorfor vil ITS-tiltaket forårsake en endring?

Endringsteorien er en beskrivende (deskriptiv) modell av de underliggende mekanismer som må inntreffe for at tiltaket skal medføre en endring. I Figur 33 illustreres dette ved de reaksjoner som skjer i pilene mellom boksene. Programteorien kan inneha ulike endringsteorier for ulike faser av tiltaket og for ulike målgrupper, men her inngår typisk faktorer som sosial aksept og brukeropplevelse. I eksempelet med samkjøringstiltaket, vil applikasjonens brukervennlighet være en av mange mekanismer som påvirker hvorvidt man oppnår en faktisk endring. Endringsteorien er gjerne tuftet på samfunnsvitenskapelige teorier (f.eks. psykologiske, sosiale, økonomiske, politiske modeller). Dersom man skal evaluere en ITS-tjeneste som forventes å påvirke trafikantenes reisemiddelvalg, vil man for eksempel benytte allerede kjente og velprøvde atferdsteorier for å vurdere sannsynligheten for at man oppnår en faktisk endring hos trafikantene (for eksempel Fishbein & Ajzen, 2010). Tilsvarende vil evaluering av trafikksikkerhetseffekter kunne dra nytte av Kulmala (2010) sin sjekklister med ni mekanismer for sikkerhetsvurdering av ITS, som presentert i kapittel 4.4. ITS-tjenestens brukeraksept studeres gjerne med utgangspunkt i modeller for aksept og bruk av teknologi, som presentert i kapittel 4.4.4.

Når programteorien er etablert, danner den grunnlaget for etterfølgende empirisk testing av de antakelser og forventninger som ligger til grunn for modellen. Rammeverket gir således en struktur for evalueringen, ved å identifisere kausale relasjoner og synliggjøre hull i kunnskapsgrunnlaget. Innledningsvis kan programteorien være forholdsvis lite detaljert og med mange kunnskapshull. Programteorien er imidlertid en foreløpig læringsramme som justeres og oppdateres etter hvert som sammenhengene i modellen bekreftes eller avkreftes. Resultatet av evalueringen blir dermed en mer sofistikert programteori som *forklarer hvordan tiltaket virker* basert på økt kunnskap og forståelse av ITS-tjenesten som studeres.

Trinn 4: Evalueringskriterier

Evalueringskriteriene gjenspeiler formålet med evalueringen (*hva man ønsker å etablere kunnskap om*) og hvordan nytten av ITS-tjenesten skal vurderes. Dette inkluderer typisk kriterier knyttet til relevans, måloppnåelse, bærekraft, produktivitet og effekter/virkninger. I denne fasen defineres hva evalueringen skal gi kunnskap om, med utgangspunkt i ITS-tjenestens programteori. Den visuelle modellen (programteorien) synliggjør kunnskapshull og vil også være et hjelpemiddel i prioritering av evalueringsressurser mellom ulike problemstillinger. Evalueringskriteriene må reflektere hvordan ITS-tjenesten er forventet å påvirke ulike målgrupper i henhold til behovsanalysen.

Eksempler på relevante evalueringskriterier⁵⁷:

- *Produktivitet*: Hvor effektivt kan man omsette ressurser til leveranser i form av resultatmål for kostnad, tid og kvalitet (herunder både teknisk kvalitet og tjenestekvalitet)?
- *Måloppnåelse*: I hvilken grad vil ITS-tjenesten bidra til at man oppnår ønskede effekter og oppfyller målsettinger innen eksempelvis trafikksikkerhet, trafikkavvikling, komfort, miljø og mobilitet?
- *Andre virkninger*: I hvilken grad vil ITS-tjenesten forårsake utilsiktede effekter (positive og negative) for ulike aktører?
- *Relevans*: I hvilken grad er ITS-tjenesten relevant i forhold til politiske prioriteringer, etterspørsel i markedet og interessegruppers behov?
- *Levedyktighet*: I hvilken grad vil de positive effektene vedvare i hele ITS-tjenestens levetid med utgangspunkt i økonomisk, miljømessig og politisk (fordelingsmessig) aksept?
- *Samfunnsøkonomisk effektivitet*: I hvilken grad vil ITS-tjenesten bidra til samfunnsøkonomisk lønnsomhet eller formålseffektivitet (sammenheng mellom måloppnåelse og ressursbruk).

Evalueringskriteriene uttrykkes i form av forskningsspørsmål eller hypoteser som videre brytes ned til et sett indikatorer, og som i neste omgang legger føringer for valg av evalueringsmetode og nødvendig datainnsamling. ITS-spesifikke guidelines (f.eks. EasyWay EEG (2012)) kan være til støtte i prosessen med å definere målbare indikatorer (KPIs). Det er viktig at formålet med evalueringen er avklart mellom involverte aktører og at det stilles forskningsspørsmål som faktisk kan besvares med hensyn til datatilgang, egnet metode og gjennomførbart evalueringsdesign.

⁵⁷ Eksempelene er basert på OECDs evalueringsmodell, videreutviklet av forskningsprogrammet Concept ved NTNU.

Trinn 5: Evalueringsdesign

Evalueringsdesignet beskriver evalueringens *fremgangsmåte* med hensyn til metodevalg, gjennomføring av datainnsamling, og analyser. Den teoribaserte evalueringsprosessen er metodenøytral og gir ikke føringer for spesifikt valg av metodeoppsett og analyseteknikk. Evalueringsdesignet defineres derfor med utgangspunkt i de metodene som er best egnet til å besvare de relevante forskningsspørsmålene, ut ifra formålet med evalueringen og ITS-tjenestens kompleksitet og teknologiske modenhet. Evalueringen kan i utgangspunktet inkludere og kombinere flere metoder og analyser (både kvalitative og kvantitative), og det forutsettes at evalueringsdesignet utformes på bakgrunn av kunnskap om de ulike metodenes styrker og svakheter.

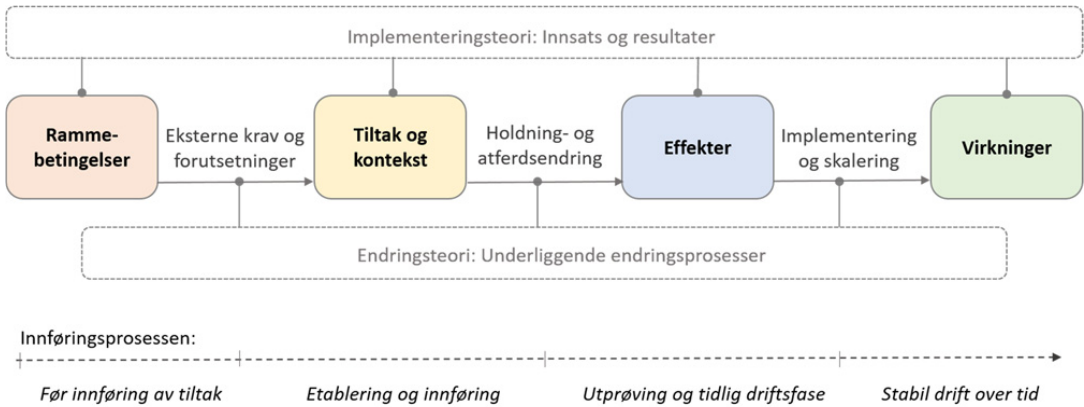
Det finnes i utgangspunktet tre hovedtyper evalueringsdesign, som alle kan benyttes i deler av den teoribaserte evalueringsprosessen (se ytterligere metodebeskrivelser i kapittel 3.4):

Randomisert eksperimentelt design: Innebærer at forsøkspersoner fordeles tilfeldig i en testgruppe og kontrollgruppe som sammenlignes både før og etter innføring av en ITS-tjeneste. Dette er en rigid metode som blant annet benyttes i kontrollerte laboratorieeksperimenter innenfor medisinsk forskning, og som ofte refereres til som "gullstandard" for metodeoppsett. Innenfor transportsektoren benyttes randomiserte eksperimenter først og fremst i kjøresimulator eller intervjustudier, da det er vanskelig å oppnå dette designet i evalueringer som foregår "i den virkelige verden".

Kvasi-eksperimentelt design: Følger det eksperimentelle metodeoppsettet med testgruppe og kontrollgruppe, men uten at forsøkspersoner fordeles tilfeldig i de to gruppene. Dette metodiske oppsettet representerer en lang tradisjon for evaluering av veg- og trafikktiltak innenfor transportsektoren.

Ikke-eksperimentelle design: Innebærer at ITS-tjenesten evalueres i en testgruppe, uten sammenligning mot en kontrollgruppe. Dette metodeoppsettet benyttes blant annet når det ikke er mulig å etablere en passende kontrollgruppe, enten på grunn av begrensede ressurser eller fordi tiltaket berører hele befolkningen. Et ikke-eksperimentelt evalueringsdesign krever at det etableres kunnskap om og forståelse av hvordan ITS-tjenesten forventes å forårsake effekter i en målgruppe (årsakssammenheng), for eksempel ved at evalueringen struktureres rundt en definert programteori. *Theory of change* og *realistisk evaluering* er eksempler på slike teoribaserte fremgangsmåter.

Dersom man sammenstiller ITS-tiltakets programteori med tiltakets innføringsprosess, gir dette en indikasjon på når det er relevant å gjennomføre datainnsamling og analyser knyttet til de ulike elementene i programteorien, som illustrert i Figur 34.



Figur 34: Generell programteori og ulike faser i ITS-tiltakets innføringsprosess

Tiltakets årsakssammenhenger vil kunne testes empirisk med ulike metoder og på ulike tidspunkt i innføringsprosessen. Eksempelvis vil måling av faktiske virkninger kreve at ITS-tiltaket har fungert stabilt over en lengre driftsperiode. Det presiseres at selv om den overordnede evalueringsmetodikken er basert på en teoribasert prosess, innebærer dette likevel en fleksibilitet i metodevalg som åpner for å kombinere konkrete eksperimentelle og ikke-eksperimentelle forskningsdesign, og både kvantitative og kvalitative metoder, for analyse av ulike forskningsspørsmål og elementer i programteorien.

Tabell 31 viser en oversikt over relevante betingelser ved valg av evalueringdesign og metoder.

Tabell 31: Valg av evalueringsdesign og metode


Evalueringsdesign	(Kvasi-) eksperimentelt	Teoribasert (ikke-eksperimentelt)	
		Theory of change	Realistisk
Metode	Kontrafaktisk analyse med statistiske metoder	Analyse av programteori	Analyse av endringsmekanismer
Formål	Bevise effekter	Finne effekter	Forklare effekter
Eksempel på forskningsspørsmål	Har man oppnådd forventet effekt i løpet av prosjektperioden?	Ex ante: Hvordan kan tiltaket bidra til å utløse effekter? Ex post: Har man oppnådd effekter og var det tiltaket som førte til disse?	Ex ante: Hva virker, for hvem, i hvilken kontekst og på hvilken måte? Ex post: Hvilke endringer har skjedd og hvordan har de oppstått?
Tidspunkt for evaluering	Etter implementering	Gjennom hele innovasjonsprosessen: Før, underveis eller etter innføring.	
Egenskaper ved tiltaket og omgivelser	Forventer å gi målbare resultater i løpet av prosjektperioden Stabilt og uforandret gjennom prosjektperioden Entydig årsak-virkning, og kan skilles fra andre tiltak Empiri etablert, få andre forklaringer på endring og robuste hypoteser	Langsiktig perspektiv på realisering av effekter Forventer små, kvalitative eller indirekte effekter Tiltaket er teknologisk umodent og/eller forventes å utvikle seg over tid Tiltaket innebærer komplekse årsakssammenhenger og/eller situasjonsavhengige effekter Dynamiske omgivelser og/eller variasjoner i implementeringen av tiltaket	
Metodiske begrensninger og muligheter	Representativt utvalg og kontrollgruppe Må være mulig å etablere baseline for kontrafaktiske analyser	Utforsker samspill mellom rammebetingelser, eksterne faktorer og utfall. Særlig egnet for svært komplekse, sammensatte tiltak (makronivå).	Utforsker underliggende mekanismer som utløser tiltakets effekter. Særlig egnet for tiltak med begrenset fokus og influensområde (mikronivå).

Trinn 6: Datainnsamling

Datainnsamlingen danner grunnlaget for analysene som skal gjøre det mulig å besvare forskningsspørsmålene, gjennom å teste antakelsene (hypotesene) som ligger til grunn for programteorien i trinn 3. Datainnsamlingen må derfor planlegges i relasjon til *evalueringsskriteriene* som er definert i trinn 4 og *evalueringdesignet* som er beskrevet i trinn 5. I trinn 6 utarbeides dermed en operativ plan for datainnsamlingen som identifiserer *hvilke* data som skal samles inn, *hvordan* dataene skal samles inn, samt en *tidsplan* for datainnsamlingen. Planen kan for eksempel benytte FESTA-metodikkens inndeling i prestasjonsindikatorer, valg av målemetoder og registreringer, samt bruk av sensorer, som presentert i kapittel 4.4.

I planleggingen av datainnsamlingen bør det inngå en vurdering av datakvalitet og egnethet for bruk i evalueringsdesignet som er beskrevet i trinn 5. Hva som er praktisk mulig å innhente av data vil være nært knyttet til ITS-tiltakets teknologiske modenhet, slik den er vurdert i trinn 1, og ikke minst til evalueringens rammebetingelser i form av tilgjengelige ressurser. Dette ble blant annet erfart i studiet av NonStop-tiltaket, hvor planen for datainnsamling ble justert flere ganger i løpet av prosjektperioden (se kapittel 6.5.5). I praksis er det sjelden mulig å innhente et komplett datagrunnlag med alle datatyper og av det omfanget som man ideelt ønsker seg. Det er derfor nødvendig å foreta en kunnskapsbasert prioritering av de mest relevante datatyper med riktig kvalitet i henhold til evalueringsdesign og med kapasitet til å besvare de viktigste forskningsspørsmålene.

Figur 35 viser et eksempel på hvordan et utvalg datainnsamlingsmetoder kan relateres til ulike faser av innovasjonsprosessen. Oversikten er utarbeidet i NordicWay-prosjektet og tidligere presentert i Lervåg (2016).

TRL	Phase of C-ITS development	Data collection method
1, 2, 3	Concept development	 <ul style="list-style-type: none"> Checklist Delphi/foresight Modeling Simulation Laboratory testing Field trials Impact monitoring
4, 5, 6	Proof of principle	
7, 8, 9	Proof of performance	
All	User needs and user acceptance	

Figur 35: Sammenhengen mellom ITS-tjenestens teknologiske modenhet og aktuelle datainnsamlingsmetoder i NordicWay-prosjektet

Innovasjonsprosessen er her representert ved nivåene på TRL-skalaen, som tidligere presentert i Figur 32. Som indikert i figuren er det ikke et entydig forhold mellom TRL-nivå og datainnsamlingsmetode, derimot anbefales det å kombinere flere datatyper for å dekke alle evalueringskriterier og indikatorer.

Evaluering av ITS-tiltak kan i mange tilfeller dra fordel av at ITS-tjenestene i seg selv samler inn en rekke datatyper som tidligere ikke har vært tilgjengelig innenfor transportsektoren. Dette åpner for analyser og evalueringer som i større grad kan forklare og dokumentere hvordan ITS-tjenesten utløser en rekke mekanismer og effekter i transportsystemet. Det stiller imidlertid også større krav til den som skal utføre analysene, fordi man nå står overfor muligheter og datatyper som man har begrenset erfaring med å håndtere (f.eks. big data, usikker datakvalitet, personvern hensyn og behov for nye analysemetoder).

Trinn 8: Analyser

Valg av analysemetoder avgjøres i realiteten ved utforming av evalueringsdesign og planlegging av datainnsamling. Hvilke analyser som prioriteres vil i stor grad avhenge av forskningsspørsmålene som er definert i en tidligere fase av evalueringsprosessen. For de fleste ITS-tjenester har man i utgangspunktet begrenset kunnskap om effektbildet, og det vil som regel være nødvendig å kombinere både kvantitative og kvalitative analysemetoder for å forstå fullt ut hvordan tiltaket fungerer. De kvalitative analysene bidrar gjerne til å identifisere relevante effekter og virkninger (både de som er forventet og eventuelt utilsiktet), og til å skape forståelse for de mekanismene som utløser disse effektene (f.eks. implementeringsstrategi, motivasjonsfaktorer og brukeraksept). De kvantitative analysene gir i større grad kunnskap om effektens størrelse og konfidensintervall, gjennom mulighet for statistisk testing.

Hvilket beviskrav som stilles med hensyn til vurdering av kausalitet avhenger av evalueringsdesignet som er valgt i trinn 5. Mens den eksperimentelle metodikken legger kontrafaktiske analyser av testsituasjoner med og uten tiltaket til grunn for tolkningen av nytteeffekter, har den teoribaserte evalueringen i større grad en mekanistisk tilnærming med fokus på å teste de ulike sammenhengene i programteorien. I praksis vil større evalueringer som regel kombinere både eksperimentelle metoder i studier av teknisk ytelse og kvalitet, atferdsendringer og effekter der dette er mulig å måle, og teoribaserte fremgangsmåter som gir kunnskap om årsakssammenhenger, rammebetingelser og eksterne faktorer som påvirker i hvilken grad man vil lykkes med ITS-tjenesten i ulike situasjoner.

Siden det foreligger lite empiri om nytten av ITS-tjenester, bør analysene også inkludere vurderinger av usikkerhet og følsomhet for endringer i tiltaket eller omgivelsene: Hva betyr det om antakelsene som ligger til grunn for tiltakets programteori er feil? Hvilket utslag gir det om størrelsen på brukergruppen (markedspenetrasjon) blir mindre enn forutsatt? Og hvordan endres lønnsomheten dersom teknologien forbedres vesentlig?

For de fleste evalueringsformål vil det være relevant å gjøre en vurdering av samfunnsøkonomisk lønnsomhet i en eller annen form. Innenfor transportsektoren er dette tradisjonelt knyttet opp mot nyttekostnadsanalyser, men det finnes også andre tilnærminger som i større grad håndterer og inkluderer kvalitative variabler som ikke kan uttrykkes monetært, eksempelvis multikriterieanalyser. Evalueringsmetodikken legger ingen begrensninger til valg av analysemetode, utover at den som gjennomfører nytteevalueringen må være kjent med de premisser og forbehold som ligger til grunn for ulike metodevalg.

Trinn 9: Rapportering av resultater

Resultatene fra effektevalueringen bør forklares og presenteres i en form som beskriver hvordan evalueringen bidrar til utvikling av ny kunnskap om ITS-tiltaket som studeres, samt også bidrar til å gjøre kunnskapen anvendbar for fremtidige implementeringer. Det anbefales at resultatrapporten som et minimum inneholder en beskrivelse av grunnlaget for beslutninger i de ulike trinnene i evalueringsmetodikken:

- a) Beskrivelse av selve ITS-tiltaket og nødvendig infrastruktur, hvilke ITS-tjenester og virkemidler som inngår i tiltaket, teknologisk funksjonalitet, begrensninger og modenhet, teknisk kvalitet og ytelse, samt brukergrensesnitt (HMI).
- b) Beskrivelse av ITS-tiltakets kontekst og målsettinger, i form av rammebetingelser og lokal implementeringsstrategi, med blant annet bruksområde, målgrupper og problemstillinger/behov som skal dekkes.
- c) Beskrivelse av programteorien som ligger til grunn for evalueringen; ITS-tiltakets hensikt og hvordan det virker (forventning om sammenheng mellom innsats og effekter/virkninger).
- d) Beskrivelse av evalueringskriterier og forskningsspørsmål som ligger til grunn for analysene.
- e) Beskrivelse av evalueringsdesign, metodevalg og datainnsamling, inkludert eventuelle avvik og begrensninger.
- f) Beskrivelse av analyseteknikk og beviskrav (f.eks. statistisk signifikans), samt analyseresultater
- g) Vurdering av usikkerhet og følsomhet
- h) Overførbarhet til andre situasjoner og implementeringer

Resultatene fra evalueringen kan gjerne visualiseres i form av en ny og oppdatert programteori som synliggjør hvilke sammenhenger som har blitt styrket gjennom empirisk testing, hvilke antagelser som eventuelt ikke er blitt oppfylt, og hvilke kunnskapshull som krever videre forskning.

Diskusjonen av resultatene må gi grunnlag for å vurdere om ITS-tiltaket er realistisk og hensiktsmessig innenfor det bruksområdet som er studert. Dette innebærer en drøfting av hvordan ITS-tiltaket fungerer på ulike organisatoriske nivå:

Operasjonelt: Fungerer ITS-tiltaket effektivt? Hvordan kan eventuelle justeringer bidra til optimalisering av teknisk og funksjonell ytelse?

Taktisk: Realiserer ITS-tiltaket de tiltenkte effektene for målgruppen? Har man avdekket uønskede effekter? Hvilke betingelser og eksterne faktorer påvirker graden av måloppnåelse?

Strategisk: Er ITS-tiltaket relevant og nyttig i et samfunnsperspektiv?

9.2 EKSEMPEL PÅ ANVENDELSE AV EVALUERINGSRAMMEVERKET

Det pågående doktorgradsarbeidet har gitt bidrag til to uavhengige forskningsprosjekter (NordicWay og SmartFeeder), hvor den teoribaserte evalueringsprosessen er lagt til grunn for evaluering av ITS-tjenester. De påfølgende avsnittene gir en kort beskrivelse av hvordan rammeverket er anvendt i prosjektene.

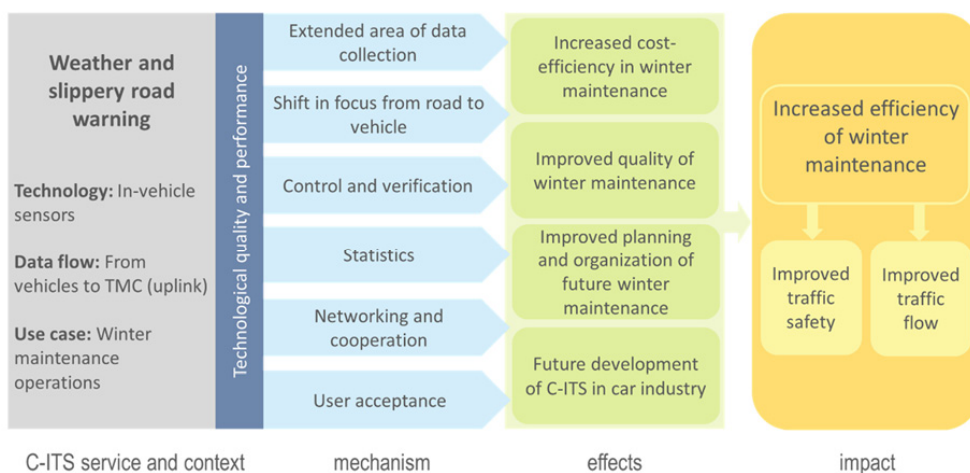
9.2.1 NORDICWAY

NordicWay er et treårig EU-prosjekt som ble gjennomført i perioden 2015-2017, med transportmyndighetene i Norge, Finland, Danmark og Sverige som samarbeidspartnere. Hensikten med prosjektet var å pilotere innovative C-ITS-tjenester på nordiske veier, med sømløs deling av trafikksikkerhetsinformasjon over mobiltelefonnett og mellom landegrensene. C-ITS-tjenestene omfattet varsling av en rekke kategorier data, med a) *varsling av glatt vegbane* og b) *varsling om farlige hendelser* som de mest sentrale. Det ble etablert en skyløsning (NordicWay Interchange) som sørget for å distribuere meldinger mellom kjøretøy, telefonapplikasjoner og baksystem. Som del av prosjektet ble det utarbeidet en NordicWay Evalueringshåndbok (Lervåg, 2016) som skisserte rammeverket for evalueringsstudier i de respektive landene. I prosjektsøknaden var det lagt opp til å følge et felles eksperimentelt evalueringsdesign med sammenlignbare før- og ettermålinger av et sett indikatorer (som f.eks. reisetid, alvorlige ulykker og CO₂-utslipp) i tråd med EasyWay Evaluation Guidelines⁵⁸. Det ble imidlertid raskt tydelig at det eksperimentelle evalueringsdesignet ville medføre en rekke metodiske utfordringer. For det første var det vesentlige forskjeller i C-ITS-tjenestenes funksjonalitet og bruksområde. I den norske piloten var målsettingen å effektivisere vinterdriften (f.eks. brøyting og strøing) og teknologien som lå til grunn for estimering av vegfriksjon var bilens egen sensorikk. Dataene ble så kommunisert via mobilnett til et baksystem (Traffic Management Center), hvor informasjonen ble bearbeidet og distribuert til målgruppen (f.eks. entreprenører). Den finske piloten var derimot basert på en applikasjon på mobiltelefonen, hvor bilførere selv rapporterte om opplevd glatt vegbane, og på grunnlag av disse meldingene ble det generert varslinger som ble distribuert til andre bilførere i samme område. Selv om både norske og finske transportmyndighetene tok utgangspunkt i en C-ITS-tjeneste for varsling av glatt vegbane, er det ulike årsaksvirkningskjeder og effekter som kan studeres i pilotene. Forskjeller i ressurstilgang og prioriteringer mellom landene la også føringer for hvordan pilotene ble gjennomført. Eksempelvis hadde den danske piloten størst fokus på teknisk funksjonalitet og dokumentasjon av hvordan trafikkmeldingene ble sendt fra baksystemet til kjøretøyet, uten at anvendelsen av disse datameldingene ble satt inn i en større kontekst. Til sist har den

⁵⁸ Se beskrivelse av EasyWay Evaluation Guidelines i kapittel 4.4.

eksperimentelle metoden også begrensninger med hensyn til hva som er realistisk å oppnå av målbare effekter i en begrenset pilotperiode, og analyser av indikatorer som alvorlige hendelser og endringer i reisetid krever en betydelig markedspenetrasjon.

For å kunne håndtere C-ITS-tjenestenes kompleksitet og ivareta fleksibilitet med hensyn til metodevalg, ble det lagt opp til en teoribasert evalueringsprosess. På denne måten kunne hvert land skreddersy et egnet evalueringsdesign, som sikret at evalueringen kunne avstedkomme relevant og anvendbar kunnskap, ut ifra den enkelte C-ITS-tjenestens beskaffenhet og bruksområde. Evalueringshåndboken beskriver en trinnvis prosess med i) vurdering av teknologisk modenhet, ii) definisjon av problemstillinger (behovsanalyse), iii) etablering av programteori, iv) definisjon av evalueringskriterier, v) valg av evalueringsdesign, vi) datainnsamling, vii) analyser og viii) rapportering av resultater, i tråd med evalueringsprosessen som er foreslått i kapittel 9.1. Samtidig ble det lagt vekt på å ivareta hensiktsmessige evalueringsprinsipper og god vitenskapelig praksis fra retningslinjene i EasyWay Guidelines (EasyWay EEG, 2012; Lervåg, 2016). Figur 36 viser et eksempel på en programteori som ble utarbeidet for tjenesten *varsling av glatt vegbane* i den norske evalueringsstudien.



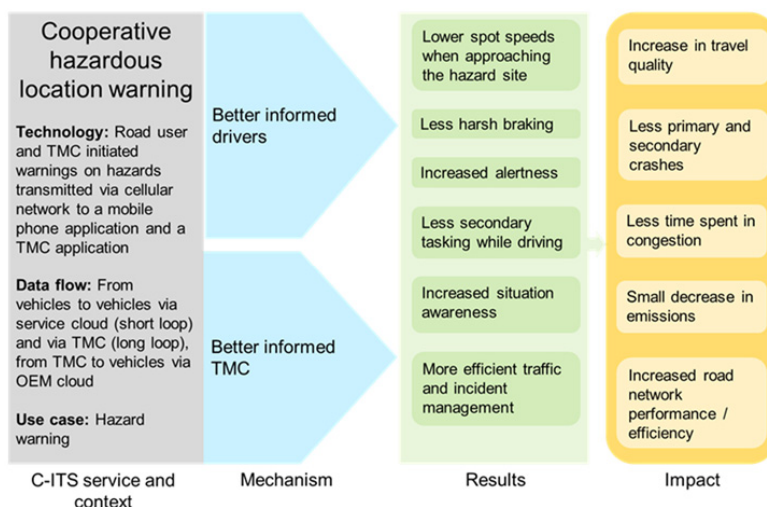
Figur 36: Eksempel på programteori for den norske C-ITS-tjenesten for varsling av glatt vegbane i NordicWay-prosjektet. Som figuren viser, er tjenesten rettet inn mot overordnet beslutningstaking og kontraktsoppfølging av vinterdrift - som indirekte og over tid kan påvirke trafiksikkerhet og fremkommelighet.

De store variasjonene i C-ITS-tjenestenes beskaffenhet og bruksområde resulterte i noe ulike evalueringskriterier og metodevalg i de nasjonale evalueringsstudiene. Dette gjorde også at effektene av C-ITS-tjenestene ikke er sammenlignbare mellom de ulike pilotene. I stedet har

pilotene bidratt med supplerende kunnskap som til sammen gir en utfyllende forståelse av C-ITS-tjenestenes funksjonalitet og muligheter.

NordicWay-prosjektet demonstrerte at kommunikasjon av trafikksikkerhetsmeldinger kan fungere sømløst over mobilnettet og mellom landegrenser, og at datautvekslingen gir grunnlag for C-ITS-tjenester som varsler om glatt vegbane og farlige hendelser. Til tross for at ulike teknologivalg og bruksområder begrenset overførbareheten mellom pilotene, har anvendelsen av et felles rammeverk for evalueringen sikret en systematisk og transparent gjennomføring av de nasjonale aktivitetene. Resultatene fra de nasjonale evalueringsstudiene ga til sammen et helhetlig bilde av C-ITS-tjenestenes potensielle effekter, suksesskriterier og barrierer (Bjerkan, 2017). Denne kunnskapen er nyttig i det videre arbeidet med utrulling av tjenestene i større skala.

I etterkant av prosjektet er erfaringene med evalueringsmetodikken i NordicWay videreført til andre relevante fagmiljø. I slutfasen av dette doktorgradsarbeidet er det blant annet registrert at det EU-finansierte nettverket C-Roads Platform har inkludert erfaringer fra NordicWay i sine anbefalinger (C-Roads, 2019). C-Roads Platform er et felles initiativ for europeiske medlemsstater og vegmyndigheter for testing og implementering av samvirkende ITS-tjenester, med hensikt å oppnå harmonisering og interoperabilitet på tvers av landegrenser. Som del av arbeidet, er det utarbeidet en felles evalueringsplan for pilotstudier av samvirkende ITS-tiltak. Evalueringsplanen inkluderer blant annet et eksempel med bruk av programteori fra den finske evalueringsstudien i NordicWay-prosjektet – som igjen er direkte basert på resultatene fra dette doktorgradsarbeid (C-Roads, 2019; Innamaa, Koskinen, & Kauvo, 2017). Se Figur 37.



Figur 37: Programteori utviklet for den finske piloten i NordicWay-prosjektet (Innamaa et al., 2017), og gjengitt i C-Roads (2019)

9.2.2 SMARTFEEDER

SmartFeeder er et pågående forskningsprosjekt (2017-2020) som skal gi kunnskap om hvordan bruk av automatiserte tilbringertjenester kan bidra til en grønn omstilling hos trafikanter (Lervåg et al., 2018). Prosjektet følger innføringen av de første norske pilotene med selvkjørende, elektriske minibusser, for å se hvordan disse kan styrke kollektivtransporten gjennom å tilby et sammenhengende og attraktivt reisetilbud fra dør til dør. Prosjektet er delfinansiert av Norges forskningsråd, og gjennomføres i samarbeid mellom offentlige og private aktører, med Jernbanedirektoratet som prosjekteier og SINTEF som prosjektleder og forskningspartner.

Gjennom studier av rammebetingelser, brukeraksept, forretningsmodeller, samt evaluering av reiseatferd og samfunnseffekter skal SmartFeeder-prosjektet utarbeide anbefalinger som muliggjør en god og smidig innføring av smarte tilbringertjenester i det norske transportsystemet.

Kunnskapsbyggingen i SmartFeeder er basert på den teoribaserte evalueringsprosessen som er foreslått i kapittel 9.1. Den viktigste begrunnelsen for valg av evalueringsmetodikk er muligheten til å håndtere en kompleks tjeneste i stadig utvikling, samt ønsket om å etablere kunnskap om de kontekstuelle faktorene som påvirker hvorvidt man lykkes med implementeringen. Selvkjørende minibusser er en forholdsvis ny tjeneste i verdenssammenheng, og det forventes at både teknologiutvikling og fortløpende tilpasning av implementeringen vil sørge for en stadig forbedring av tjenesten i løpet av prosjektperioden. For øvrig finnes foreløpig begrenset empiri om årsakssammenhenger, og det er derfor behov for en eksplorativ tilnærming til mekanismene som forårsaker effekter og virkninger.

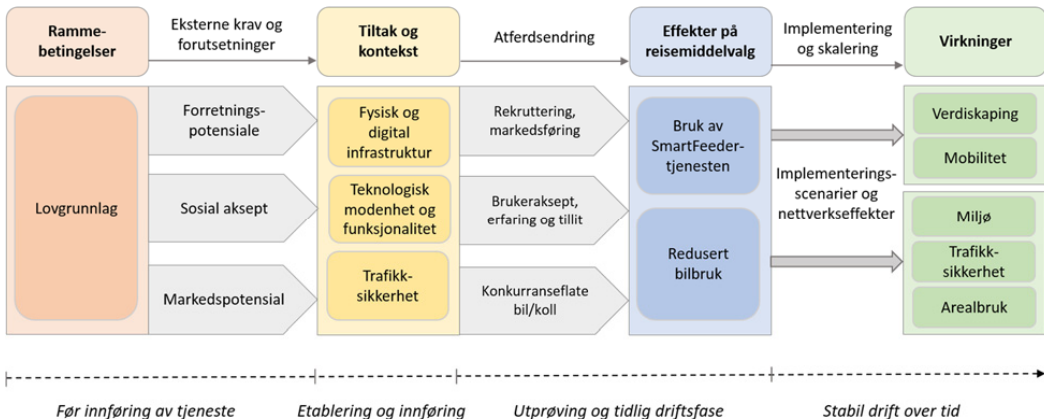
SmartFeeder-prosjektet følger fem utvalgte norske piloter i oppstartsfasen. Selv om alle disse omfatter et konsept som har til hensikt å styrke kollektivtrafikken gjennom etablering av et dør-til-dør-tilbud, har de likevel noe ulik innretning med hensyn til omfang og målsettinger, som vist i Figur 38. *Foruspiloten* i Stavanger var den første norske selvkjøringpiloten som ble satt i drift etter at lovverket åpnet for uttesting av selvkjørende kjøretøy på offentlig veg, og pilotaktivitetene har lagt vekt på å demonstrere trafiksikkerhet og oppnå brukeraksept. På *Fornebu* skulle piloten avhjelpe et eksisterende trafikk- og parkeringsproblem, mens ambisjonen på lengre sikt har vært å utforske hvordan smarte tilbringertjenester kan bidra til en mindre privatbilavhengig by- og boligutvikling. På *Gjøvik* opplevde man vesentlige driftsutfordringer på grunn av krevende topografi, og fokuset ble dermed rettet mot den fysiske og digitale infrastrukturen, med teknologisk funksjonalitet og modenhet. *Kongsbergpiloten* er blant de mest avanserte selvkjøringpilotene på verdensbasis, med kjøring i et komplekst trafikkbilde under norske vinterforhold. På lengre sikt har denne piloten en tydelig målsetting om å stimulere til verdiskaping for norsk industri, med satsing

på innovasjon og utvikling av bærekraftige forretningsmodeller. *Ruterpiloten* har hatt en annen tilnærming enn de øvrige, ved at den sikter mot en storskala systemtest, med vekt på hvordan selvkjørende minibusser påvirker mobilitetsbehov og byutvikling i et mer overordnet perspektiv.



Figur 38: Innretning og omfang på de fem første pilotene med selvkjørende minibusser i Norge. Passasjertall og utkjørt distanse per november 2019.

Det forventes ikke at de ulike pilotstudiene vil produsere direkte sammenlignbare resultater, men at de heller vil bidra med supplerende erfaringer som til sammen gir utfyllende kunnskap om suksesskriterier og barrierer for vellykket implementering av smarte tilbringertjenester i det norske transportsystemet. Som del av evalueringsprosessen (trinn 3) er det utarbeidet en programteori for smarte tilbringertjenester, som vist i Figur 39.



Figur 39: SmartFeeder evalueringsmodell – programteorien for konseptet Smarte tilbringertjenester

Programteorien tar utgangspunkt i at visse rammebetingelser, eksterne krav og forutsetninger må være oppfylt for at det i hele tatt skal være aktuelt å etablere en SmartFeeder-tjeneste. De viktigste rammebetingelsene er knyttet til etableringen av et lovgrunnlag for testing av selvkjørende kjøretøy (som trådte i kraft i januar 2018), med en tilfredsstillende søknadsprosess mot tillatelse for kjøring på offentlig veg for pilotene. Videre legges det til grunn at det må finnes et forretningspotensial med kommersielle verdier som gjør det attraktivt for private tilbydere å etablere og drifte SmartFeeder-tjenesten. Dersom man skal lykkes med å attrahere brukere, må det finnes en viss sosial aksept for selvkjørende teknologi, samtidig som tjenesten dekker et reelt transportbehov i befolkningen.

Vellykket innføring av SmartFeeder-tjenesten er avhengig av teknologisk funksjonalitet og modenhet – at tjenesten fungerer under norske vær- og trafikkforhold, og at den etableres i et helhetlig system med nødvendig infrastruktur (f.eks. ladestasjoner for bussene) og støttende ITS-tjenester (f.eks. ruteopplysning og billettbetaling). Hensikten med innføring av SmartFeeder-tjenester er å oppnå effekter på reisemiddelvalg, i form av SmartFeeder-passasjerer og redusert bruk av privatbil. Dette forutsetter en rekke mekanismer som utløser atferdsendring hos trafikantene; for det første må man lykkes med å informere og rekruttere brukere til den nye tjenesten, man må oppnå aksept og tillit hos faktiske brukere, og helst må tjenesten oppleves så attraktiv at den velges på bekostning av privatbil, og ikke av de som ellers ville syklet eller gått til fots (må altså påvirke konkurranseflaten mellom bil og kollektiv). På lengre sikt vil potensielle samfunnseffekter i form av verdiskaping for norsk industri, økt mobilitet for befolkningen, redusert miljøbelastning, bedre trafiksikkerhet og gunstig arealbruk/byutvikling være avhengig av hvordan tjenesten implementeres i større skala, når pilotperioden med innføring og tidlig driftsfase er over.

Kunnskap om SmartFeeder-konseptet etableres ved at de ulike mekanismene og betingelsene i programteorien studeres og testes empirisk i pilotprosjektene, eksempelvis på grunnlag av data fra dokumentgjennomgang, intervju med sentrale aktører, datalogg fra de selvkjørende minibussene, drifts- og passasjerstatistikk, samt ulike brukerundersøkelser.

9.3 HVA KAN RAMMEVERKET TILFØRE FAGOMRÅDET?

Dette kapittelet har presentert et analytisk verktøy for evaluering av ITS-tiltak, med utgangspunkt i et teoribasert evalueringsperspektiv. Rammeverket tilbyr fagfeltet en alternativ overordnet evalueringsstrategi til den tradisjonelle metodestyrte evalueringspraksisen som er dominerende innenfor transportområdet. Bruk av denne teoribaserte evalueringsprosessen vil styrke kunnskapsgrunnlaget for utvikling og innføring av ITS-tiltak på flere måter:

- Det bidrar til struktur og systematikk rundt metodevalg og prioritering av evalueringsressurser
- Det utvider metodegrunnlaget ved å inkludere flere former for validitet og kausalteori
- Det skifter fokus fra evaluering som verktøy for kontroll og styring, til evaluering som verktøy for innovasjon og kompetanseutvikling

Selv om mange tradisjonelle ITS-evalueringer i praksis kombinerer et bredt spekter av metoder, er det likevel forskjell på den overordnede evalueringsstrategien som igjen er forankret i litt ulike vitenskapsteoretiske utgangspunkt. Klassiske effektevalueringer undersøker om et tiltak virker gjennom etablering av en kontrafaktisk situasjon ved bruk av kontrollgrupper. Forsøksdesignet er innrettet mot å finne ett gyldig resultat: Effekten av tiltaket på et gitt tidspunkt (om tiltaket virker eller ikke). I denne tilnærmingen er de kvantitative analysene dominerende, og kvalitative tilleggsanalyser har ofte som formål å øke de kvantitative resultatenes validitet når metodiske betingelser ikke er oppfylt.

I det teoribaserte evalueringsperspektivet vurderes tiltaket etter flere kriterier enn ren effektmåling, og innsatsen styres i større grad mot å etablere kunnskap om suksessfaktorer som må styrkes eller barrierer som må overvinnes for å lykkes med tiltaket. Metodevalg og utvikling av (flere) forsøksdesign er en integrert del av evalueringsprosessen, som åpner for at hele eller deler av evalueringen kan baseres på en mekanistisk kausalitetsforståelse. Dette innebærer at evalueringen ikke nødvendigvis produserer *effektbevis*, men heller sannsynliggjør at tiltaket har hatt eller vil utløse en effekt under gitte forutsetninger. Samtidig gir den innsikt i tiltakets forbedringspotensial. Dette tilrettelegger for en dynamisk innovasjonskultur hvor man lærer av feil og hele tiden søker etter nye og bedre løsninger.

Rammeverket har vist seg å være anvendbart; det er mulig å benytte for tiltak med høy kompleksitet, også når forutsetningene for klassiske effektstudier ikke er oppfylt. Det kan benyttes gjennom hele beslutningsprosessen fra tidlig konseptutvikling til tiltaket er implementert og fungerer i stabil drift. Hvor tiltaket befinner seg i dette utviklingsløpet påvirker imidlertid hvilke evalueringskriterier som er relevante, hvilke forskningsspørsmål som er mulig å besvare, og hvilke metoder som er hensiktsmessig å benytte i evalueringsdesignet. Etter hvert som ITS-tiltaket utvikles, øker også utbudet av hensiktsmessige evalueringsmetoder.

Rammeverket gir evalueringsarbeidet struktur og synliggjør metodevalg og prioritering av utvalgte forskningstema. Tilnærmingen er pragmatisk i den forstand at den styrer innsatsen fra et ideelt (men kanskje uopnåelig) til et best mulig metodeoppsett. På dette viset tas det høyde for at det ikke vil være praktisk mulig å besvare alle relevante forskningsspørsmål til enhver tid, for eksempel å

fremskaffe effektbevis for en fremtidig implementering av et ITS-tiltak under utvikling. Den teoribaserte evalueringsprosessen sikrer at beslutningstakere og tjenesteutviklere får kunnskap som de kan nyttiggjøre seg på det stadiet man til enhver tid befinner seg i innføringsløpet.

I kapittel 4.1 ble det presentert utfordringer knyttet til evaluering og innføring av ITS-tiltak på tre ulike nivå: i) Beslutningsparadigmet i transportsektoren som favoriserer kunnskap fremskaffet gjennom klassiske effektstudier, ii) en sterkt metodestyrt evalueringsskulture som i noen tilfeller kan lede til ukritisk bruk av den eksperimentelle evalueringsmodellen, og iii) den eksperimentelle modellens egnethet med hensyn til håndtering av typiske egenskaper og relevante problemstillinger ved ITS-tiltak. Dette rammeverket adresserer utfordringene i de to sist nevnte kategoriene.

10 BEMERKNINGER OG REFLEKSJONER

Avslutningsvis anføres det i dette kapittelet noen bemerkninger og refleksjoner rundt det foreslåtte evalueringsrammeverket og kunnskapen som er fremkommet av doktorgradsstudiet. I første del presenteres noen tanker om hvordan bruk av den teoribaserte evalueringsmetodikken kan påvirke fagfeltets evalueringspraksis, med hensyn til metodebruk og kunnskapsproduksjon. Deretter gjøres det rede for hvordan innledende forskningsspørsmål er blitt besvart og hvordan arbeidet har bidratt med original kunnskap på fagfeltet. Til sist drøftes metodiske erfaringer og begrensninger, før kapittelet rundes av med noen synspunkter på videre forskning og utvikling av fagområdet.

10.1 IMPLIKASJONER FOR FAGOMRÅDET

Den teoribaserte evalueringsprosessen er et nyttig rammeverk for gjennomføring av ITS-evalueringer. Først og fremst tilbyr den en struktur som sikrer at ITS-tiltaket settes inn i en relevant kontekst, med en vitenskapelig forankring av metodevalg i henhold til forskningsspørsmålene som søkes besvart. Tilnærmingen åpner det vitenskapelige mulighetsrommet for effektstudier utover det eksperimentelle evalueringsdesignet som har stått sterkt i den transportfaglige tradisjonen. I praksis betyr dette at evalueringen kan besvare komplekse problemstillinger – og dermed bidra til mer helhetlig kunnskap både om tiltakets (potensielle) effekter og hvordan tiltaket bør innføres i transportsystemet. Verktøykassa utvides altså med flere ikke-eksperimentelle metoder (f.eks. theory of change og realistisk evaluering), uten at dette begrenser muligheten til å inkludere eksperimentelle studier når dette finnes hensiktsmessig.

Fleksibiliteten i metodevalg stiller imidlertid visse krav til den som skal gjennomføre evalueringen. Evalueringsmetodikken har ikke ett predefinert evalueringsdesign med evalueringsanalyser som kan gjennomføres oppskriftsmessig. Rammeverket tilbyr derimot en standardisert og veiledet prosess mot en individuelt tilpasset evalueringsmetodikk basert på tiltakets teknologiske beskaffenhet, relevante problemstillinger, hensiktsmessige analysemetoder, tilgjengelig datagrunnlag og praktiske muligheter i hvert tilfelle. Den som gjennomfører evalueringen må dermed inneha tilstrekkelig metodekunnskap, samt teknologisk- og transportfaglig kompetanse for å kunne foreta gode, kunnskapsbaserte valg underveis i evalueringsprosessen. Eksempelvis er etableringen av tiltakets programteori et kritisk punkt som forutsetter god forståelse for teknologiens funksjonalitet og begrensninger, samt kunnskap om hvordan trafikanter og transportsystemet responderer. Til gjengjeld sikrer dette at evalueringen avstedkommer velfundert og relevant kunnskap for beslutningstakere, også når ITS-tiltaket er umodent eller det empiriske datagrunnlaget er begrenset.

Den teoribaserte evalueringsmetodikken kan benyttes gjennom hele beslutningsprosessen fra tidlig konseptutvikling til tiltaket er implementert i ordinær drift. Selv om tidlige evalueringer ikke gir grunnlag for å dokumentere faktiske effekter, kan de ha stor praktisk og økonomisk betydning. I begynnelsen av innføringsløpet har man større mulighet til å prioritere utviklingsressursene, ved å foreta nødvendige justeringer eller forkaste lite bærekraftige konsepter, før det legges inn store investeringer. På denne måten er den teoribaserte metodikken godt egnet for ITS-faget, som i stor grad karakteriseres av innovasjonsprosjekter med kompetanseutvikling gjennom pilotering, prøving og feiling. Metodikkens vektlegging av tiltakets underliggende logikk, bidrar dessuten til å sikre tiltakets relevans i et samfunnsperspektiv (fra technology push til fokus på brukerbehov). Evalueringer i en tidlig fase av beslutningsprosessen kan således hindre at man låser seg til ITS-tiltak som viser seg å være vanskelig å realisere eller lite relevante. Evalueringsmetodikkens anvendbarhet gjennom hele tiltakets livssyklus, kan få konsekvenser for hvordan man foretar evalueringer i transportsektoren. I dag gjennomføres de fleste evalueringer som punktvis hendelser som gir et øyeblikksbilde av tiltakets observerbare effekter i transportsystemet. Dette har begrenset relevans og gyldighet for ITS-tiltak som utvikles fortløpende og som påvirkes av en rekke eksterne faktorer. Når formålet er læring og kunnskapsbygging, vil det være mer hensiktsmessig å gjennomføre evalueringen som en pågående aktivitet gjennom hele tiltakets innføringsløp. På denne måten nyttiggjøres evalueringens formative kapasitet, ved at resultater og ny kunnskap fortløpende gir innspill til veivalg og forbedringer underveis i prosessen.

Når evalueringsaktivitetene trekkes tettere inn i utviklings- og beslutningsprosessen for ITS-tiltaket, kan det stilles spørsmål ved evalueringens objektivitet. Det er grunn til å reflektere over hvordan denne ivaretas når evalueringen skifter karakter fra å være en bedømmende kraft på utsiden av tiltaket, til å være med å legge premissene for hvordan tiltaket utformes og implementeres. Den teoribaserte evalueringsprosessen gir i tillegg flere frihetsgrader i gjennomføringen, blant annet med hensyn til definisjon av forskningsspørsmål og metodevalg. Til sist vil et tett samarbeid mellom prosjekteier og den som gjennomfører evalueringen gi risiko for personlige bindinger. Kravet til objektivitet er til stede i all vitenskapelig forskning, og dels foregår disse diskusjonene på overordnet nivå; i motsetningen mellom naturvitenskapelig og samfunnsvitenskapelig forskning, eller ved bruk av kvantitative versus kvalitative metoder. Den teoribaserte evalueringens styrke er knyttet til etableringen av en åpen og transparent programteori, som anerkjenner og synliggjør det faglige utgangspunktet og antakelser som ligger til grunn for evalueringen. Dette bidrar til å gi forskningen større overføringsverdi, sammenlignet med eksperimentelle evalueringer hvor den subjektive faktoren er skjult bak valg av måleindikatorer og studieobjekt (black box) – og dermed kan gi et urettmessig inntrykk av fullstendig objektivitet. Den teoribaserte evalueringsmetodikken

stiller for øvrig samme krav til god forskningspraksis som øvrig naturvitenskapelig og samfunnsfaglig forskning. Det vitenskapelige idealet er sterkt knyttet til reproduerbare resultater, gjentakende studier og universell lovmessighet i vitenskapelige forklaringer. Fortolkningen av disse prinsippene blir imidlertid utfordret når komplekse ITS-tiltak gjør sitt inntog i transportsystemet. Man må erkjenne og akseptere at det er få vedvarende sannheter innenfor dagens transportfelt. Tiltakene endrer seg. Omgivelsene endrer seg. Folk endrer seg. Politikken endrer seg. Markedskreftene endrer seg. Forskere og beslutningstakere må også være villig til å innta nye perspektiv. Med et endret verdensbilde, rask utvikling og disruptiv innovasjon, trengs evalueringsprosesser som gjør oss i stand til å ta gode strategiske valg, også når det empiriske beslutningsgrunnlaget er usikkert. Vi må godta at vi ikke sitter på alle svar, men vi må forstå hvordan ulike mekanismer påvirker transportsystemet og lære av gode eksempler (beste praksis). Den teoribaserte evalueringsmetodikken har således vist seg å være et nyttig verktøy for å frembringe relevant og anvendbar kunnskap for den som skal være med å utvikle fremtidens transportsystem.

10.2 HAR FORSKNINGSPØRSMÅLET BLITT BESVART?

Doktorgradsarbeidet har vært en lærerik prosess som har avstedkommet tre hovedtyper resultater:

1. Økt metodisk kompetanse i form av kunnskap om evalueringsteorier og metodetradisjoner i naturvitenskap og samfunnsforskning. I særdeleshet omfatter dette ny erfaring med bruk av teoribaserte evalueringsprosesser.
2. Bedre innsikt i ITS-tiltak som virkemiddel i transportsystemet, med økt forståelse for underliggende mekanismer og relevant kunnskapsproduksjon.
3. Et konkret forslag til fremgangsmåte for evaluering av ITS-tiltak.

Hovedmålet med doktorgradsarbeidet har vært å øke kunnskapen om det metodiske fundamentet for ITS-evalueringer, med utgangspunkt i følgende forskningsspørsmål:

Kan man med utgangspunkt i generell evalueringsteori utvikle eller tilpasse en evalueringsmetodikk som bidrar til å styrke kunnskapen og beslutningsgrunnlaget for implementering av ITS i transportsystemet?

Hovedresultatet fra dette doktorgradsarbeidet er et analytisk rammeverk for teoribasert evaluering av ITS-tiltak. Rammeverket gir et vitenskapelig grunnlag for hensiktsmessige metodevalg og analyser, og legger til rette for relevant kunnskapsbygging gjennom hele beslutningsprosessen, fra tidlig konseptutvikling til tiltaket er implementert i ordinær drift.

Innledende valg av tilnærming og videre forslag til utforming av evalueringsprosessen er konkretisert gjennom studier av følgende problemstillinger:

- a) Hvilken metodikk legges til grunn for evalueringer av ITS i dag? Og hvordan er denne metodikken forankret i generell evalueringsteori?*
- b) I hvilken grad er dagens evalueringstilnærming i stand til å oppfylle kunnskapsbehovet knyttet til implementering av ITS-tiltak?*
- c) Finnes det andre innfallsvinkler som kan øke nytten av ITS-evalueringer? Hvordan kan dette eventuelt bidra til å styrke kunnskapen om og beslutningsgrunnlaget for implementering av ITS i transportsystemet?*

Dagens evalueringer i transportsektoren er i stor grad styrt av en metodisk tilnærming, med den økonomiske nyttekostnadsmodellen som verktøy for samfunnsøkonomiske analyser, og den (kvasi-)eksperimentelle modellen som gullstandard for effektevalueringer. Transportfaget preges således av en positivistisk, naturvitenskapelig forskningstradisjon. Selv om det finnes eksempler på egne rammeverk for evaluering av ITS-tiltak (f.eks. FESTA-metodikken og EasyWay guidelines), er disse bygd opp rundt det samme eksperimentelle vitenskapsidealet med kontrafaktiske analyser av observerbare effekter og ekskludering av eksterne faktorer som påvirker implementeringen av tiltaket.

De viktigste begrensninger i dagens metodestyrt evalueringer er at metodikken legger til grunn at studieobjektet er a) et teknologisk modent tiltak, b) i stabil drift, c) upåvirket av eksterne faktorer og d) at det foreligger solid empiri om tiltakets årsaksvirkningskjeder. Når disse betingelsene er oppfylt, er den eksperimentelle modellen en vitenskapelig sterk og svært anerkjent metode for å stadfeste i hvilken grad tiltaket har medført målbare effekter på en gitt variabel. ITS-tiltak har imidlertid en rekke egenskaper som gjør at betingelsene for bruk av det klassiske eksperimentet sjelden er oppfylt. Mens tradisjonelle veg- og trafikktiltak prioriteres og implementeres med utgangspunkt i empiri etablert i løpet av lang tids bruk, finnes det ikke tilsvarende erfaring eller historisk kunnskap om nytte- og kostnadsbildet for ITS-tiltakene. Videre vanskeliggjøres etablering av empiri ved at ITS-tiltakene er i kontinuerlig endring etter hvert som teknologien utvikles og modnes. Dette gjør at det er behov for å bygge kunnskap om hvordan ulike ITS-tiltak fungerer, parallelt med utprøving og pilotering i transportsystemet. Til tross for liten evne til å håndtere komplekse ITS-tiltak og kompliserte årsaksvirkningskjeder, er den tradisjonelle metodestyrt tilnærmingen nærmest enerådende i fagfeltet. Ofte forsømmes det faktum at premissene for bruk av metodikken ikke er oppfylt, og en del ITS-evalueringer gjennomføres med en tillemping av datainnsamling og analyser

som til slutt gjør det vanskelig å trekke sikre konklusjoner. Dette gir lite anvendbar kunnskap både for den som skal ta beslutninger og for den som skal implementere ITS-tiltaket i transportsystemet.

Disse betraktningene har gitt grunnlag for å undersøke om det finnes andre metodiske innfallsvinkler, som kan øke nytten av ITS-evalueringer i transportfaget. Det *teoribaserte evalueringsperspektivet* representerer et vitenskapelig alternativ som skiller seg fra dagens metodestyrt tilnærming på to vesentlige områder: Behandlingen av kausale årsakssammenhenger og inkludering av kontekst (rammebetingelser, suksesskriterier og barrierer) som en relevant faktor i evalueringen. Det er gjort rede for det teoretiske grunnlaget for disse forholdene i kapittel 5, men i praksis innebærer dette at den teoribaserte evalueringprosessen har stor forklaringskraft med hensyn til å vurdere hvilken kombinasjon av innsats og virkemidler som gir ønsket effekt. Dette er relevant kunnskap både for den som skal ta strategiske beslutninger om utvikling av fremtidens transportsystem, og for den som skal implementere løsningene i praksis.

Selv om den teoribaserte fremgangsmåten ikke løser alle metodiske utfordringer knyttet til effektevaluering av ITS, legger den til rette for en smidig evalueringssprosess, hvor metodiske svakheter synliggjøres og kan håndteres på en konstruktiv måte i evalueringssprosessen. Det teoribaserte metodegrunnlaget har sin styrke i evalueringer med læring og kunnskapsbygging som formål, og kan bidra til økt forståelse også for komplekse tiltak med mange involverte aktører. Videre har den teoribaserte fremgangsmåten vist seg å være fleksibel og anvendbar i en tidlig fase av innføringsløpet, også når tiltaket har lav teknologisk modenhet, og selv om det foreligger lite empiri om årsakssammenhenger. Dette øker evalueringens formative kapasitet gjennom etablering av verdifull kunnskap og mulighet for effektiv ressursstyring allerede før tiltaket implementeres i stor skala. Teoribaserte evalueringer synes dermed å være særlig egnet for kunnskapsproduksjon i et anvendt perspektiv. Det foreliggende doktorgradsarbeidet har således vist at en dreining fra dagens metodestyrt tilnærming til bruk av en teoribasert evalueringssprosess, kan bidra til å styrke kunnskap om og beslutningsgrunnlaget for implementering av ITS i transportsystemet.

Doktorgradsarbeidet har således bidratt med original kunnskap innenfor fagfeltet. Avhandlingen gir en grunnleggende og samlet oversikt over relevante evalueringsperspektiver for kunnskapsbygging innenfor transportområdet, med innsikt i styrker og svakheter ved et utvalg fremgangsmåter for evaluering av ITS-tiltak. Arbeidet har a) synliggjort utfordringer med bruk av eksperimentell metodikk til evaluering av ITS, b) drøftet styrker og begrensninger med bruk av teoribaserte modeller, c) utforsket og vist hvordan en teoribasert tilnærming kan benyttes til evaluering av ITS i praksis, og d) gitt en systematisert og generisk fremgangsmåte for utvikling av en teoribasert evalueringsmetodikk for ITS-tiltak. Erfaringer fra dette arbeidet har gitt grunnlag for å anbefale

bruk av en teoribasert tilnærming for evaluering av ITS. Avhandlingen presenterer en oppskrift på organisering og gjennomføring av en teoribasert evalueringsprosess, i form av et rammeverk for utvikling av en helhetlig, fleksibel og anvendbar evalueringsmetodikk. Resultatet av dette arbeidet endrer ikke den grunnleggende evalueringsteorien, men kombinerer eksisterende verktøy på en ny måte som kan gi nytteeffekter både for fagfeltet og transportmyndigheter. Innenfor transportsektoren finnes det knapt eksempler på eksplisitt bruk av teoribasert metodikk i evalueringsstudier, og en utvidelse av det metodiske grunnlaget som anvendes vil bidra til å skape et mer helhetlig kunnskapsgrunnlag og en bedre utnyttelse av evalueringsressurser. Bruk av rammeverket gir transportmyndighetene større mulighet til å vurdere teknologi på et lavere modenhetsnivå, samt bedre innsikt i mekanismene som styrer hvorvidt man lykkes eller ikke med iverksettelsen av nye ITS-tiltak. Tidlig involvering og aktiv tilrettelegging av rammebetingelser kan bidra til å fremskynde innføringen av bærekraftige og fremtidsrettede tjenester. Samtidig kan man unngå tapsinvesteringer dersom feilslått teknologi velges bort på et tidligere tidspunkt i innovasjonsløpet.

10.3 METODISKE ERFARINGER OG BEGRENSNINGER

I arbeidet med utviklingen av et teoribasert rammeverk for evaluering av ITS-tiltak, er det benyttet en kombinasjon av *litteraturstudier* som har gitt kunnskap om tidligere forskning og teoretisk grunnlag for fagområdet, og *empiriske case-studier* som har gitt praktisk innsikt i problemområdet.

Litteraturstudiene har omfattet to forskjellige fagfelt: For det første er det utarbeidet en oversikt over kunnskap og rådende perspektiver innenfor generell evalueringsteori. Evalueringsforskning er et etablert fagfelt med pågående vitenskapelige debatter og en tradisjonell akademisk kunnskapsproduksjon, hvor relevant litteratur finnes tilgjengelig i vitenskapelige publiseringskanaler som tidsskrift, serier og fagbokforlag. I tillegg er det gjennomført en litteraturstudie knyttet til intelligente transportsystemer generelt, og evaluering av ITS-tiltak spesielt. Dette fagområdet er forspent med helt andre referanser, hvor forskningsfronten i liten grad finnes tilgjengelig innenfor academia. Dette skyldes blant annet at prosessen med fagfellebedømming og publisering av vitenskapelige artikler er beheftet med en treghet som gjør at den tradisjonelle akademiske kunnskapsproduksjonen ikke holder tritt med den raske utviklingen av fagfeltet. Dertil er det også grunn til å tro at det faktisk er en relativt ny og teoretisk umoden vitenskap, med større vekt på problemløsning og praktisk anvendelse av forskningsresultater, også har en avdempende effekt på det akademiske publiseringsarbeidet. Dette har medført at den aktuelle kunnskapen og resultatene på fagfeltet i større grad finnes på konferanser, i prosjektrapporter og myndighetsdokumenter. Konsekvensene av dette er at litteraturen som foreligger i mindre grad har

gjennomgått de samme kvalitetsvurderingene som gjelder innenfor vitenskapelig kunnskapsproduksjon. Det er også en fare for skjevhet i materialet, ved at det som frontes fra myndigheter og prosjektgrupper stort sett er de vellykkede tiltakene og implementeringene – slik at man kan sitte igjen med en glansbildeversjon av virkeligheten.

Case-studiet er en empirisk undersøkelse som studerer et aktuelt fenomen i dets virkelige kontekst, med hensikt å samle inn så mye data som mulig for å utforske, beskrive og forklare et tema (Yin, 2009). Metoden er egnet til å besvare forskningsspørsmål av typen *hvordan* og *hvorfor*, og benyttes for å kaste lys over en problemstilling som kan være uklar og uten tydelig definerte hypoteser. I så måte har case-studiet vært et anvendelig forskningsdesign for dette arbeidet, som har hatt en klar eksplorativ tilnærming uten særlig anledning til å stille opp konkrete testbare hypoteser i forkant. Case-studiets styrker er knyttet til muligheten for å basere seg på et bredt utvalg datakilder og evnen til å belyse komplekse problemstillinger gjennom både kvalitative og kvantitative analyser. Det har gitt mulighet for å gå i dybden av problemstillingen og komme frem til inngående beskrivelser av temaet som studeres. Case-studiet gir derimot ikke grunnlag for statistisk generalisering, og anses dermed mindre pålitelig i et evidenshierarki (som vist i kapittel 3.4.5). Selv om et case-studium viser at den teoribaserte evalueringsmetodikken virker for ett ITS-tiltak, er ikke det nødvendigvis representativt for alle andre ITS-evalueringer. Resultatene fra dette arbeidet er imidlertid noe styrket av at evalueringsmetodikken er anvendt på to vidt forskjellige ITS-tiltak med ulik teknologi, kontekst og målgrupper (to case-studier med samsvarende resultater). Det må også anses å være til støtte for resultatene at evalueringsmetodikken er videre anvendt i andre prosjekter i ettertid (se kapittel 9.2.1 og 9.2.2).

Jeg har i dette arbeidet konsentrert meg om den overordnede evalueringstilnærmingen som legges til grunn for effektevaluering av ITS. Jeg har valgt å gjennomføre to separate case-studier, for å gi en helhetlig og variert innsikt i problemområdet. Jeg har deltatt aktivt i begge primærstudiene, og tilegnet meg data og erfaring gjennom praktisk deltakelse. Som metode, krever case-studiet mer tid og innsats fra forskeren, enn om arbeidet skulle vært gjennomført som en ren teoretisk øvelse. Til gjengjeld får man dypere innsikt og forståelse for et komplekst problemområde. Det er i dette arbeidet ikke gjennomført komparative studier, utover noen enkle betraktninger av hvordan en tradisjonell evalueringsmetodikk ville påvirket kunnskapsproduksjonen i evalueringsstudiene. Det er heller ikke gått i dybden på problemstillinger knyttet til spesifikke analyseteknikker, som f.eks. hvordan man tester de ulike årsaksvirkningskjedene i programteorien, eller hvordan man prissetter typiske ITS-effekter i nyttekostnadsanalysen. Det er likevel min oppfatning at det har vært fruktbart å starte med å etablere en felles tilnærming for ITS-evalueringer, som har kapasitet til å produsere

nyttig og anvendbar kunnskap for utviklingen av fagområdet. Innenfor dette rammeverket kan man så avstemme og justere metodevalg og teknikker etter hvert som man oppnår mer erfaringer og større forståelse for de underliggende mekanismene som påvirker hvorvidt man lykkes med innføringen av ITS-tiltak i et lengre tidsperspektiv.

10.4 VIDERE FORSKNING OG UTVIKLING

Evaluering av ITS-tiltak er fortsatt et fagfelt med mange ubesvarte spørsmål og interessante problemstillinger. Som tidligere beskrevet, har dette arbeidet fokusert på utviklingen av et strategisk rammeverk for selve evalueringsprosessen, mens det fortsatt er opp til den enkelte forsker å etablere tiltakets programteori og gjøre fornuftige metodevalg underveis. Det er et klart behov for videreutvikling av metodegrunnet som benyttes for å teste (og bekrefte) sammenhengene i ITS-tiltakenes programteori. Det er også etterspørsel etter forskning som bidrar til økt forståelse av hvordan ulike ITS-tiltak fungerer under ulike rammebetingelser; empiri som gjør det lettere å sette opp riktige programteorier. Etter hvert som forskningsfeltet utvikler seg, vil også omfanget av tilgjengelige evalueringsresultater øke. Mangfoldet av ITS-tiltak og installasjoner kan imidlertid gjøre det krevende å sammenfatte et relevant og riktig kunnskapsgrunnlag. Et anvendbart og nyansert evidenshierarki må inkludere flere forskningsspørsmål og ta hensyn til kunnskapens relevans for aktuelle implementeringer. Forskning som bidrar til utvikling av standarder for kvalitetsbedømming av slike forskningsresultater vil styrke fagområdet. Til sist har erfaringene fra dette arbeidet synliggjort et behov for å se nærmere på metoder for scenariebygging og fremtidsforskning. Hvis man tar utgangspunkt i programteorien, har forskningen hittil i stor grad vært fokusert rundt mekanismer og sammenhenger mellom innføringen av tiltaket og effekter som inntreffer på relativt kort sikt (starten av programteorien). Når man beveger seg fra de umiddelbare effektene og videre til samfunnsmessige virkninger i et lengre perspektiv, inntreffer en rekke mekanismer som er vanskeligere å predikere eller studere. ITS-tiltakenes potensial i et fremtidig transportsystem er avhengig av komplekse strukturer knyttet til blant annet politikk, regulering og samspillet mellom de som tilbyr og de som bruker tjenestene. Det er behov for metoder som kan illustrere hvordan kombinasjonen av slike mekanismer vil kunne slå ut i ulike framtidsscenarioer. Foreløpig er slike metoder lite beskrevet innenfor transportsektoren.

REFERANSER

- Aakre, E., Engen, T., & Roche-Cerasi, I. (2014). Targeted selection of overweight vehicles in Norway. Paper presented at the Transport Research Arena 2014, Paris.
- Agatz, N., Erera, A., Savelsbergh, M., & Wang, X. (2012). Optimization for dynamic ride-sharing: A review. *European Journal of Operational Research*, 223(2), 295-303.
- Ajzen, I. (1991). The Theory of Planned Behavior. *Organizational Behavior and Human Decision Processes*, 50, 179-211.
- Alkins, M. C. (2013). *Evaluation Roots. A wider perspective of theorists' views and influences.* 2 ed. Los Angeles: Sage.
- Alvsvåg, H. (2009). Essay: Kunnskapsbasert praksis er ikke noe nytt. *Sykepleien Forskning*, 4(3), 216-220.
- Amey, A. (2010). Real-time ridesharing : exploring the opportunities and challenges of designing a technology-based rideshare trial for the MIT community. Master in City Planning and Master of Science in Transportation.
- Amey, A., Attanucci, J., & Mishalani, R. (2011). Real-Time Ridesharing: Opportunities and Challenges in Using Mobile Phone Technology to Improve Rideshare Services. *Transportation Research Record*, 2217(1),
- Memorandum of Understanding - Deployment strategy for Cooperative ITS in Europe, Car 2 Car Communication Consortium (2013).
- Amundsen, A. (2011). Tiltakskatalog for transport, miljø og klima. Transportøkonomisk institutt.
- Andreone, L., Brignolo, R., Damiani, S., Sommariva, F., Vivo, G., & Marco, S. (2010). SAFESPOT Final Report - Public version. IST-4-026963-IP (D8.1.1).
- Arbib, J., & Seba, T. (2017). Rethinking Transportation 2020-2030. The Disruption of Transportation and the Collapse of the Internal-Combustion Vehicle and Oil Industries.
- Askim, J., Døving, E., & Johnsen, Å. (2013). Evalueringspraksis i den norske statsforvaltningen 2005-2011. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering: Tradisjoner, praksis, mangfold.* Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Aust, M. L. (2012). Improving the Evaluation Process for Active Safety Functions. Doktorsavhandling, Chalmers tekniska högskola, 2012. Serie 3344.
- Baklien, B. (1993). Evalueringsforskningens historie og utvikling i Norge. *Politica. Tidsskrift for politisk videnskap* 25, 56-63.
- Baklien, B., & Skatvedt, A. (2013). Forventning og forvirring - følgforskningens rotete roller. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering. Tradisjoner, praksis, mangfold.* Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Balla, S. J., Lodge, M., & Page, E. C. (Eds.). (2015). *The Oxford handbook of classics in public policy and administration:* Oxford University Press.
- Barnard, Y., Innamaa, S., Koskinen, S., Gellerman, H., Svanberg, E., & Chen, H. (2016). Methodology for Field Operational Tests of Automated Vehicles. Paper presented at the 6th Transport Research Arena, April 18-21, Warsaw, Poland.
- Belin, M.-Å., Tillgren, P., Vedung, E., Cameron, M., & Tingvall, C. (2010). Speed cameras in Sweden and Victoria, Australia - A case study. *Accident analysis and prevention*, 42(2010), 2165-2170.
- Bengler, K. (2007). Subject testing for evaluation of driver information systems and driver assistance systems - learning effects and methodological solutions. In Cacciabue (Ed.) *Modelling driver behaviour in an autonomous environments: critical issues in driver interactions with intelligent transport systems.*
- Berger, N. P., & Vrangbæk, K. (2011). Kan MTV styrkes ved brug af realistisk evaluering? - Udfordringer, barrierer og muligheder ved brug af realistisk evaluering til vurdering av komplekse behandlingsteknologier. *Tidsskrift for Forskning i Sygdom og Samfund*, 15, 55-76. Danmark.
- Bhaskar, R. (2008). *A realist theory of science.* London: Routledge, Taylor and Francis Group.
- Bjerkan, K. (2017). *NordicWay Evaluation Outcome Report.* Version 3.0.
- Blamey, A., & Mackenzie, M. (2007). Theories of Change and Realistic Evaluation. Peas in a Pod or Apples and Oranges? *Evaluation*, 13(4), 439-455.
- Bly, P., & Lawson, M. V. (2007). CityMobil Towards advanced transport for the urban environment. Summary specification for the Heathrow Pilot Scheme.
- Bredgard, T., Salado-Rasmussen, J., & Sieling-Monass, S. M. (2016). *Virkningsevaluering. Evaluering af offentlig politik og organisationer:* Hans Reitzels Forlag.
- Bristow, A. L., Pearman, A. D., & Shires, J. D. (1997). An assessment of advanced transport telematics evaluation procedures. *Transport Reviews*, 17(3), 177-205.
- Brown, V. J., Balducci, P., Mahadevan, K., Greenberg, A., Murray, D., Zirker, M., & Capecci, S. (2007). *CVISN Business Case (Commercial Vehicle Information Systems and Networks). Final Summary Report.*
- Browne, D., & Ryan, L. (2011). Comparative analysis of evaluation techniques for transport policies. *Environmental Impact Assessment Review*, 31, 226-233.
- Böhm, M., Flechl, B., Aigner, W., & Visser, C. (Eds.). (2016). *ITS Evaluation Policy - Culture and Needs.* London: Institution of Engineering and Technology (IET).
- C-Roads. (2019) *Evaluation and Assessment Plan (Version 1.1).* Vol. June 2019: C-Roads Platform Working Group 3.
- Campbell, D. T., & Stanley, J. C. (1966). *Experimental and Quasi-Experimental Designs for Research (Reprint from Handbook of Research on Teaching):* Rand McNally & Company (1963)/Houghton Mifflin Company (1966).
- Carlin, M. (2015) *Effekter av teknologiske endringer på norsk nærings- og arbeidsliv.* SINTEF Rapport A27222: SINTEF IKT.
- Carsten, O. (2007). From driver models to modelling the driver - what do we really need to know about the driver? In Cacciabue (Ed.), *Modelling driver behaviour in an autonomous environments: critical issues in driver interactions with intelligent transport systems.*
- Chan, N. D., & Shaheen, S. A. (2012). Ridesharing in North America: Past, Present, and Future. *Transport Reviews*, 32(1), 93-112.

- Chen, H. T. (1990). *Theory-Driven Evaluations*: Sage Publications.
- Chen, H. T. (2005). *Practical Program Evaluation. Assessing and Improving Planning, Implementation and Effectiveness*: Sage Publications.
- Chen, H. T. (2012). Theory-driven evaluation: Conceptual framework, application and advancement. In R. Strobl, O. Lobermeier, & W. Heitmeyer (Eds.), *Evaluation von Programmen und Projekten für eine demokratische Kultur*.
- Chen, H. T. (2016). Interfacing theories of program with theories of evaluation for advancing evaluation practice: Reductionism, systems thinking, and pragmatic synthesis. *Evaluation and Program Planning*, 59, 109-118.
- Clough Harbour & Associates. (2012). *Commercial Vehicle Information Exchange Window (CVIEW) Roadside Enforcement/Compliance Project: Comprehensive and Detailed Final Project Report*.
- Copenhagen Institute for Future Studies. (2016). *How to be Resilient in the 21st Century*. Retrieved from [http://cifs.dk/radar-shield-and-sword/](http://cifs.dk/radar-shield-and-sword/radar-shield-and-sword/)
- Correia, G., & Viegas, J. M. (2011). Carpooling and carpool clubs: Clarifying concepts and assessing value enhancement possibilities through a Stated Preference web survey in Lisbon, Portugal. *Transportation Research Part A: Policy and Practice*, Elsevier, 45(2), 81-90.
- COST 323. (2001). *Weighing-in-motion of Axles and Vehicles for Europe (WAVE) General report*.
- Dahler-Larsen, P. (2013). Evaluering i Norge set udenfra. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering: Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Dailey, D., Loseff, D., & Meyers, D. (1999). Seattle smart traveler: Dynamic ridesharing on the World Wide Web. *Transportation Research Part C*, 7(1999), 17-32.
- Davies, H., Nutley, S., & Smith, P. (Eds.). (2000). *What works? Evidence based policy and practice in public services*. Bristol, UK: The Policy Press.
- Davis, F. D. (1989). Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology. *MIS Quarterly*, 13(3), 319-340.
- Davis, M. (2003). Operational benefits of the Longs Creek weigh-in-motion system. Referert i Zhang, L. A., Haas, C., & Tighe, S. L. (2008). *Evaluation of the Economic Feasibility of Weigh-In-Motion in Canada*.
- Deakin, E., Frick, K. T., & Shively, K. M. (2010). Markets for Dynamic Ridesharing? Case of Berkeley, California. *Transportation Research Record*, 2187(1), 131-137.
- Department for Transport. (2015). *The Pathway to Driverless Cars. Summary report and action plan*. London.
- Direction generale des Infrastructures des Transports et de la Mer. (2014). *Test deployment project of cooperative intelligent transport systems (SCOOP@F)*.
- Directive 2010/40/EU on the framework for the deployment of Intelligent Transport Systems in the fields of road transport and for interfaces with other modes of transport, 2010.
- Downes, L. (2009). *The Laws of Disruption: Harnessing the New Forces that Govern Life and Business in the Digital Age*. New York: Basic Books.
- Dziekan, K., Riedel, V., Müller, S., Abraham, M., Kettner, S., & Daubitz, S. (2013). *Evaluation matters. A practioners' guide to sound evaluation for urban mobility measures*. Münster, NewYork, München, Berlin.
- EasyWay EEG. (2012). *EasyWay Handbook on Evaluation Best Practice*. Retrieved from www.easyway-its.eu
- Elvik, R. (2004). To what extent can theory account for the findings of road safety evaluation studies? *Accident analysis and prevention*, 36(5), 841-849.
- Elvik, R. (2008) Making sense of road safety evaluation studies. Developing a quality scoring system. Rapport 984/2008. Oslo: Transportøkonomisk institutt.
- Elvik, R. (2011). Effects of Mobile Phone Use on Accident Risk: Problems of Meta-Analysis When Studies Are Few and Bad. *Transportation Research Record*, 2236(1), 20-26.
- Elvik, R., & Vaa, T. (2004). *The Handbook of Road Safety Measures*. Chapter 5: Assessing the quality of evaluation studies (Vol. 44): Elsevier.
- Engström, J., & Hollnagel, E. (2007). A general conceptual framework for modelling behavioural effects of driver support functions. In Cacciabue (Ed.), *Modelling driver behaviour in an autonomic environments: Critical issues in driver interactions with intelligent transport systems*.
- ERTRAC. (2013). *Roadmap for Cross-Modal Transport Infrastructure Innovation. Towards a Performing Infrastructure*.
- ERTRAC. (2015). *Automated Driving Roadmap. Version 5.0 (21/07/2015)*.
- European Commission. (2011). *Roadmap to a single European transport area - towards a competitive and resource efficient transport system*.
- European Commission. (2016a). *C-ITS Platform. Final report*.
- European Commission. (2016b). *Digital Single Market. Digital Economy and Society database*. Retrieved 26 April 2016, from European Commission <https://ec.europa.eu/digital-single-market/node/66889>
- European Commission. (2016c). *EU road fatalities. Statistics Accident data*. Retrieved from http://ec.europa.eu/transport/road_safety/specialist/statistics/index_en.htm
- European Commission. (2016d). *A masterplan for the deployment of interoperable cooperative intelligent transport systems in the EU. Roadmap (07/04/2016)*. European Commission.
- European Commission. (2016e). *What is the Digital Economy and Society Index? Factsheet 25 Feb 2016*. Brussels: European Commission.
- FHWA. (2007). *Commercial Vehicle Size and Weight Enforcement in Europe*. (FHWA-PL-07-002). Federal Highway Administration.
- FHWA. (2014). *USDOT's Intelligent Transportation Systems (ITS) ITS Strategic Plan 2015-2019*. (FHWA-JPO-14-145). Federal Highway Administration.
- Finansdepartementet. (2005). *Veiledning til gjennomføring av evalueringer*.

- Finansdepartementet. (2016). Den fjerde industrielle revolusjon - muligheter til å bedre ressursutnyttelsen. Paper presented at IKT Norges konferanse 12.04.2016.
- Finne, H., Bull-Berg, H., & Samset, K. F. (2012). Klarering, kontroll og kø på grensen. Evaluering av felles statlig kontrollområde for toll, vegvesen og politi på Svinnesund. Rapport A23491: SINTEF Teknologi og samfunn.
- Finseraas, H., & Kotsadam, A. (2013). Hvorfor identifisere årsakssammenhenger i ikke-eksperimentelle data? En ikke-teknisk introduksjon. Tidsskrift for samfunnsforskning, Universitetsforlaget, 54(3), 371-387.
- Fishbein, M., & Ajzen, I. (2010). Predicting and Changing Behavior: The Reasoned Action Approach.
- Fitzgerald, P. C., Sevillano, E., Obrien, E. J., & Malekjafarian, A. (2017). Bridge weigh-in-motion using a moving force identification algorithm. *Procedia Engineering*, 199, 2955-2960.
- Fitzpatrick, J. L., Sanders, J. R., & Worthen, B. R. (2012). Program evaluation : alternative approaches and practical guidelines (4th ed., International ed.). Harlow: Pearson Education.
- Flechl, B. (2017). Why evaluating intelligent transport systems? AustriaTech.
- Forskningsrådet. (2015). Sluttrapport fra SMARTRANS 2007-2014.
- Foss, T. (2015) Fra ARKTRANS til kravspesifikasjoner for ITS-applikasjoner. Rapport A26423: SINTEF Teknologi og samfunn.
- Foss, T. (2017) Automatisert kjøring på veg. Konsept, terminologi og klassifisering av automatiseringsnivå. Rapport 2017-00264. SINTEF Teknologi og samfunn.
- Foss, T. (2017b). Statens vegvesen ITS-veileder. SINTEF Teknologi og samfunn. Web-løsning for Statens vegvesen.
- Foss, T., Lervåg, L.-E., & Moen, T. (2016). Challenges in commercial vehicles pre-clearance. Paper presented at the 11th ITS European Congress, Glasgow.
- FOT-NET. (2014). The FESTA Handbook version 5. FOT-NET Data, version 28.02.2014.
- FOT-NET. (2016). A common methodology for road automation and FOTs.
- Funnel, S. C., & Rogers, P. J. (2011). Proposeful program theory: Effective use of theories of change and logic models: ISBN: 978-0-470-47857-8.
- Furre, H., & Hørringmo, A. M. J. (2013). Om forholdet mellom bestiller og utfører av evalueringer. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering. Tradisjon, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad og Bjørke AS.
- Furuhata, M., Dessouky, M., Ordóñez, F., Brunet, M.-E., Wang, X., & Koenig, S. (2013). Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57, 28-46.
- Gamble, J. A. A. (2006). A developmental evaluation primer. Canada: The J.W McConnell Family Foundation.
- Genus, A., & Stirling, A. (2018). Collingridge and the dilemma of control: Towards responsible and accountable innovation. *Research Policy*, 47(1), 61-69.
- Gjerstad, B., & Bayer, S. B. (2012) Evaluering av pilotprosjekter for samkjøring i Oslo og Bergen. Stavanger: International Research Institute of Stavanger (IRIS).
- Government of Canada. (2012). Theory-Based Approaches to Evaluation: Concepts and Practices. Treasury Board of Canada Secretariat.
- Government of the Netherlands. (2016). Mobility, public transport and road safety. Self-driving vehicles. Retrieved 15.04.2016 from www.government.nl/topics/mobility-public-transport-and-road-safety/contents/self-driving-vehicles.
- Griffin, G., Wood, N., Stoeltje, G., Simek, C., Jones, N., & Ettelman, B. (2015) Real-Time Ridesharing Technology to Support Differential Tolling by Occupancy. Evaluation of 2014 Pilot and 2015 Extension. Central Texas Regional Mobility Authority.
- Groot, K. M. d. (2005). Project Remove Weigh in Motion Systems: Future Enforcement Strategy. Remove Project WP 3 Final Report.
- Guba, E. G., & Lincoln, Y. S. (1989). Fourth generation evaluation. Newsbury Park, California: Sage Publications.
- Guba, E. G., & Lincoln, Y. S. (2001). Guidelines and checklist for constructivist (a.k.a. fourth generation) evaluation. Retrieved 10.03.2017 from Evaluation Checklist Project website: www.wmich.edu/evalctr/checklists
- Guyatt, G., Cairns, J., Churchill, D., Cook, D., Haynes, B., Hirsh, J., . . . Tugwell, P. (1992). Evidence-Based Medicine: A New Approach to Teaching the Practice of Medicine. *JAMA*, 268(17), 2420-2425.
- Halvorsen, A. (2013). Forskningsbaserte evalueringen og andre evalueringer. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering: Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad og Bjørke AS.
- Hansen, H. F. (1991). Programevaluering: Videnskap eller politik? Belyst ved eksemplet midtvejsvalueringen af Det Biotechniske Forsknings- og Udviklingsprogram. *Politica. Tidsskrift for politisk videnskap*, 23(1), 84.
- Haugen, T., Levy, J. R., Aakre, E., & Tello, M. E. P. (2016). Weigh-in-Motion Equipment – Experiences and Challenges. *Transportation Research Procedia*, 14, 1423-1432.
- Herrala, M. (2007). The value of transport information. VTT. Helsinki.
- Herrera, J. C., & Bayen, A. M. (2010). Incorporation of Lagrangian measurements in freeway traffic estimation. *Transport Research Part B: Methodological*, 44(4), 460-481.
- Hessels, & van Lente. (2008). Re-thinking new knowledge production: A literature review and a research agenda. *Research Policy*, 37, 740-760.
- Hills, D., & Junge, K. (2010). Guidance for transport impact evaluations. Choosing an evaluation approach to achieve better attribution. Tavistock Institute/ UK Department for Transport
- Hjørthol, R., Engebretsen, Ø., & Uteng, T. P. (2014) Den nasjonale reisevaneundersøkelsen 2013/14-nøkkelfrapport. TØI Rapport 1383/2014. Oslo: Transportøkonomisk institutt.
- Hjørland, B. (1996). Forskningsevaluering i videnskapsteoretisk belysning. Biblioteksarbejde, Nr 48.
- Hogan, R. L. (2007). The historical development of program evaluation: Exploring the past and present. *Online Journal of Workforce Education & Development*, 2(4), 1-14.
- Hughes, B. P., Falkmer, T., Anund, A., Black, M. H. (2019). The relevance of U. S. strategic Highway Safety Plans in a future context. *PLoS ONE* 14 (10): e0223646

- Høye, A., Elvik, R., & Sørensen, M. W. J. (2011) Trafikksikkerhetsvirkninger av tiltak. TØI rapport 1157/2013. Oslo: Transportøkonomisk institutt.
- Høye, A., Elvik, R., Sørensen, M. W. J., & Vaa, T. (2012). Trafikksikkerhetshåndboken (Vol. 4. utgave). Oslo: Transportøkonomisk institutt.
- Høye, A., Sørensen, M. W. J., Elvik, E., Akhtar, J., Nævestad, T.-O., & TrulsVaa. (2011) Evaluering av friteksttavler i Trondheim. Oslo: Transportøkonomisk institutt.
- IEEE Spektrum. (2020). White paper: Testing is critical for adoption of autonomous vehicles. Retrieved from <https://www.keysight.com/us/en/assets/7018-06408/white-papers/5992-3478.pdf>.
- Innamaa, S., Koskinen, S., & Kauvo, K. (2017). Evaluation Outcome Report - Finland. NordicWay.
- International Transport Forum. (2017a). Shared Mobility Simulations for Helsinki.
- International Transport Forum. (2017b). Transition to Shared Mobility. How large cities can deliver inclusive transport services.
- ITS Norge. (2011). Spontan samkjøring, rapport til Transnova etter første fase.
- ITS Norge. (2013). Spontan samkjøring, rapport til Transnova etter fase 2.
- Jabeen, S. (2016). Do we really care about unintended outcomes? An analysis of evaluation theory and practice. *Evaluation and Program Planning*, 55, 144-154.
- Jacob, B. (2015). Overweight screening by WIM and French initiative for using of WIM for directweight enforcement. Paper presented at the DNIT Workshop 24.06.2015.
- Jacob, B., & Cottineau, L.-M. (2016). Weigh-in-motion for Direct Enforcement of Overloaded Commercial Vehicles. *Transportation Research Procedia*, 14, 1413-1422.
- Jacob, B., & Feypell-de La Beaumelle, V. (2010). Improving truck safety: Potential of weigh-in-motion technology. *IATSS Research*, 34(1), 9-15.
- Jenssen, G. D. (2010). Behavioural Adaptation to Advanced Driver Support Systems. Steps to Explore Safety Implications. (PhD Thesis), NTNU, Trondheim.
- Johansson, K., Denvall, V., & Vedung, E. (2015). After the NPM Wave. Evidence-Based Practice and the Vanishing Client. *Scandinavian Journal of Public Administration*, 19(2), 69-88.
- Jones, M. (2008). Targeted roadside enforcement using WIM and ANPR International Conference on Heavy Vehicles HVPariS 2008 (pp. 359-368).
- Jordanger, I., Malerud, S., Minken, H., & Strand, A. (2007) Flermålsanalyser i store statlige investeringsprosjekt. Rapport nr 18: Concept-programmet, NTNU.
- Jordbakke, A., Salte, M. L., & Hammer, B. S. (2017) How to embrace the potential of autonomous vehicles? Scenario analysis of how autonomous vehicles may affect mobility in Oslo and Akershus. Analyse & Strategi AS, Multiconsult AS.
- Kaplan, S. A., & Garrett, K. E. (2005). The use of logic models by community-based initiatives. *Evaluation and Program Planning*, 28, 167-172.
- Karim, M. R., Ibrahim, N. I., Saifuzil, A. A., & Yamanaka, H. (2014). Effectiveness of vehicle weight enforcement in a developing country using weigh-in-motion sorting system considering vehicle by-pass and enforcement capability. *IATSS Research*, 37(2), 124-129.
- Karlsen, T. K., & Jentoft, N. (2013). Programteori versus samfunnsvitenskapelig teori. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering: Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Kelly, K. (2009). *What technology wants*. New York, USA: Viking Penguin.
- Kjensli, A. (2016). Transportrevolusjonen, *Kronikk* 15.04.2016. Dagens Næringsliv.
- Kolosz, & Grant-Muller. (2015). Extending cost-benefit analyses for the sustainability impact of inter-urban Intelligent Transport Systems. *Environmental Impact Assessment Review*, 50, 167-177.
- Kompfner, P. (2010). Cooperative Vehicle Infrastructure Systems. Final activity report. FP6-2004-IST-4-027293-IP. Report D.CVIS.1.3, ERTICO.
- KPMG. (2018). *Fremsyn 2050 – Trender innen samferdsel frem mot 2050*.
- KPMG. (2019). *Autonomous Vehicles Readiness Index. Assessing countries' preparedness for autonomous vehicles*.
- Kubisch, A. C., C. Weiss, L. B. Schorr and J. P. Connell. (1995). Introduction. In J. P. Connell, A. C. Kubisch, L. B. Schorr, & C. H. Weiss (Eds.), *New Approaches to Evaluating Community Initiatives (Vol. 1)*. Washington DC: The Aspen Institute.
- Kujala, T. (2010). Capacity, workload and mental contents. Exploring the foundations of driver distraction. University of Jyväskylä.
- Kulmala, R. (2010). Ex-ante assessment of the safety effects of intelligent transport systems. *Accident analysis and prevention*, 42, 1359-1369.
- Kyriakidis, M., de Winter, J. C. F., Stanton, N., Bellet, T., van Arem, B., Brookhuis, K., . . . Happee, R. (2017). A Human Factor Perspective on Automated Driving. *Theoretical Issues in Ergonomics Science*, February 2017.
- Larsen, G. (2006). Why megatrends matter. *Scenario Magazine*, 5/2006. Retrieved 19.04.2016 from <http://cifs.dk/publications/scenario-magazine/2006/fo-52006/futureorientation-52006/why-megatrends-matter/>.
- Larsen, O. (2012) Samfunnsøkonomisk vurdering av reduksjon i tillatt totalvekt for vogntog fra 50 til 40 tonn og utvidet veinett for modulvogntog. Rapport nr 2017: Møreforskning Molde AS.
- Lervåg, L.-E. (2012) Workshop om evaluering av ITS-tiltak, 12. mars 2012. SINTEF Notat. Trondheim: SINTEF Teknologi og samfunn, Transportforskning.
- Lervåg, L.-E. (2016). *NordicWay Evaluation Handbook*. NordicWay.
- Lervåg, L.-E., Foss, T., Moen, T., & Aakre, E. H. (2015) NonStop innovasjonsprosjekt. Effektivisering av kontroll av tunge kjøretøy. SINTEF Rapport A-27254. Trondheim: SINTEF Teknologi og samfunn.
- Lervåg, L.-E., Foss, T., & Wahl, R. (2018). SmartFeeder-seamless, connected and automated feeder and shuttle services. Paper presented at the 25th ITS World Congress.

- Lervåg, L.-E., & Meland, S. (2014). Impact assessment of dynamic ridesharing - A Norwegian case study. Paper presented at the European Transport Conference, Frankfurt.
- Lervåg, L.-E., Moen, T., & Nordtømme, M. (2011). Dynamisk varsling av farlige situasjoner - førerstøtte eller distraksjon? En effektstudie av hendelsesvarsler i bil med bruk av kjøresimulator. SINTEF Rapport A19565: SINTEF Teknologi og samfunn.
- Linden, A & Fenn, J. (2003). Understanding Gartner's Hype Cycles. Strategic Analysis Report 30 May 2003. Gartner R-20-1971.
- Linder, Kircher, Vadeby, & Nygårdshus (2007). Intelligent Transport Systems (ITS) in passenger cars and methods for assessment of traffic safety impact. Swedish National Road and Transport Research Institute VTI.
- Little, A. D. (2018). The Future of Mobility 3.0. Reinventing mobility in the era of disruption and creativity.
- Marchal, B., Van Belle, S., De Brouwere, V., & Witter, S. (2013). Studying complex interventions: reflections from the FEMHealth project on evaluating fee exemption policies in West Africa and Morocco. *BMC Health Services Research*, 13(1), 469.
- Martens, M., & Jenssen, G. D. (2012). Behavioral Adaptation and Acceptance. In Eskandarian (Ed.), *Handbook of Intelligent Vehicles* (Vol. 1): Springer.
- Martinussen, M., Reedtz, C., Eng, H., Neumer, S. P., Patras, J., & Mørch, W. T. (2016) Kriterier og prosedyrer for vurdering og klassifisering av tiltak. *Ungsinn: Vol. 2. Tromsø: Universitetet i Tromsø.*
- McKinsey, & Bloomberg. (2016). An integrated perspective on the future of mobility. McKinsey&Company and Bloomberg New Energy Finance.
- Meland, S. (2012) Arbeidsreisen - reisevaner og utvikling. En undersøkelse knyttet til Statens hus og Teknobyen i Trondheim. Rapport A23625. Trondheim: SINTEF Teknologi og samfunn.
- Meland, S. (2017). Cooperative ITS for mobility in European cities. Final project report. Cimec deliverable 5.3.
- Meland, S., Lervåg, L.-E., & Roche-Cerasi, I. (2015) Evaluering av samkjøring. Erfaringer fra samkjøringsaktiviteter i Bergensområdet. SINTEF Rapport A26695. Trondheim: SINTEF Teknologi og samfunn.
- Meland, S., & Nordtømme, M. E. (2014) Reisevaneundersøkelse for Bergensområdet 2013. Datagrunnlag og hovedresultater. Rapport A23625. Trondheim: SINTEF Teknologi og samfunn.
- Miljødirektoratet. (2016). Klimagassutslipp fra transport. Retrieved 20.04.2016 from www.miljostatus.no
- Mitsakis, E., Grau, J. M. S., Iordanopoulos, P., & Mintsis, E. (Eds.). (2016). *Frameworks and methods for ITS evaluation*. London: Institution of Engineering and Technology.
- Moen, T., Lervåg, L.-E., & Jenssen, G. D. (2009) Studie av brukergrensesnitt for systemer for kjørefeltstøtte. Evaluering av WayPilot HMI-løsning. SINTEF Rapport 13765: SINTEF Teknologi og samfunn.
- Murray, G., & Chase, M. (2012) Ridesharing as a Complement to Transit. A Synthesis of Transit Practice. Transit Cooperative Research Program: Transportation Research Board.
- National Highway Traffic Safety Administration. (2013). U.S. Department of Transport Releases Policy on Automated Vehicle Development. Press Release (30 May 2013).
- Natvig, M., Westerheim, H., Moseng, T. K., & Vennesland, A. (2009) ARKTRANS - The Multimodal ITS framework architecture. Version 6. SINTEF-rapport A12001. SINTEF IKT.
- Newman-Askins, R., Ferreira, L., & Bunker, J. (2003). Intelligent transport systems evaluation: From theory to practice. Paper presented at the 21st ARRB and 11th REAAA Conference, Cairns.
- Nordhoff, S., Kyriakidis, M., van Arem, B., Happee, R. (2019). A multi-level model on automated vehicle acceptance (MAVA): a review-based study. *Theoretical Issues in Ergonomics Science*
- Nordtømme, M. E., Jenssen, G. D., Lervåg, L.-E., Hjelkrem, O. A., & Kummeneje, A.-M. (2014) Adaptiv cruisekontroll (ACC) i Norge. Utbredelse, bruksmønster og konsekvenser for trafikksikkerhet og trafikkavvikling. SINTEF Rapport A 26202. Trondheim: SINTEF Teknologi og samfunn.
- NORPASS. (2012). NORPASS - North American Preclearance and Safety Systems.
- Nova Scotia Transportation and Infrastructure Renewal. (2008). Aulds Cove Weigh-In-Motion Evaluation Report.
- Nowacki, G. (2012). Development and Standardization of Intelligent Transport Systems. *International Journal on Marine Navigation and Safety of Sea Transportation*, 6(3).
- Nowotny, Scott, & Gibbons. (2003). Introduction. "Mode 2" Revisited: The Knew Production of Knowledge. *Minerva*, 41, 179-194.
- NVF. (2012). ITS Terminology. *Nordisk Vegforbund*. Vol. 3/2012: Nordisk Vegforbund, ITS-utvalg.
- O'Sullivan, R. G. (2012). Collaborative Evaluation within a framework of stakeholder-oriented evaluation approaches. *Evaluation and Program Planning*, 35(4), 518-522.
- Odeck, J., & Welde, M. (2010). Economic evaluation of intelligent transportation systems strategies: The case of the Oslo toll cordon. *IET Intelligent Transport Systems*, 4(3), 221-228.
- OECD. (1991). Principles for evaluation of development assistance.
- OECD. (2002). Glossary of key terms in evaluation and results based management.
- OECD. (2015). Road Safety Annual Report 2015.
- OECD. (2016). Decarbonising Transport. Helping decision makers establish pathways to carbon-neutral mobility. In OECD International Transport Forum.
- Osland, O., Leiren, M. D., Hoff, A. H., Tennøy, A., & Strand, S. (2007) Evalueringstradisjoner og kvalitetskriterier i samferdselssektoren. TØI Rapport 929/2007. Oslo: Transportøkonomisk institutt.
- Patton, M. Q. (2006). Developmental evaluation. *Evaluation for the way we work*. *Nonprofit Quarterly*, 13(1).
- Patton, M. Q. (2012). *Essentials of utilization-focused evaluation*. Saint Paul, MN: Sage Publications.
- Patton, M. Q., McKegg, K., & Wehipeihana, N. (Eds.). (2015). *Developmental Evaluation Exemplars. Principles in Practice*. New York: The Guilford Press.
- Paulweber, M., Rom, W., & Affenzeller, J. (2016). Austrian R&D&I Roadmap "Automated Vehicles". ECSEL, Austria.
- Pawson, R. (2002a). Evidence-based Policy: In Search of a Method. *Evaluation*, 8(2), 157-181.
- Pawson, R. (2002b). Evidence-based Policy: The Promise of 'Realist Synthesis'. *Evaluation*, 8(3), 340-358.

- Pawson, R. (2013). *The Science of Evaluation. A Realist Manifesto*. London: Sage Publications.
- Pawson, R., & Tilley, N. (1997). *Realistic Evaluation*. London: Sage Publications.
- Pou, A., Nieto, J. L., Cacheiro, E., Garcia-Linares, F., Alfonso, J., Sánchez, N., . . . Kowalska, A. (2015). European Field Operational Test on Safe, Intelligent and Sustainable Road Operation (FOTsis) - Final Report. Version 3. Grant Agreement no 270447 (D1.4).
- Poulikakos, L. D., Heutschi, K., Arraigada, M., Anderegg, P., & Soltic, P. (2010). Environmental footprint of road freight: Case studies from Switzerland. *Transport Policy*, 17(5), 342-348.
- Ranney, T. A. (1994). Models of driving behaviour: A review of their evolution. *Accident analysis and prevention*, 26(6), 733-750.
- Ravndal, E. (2007) Randomiserte kontrollerte forsøk i klinisk rusmiddelforskning - vitenskapeelig hybrid? , *Utposten. Fagblad for allmenn- og samfunnsmedisin*: Vol. 1/2007.
- Riché, M. (2012). Theory Based Evaluation: A wealth of approaches and an untapped potential. Conference paper. Retrieved 15.05.2016 from https://ec.europa.eu/regional_policy/sources/impact/evaluation/conf_doc/helsinki_mri_2012.pdf
- Ropohl, G. (1999). Philosophy of Socio-Technical Systems. *Techné: Research in Philosophy and Technology*, 4(3), 186-194.
- Rossi, P. H., Freeman, H. E., & Lipsey, M. W. (2003). *Evaluation: A systematic approach* (6th edition). Sage Publications.
- Rossi, P. H., Lipsey, M. W., & Freeman, H. E. (2007). *Evaluation. A systematic approach* (7 ed.): Sage.
- Rothengatter, T. (1988). Risk and the absence of pleasure: a motivational approach to modeling road user behaviour. *Ergonomics*, 31(4), 599-607.
- Rudin-Brown, C. M., Jenkins, R. W., Whitehead, T., & Burns, P. C. (2009). Could ESC (Electric Stability Control) Change the Way We Drive? *Traffic Injury Prevention*, 10(4), 340-347.
- Sackett, D. L., Rosenberg, W. M., Gray, J. A., Haynes, R. B., & Richardson, W. S. (1996). Evidence based medicine: What it is and what it isn't. *British Medical Journal (BMJ)*, 312(7023).
- SAE International. (2014). Taxonomy and definitions for terms related to on-road motor vehicle automated driving systems (SAEJ3060): SAE International.
- SAE International. (2018). Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles (SAEJ3016-201806): On-Road Automated Driving (ORAD) Committee.
- Sager, F., & Andereggen, C. (2012). Dealing with Complex Causality in Realist Synthesis: The Promise of Qualitative Comparative Analysis (QCA) (Vol. 33).
- Samferdselsdepartementet. (2010). Intelligente transportsystemer. Strategi. (N-0547B).
- Samset, K., & Christensen, T. (2013). Evaluering av prosjekter ex ante og ex post - og beslutningsprosessens kompleksitet og betydning. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering. Tradisjoner, praksis og mangfold*. Bergen: Fagbokforlaget Vigmostad og Bjørke.
- Sanderson, I. (2002). Evaluation, policy and evidence-based policy making. *Public Administration*, 80(1), 1-22.
- Saunders, R. P., Evans, M. H., & Joshi, P. (2005). Developing a process-evaluation plan for assessing health promotion program implementation: A how-to guide. *Health Promotion Practice*, 6(2), 134-147.
- Schade, J., & Schlag, B. (2003). Acceptability of urban transport pricing strategies. *Transportation Research Part F: Traffic Psychology and Behaviour*, 6(1), 45-61.
- Schwab, K. (2016). The Fourth Industrial Revolution: World Economic Forum.
- Scriven, M. (1991). *Evaluation Thesaurus* (4 ed.): Sage.
- Shadish, W. R. (1998). Evaluation theory is who we are. *American Journal of Evaluation*, 19(1), 1-19.
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2002). *Experimental and quasi-experimental designs for generalized causal inference*. Boston: Houghton Mifflin.
- Shapin, S. (Ed.) (2008). *Science and the Modern World*. Cambridge: The MIT Press.
- Skryseth, H. (2016). Den digitale folkebevegelsen! Klimarevisjonen, *Kronikk 07 April 2016*. *Aftenposten*.
- Skulberg, H. (2008) Evidens og evidensdebattens betydning for utdanningssystemet. *Temanotat 6/2008*.
- Stam, D., & Alessandrini, A. (2012). CityMobil Towards advanced transport for the urban environment. Final ex-post reports (focus on Rome and small demonstrations) (D5.2.4).
- Stam, D., Alessandrini, A., & Marco, S. (2010). CityMobil Towards advanced transport for the urban environment. Field trial B ex ante evaluation report.
- Stanczyk, D., Geroudet, B., Thiounn, C., & Millot, A. (2008). Pre-selection of overloaded vehicles. Paper presented at the Int. Heavy Vehicle Conference HVParis2008 (HVTT10-ICWIM5).
- Stanczyk, D., & Klein, E. (2012). Heavy Traffic Data Collection and Detection of Overloaded HGV. *Procedia - Social and Behavioral Sciences*, 48, 133-143.
- Standard Norge. (2015). SN/K 175 Intelligente transportsystemer. Draft mandate on Urban ITS: Standard Norge.
- Statens vegvesen. (2009). Spontan samkjøring. Revidert forprosjekt.
- Statens vegvesen. (2011a). ITS på veg. En veileder for innføring av vegbaserte ITS-løsninger. ISBN: 978-82-91228-28-0. Statens vegvesen.
- Statens vegvesen. (2011b). Utekontrollveiledning. Versjon 5.0.
- Statens vegvesen. (2012a). Håndbok V712 Konsekvensanalyser.
- Statens vegvesen. (2012b). ITS-strategi for Statens vegvesen. Revisjon 2012. Vegdirektoratet.
- Statens vegvesen. (2013). ITS-standardisering.
- Statens vegvesen. (2014). Oversiktsplanlegging. Veg- og transportplanlegging etter plan- og bygningsloven.
- Statens vegvesen. (2015). Pressemelding 2015-02-12: Tester biler som rapporterer førerforhold i Oslo.
- Statens vegvesen. (2017). Slik skal E8 bli en intelligent veg. Statens vegvesen nyhetsarkiv (12.02.2017).
- Statens vegvesen. (2018). ITS-strategi for Statens vegvesen 2018-2023. Et vegkart mot fremtidens transportsystem.
- Stephens, J., Carson, J., Hult, D. A., & Bisom, D. (2003). Preservation of Infrastructure by Using Weigh-in-Motion Coordinated Weight Enforcement. *Transportation Research Record*, 1855(1), 143-150.

- Stevens, A. (2000). UK Perspective on Cost-Benefit Assessment of Intelligent Transport Systems.
- Stevens, A. (2004). The application and limitations of cost-benefit assessment (CBA) for intelligent transport systems. *Transportation Economics, Vol 8 (Economic impacts of Intelligent Transportation Systems: Innovations and case studies)*, 91-111.
- Stevens, A. (2016). Towards a methodology for Field Operational Tests (FOTs) for automated vehicles. Paper presented at the 11th ITS European Congress Glasgow.
- Stortinget (2015). Lov om intelligente transportsystemer innenfor vegtransport m.m. (ITS-loven), LOV-2015-12-11-101.
- Stortinget (2017) Lov om utprøving av selvkjørende kjøretøy, 2017-12-15-112.
- Stortinget (2017) Prop. 152L (2016-2017) Lov om utprøving av selvkjørende kjøretøy. Forslag til lovvedtak., (2017b).
- Stortingsmelding 26. (2012-2013). Nasjonal transportplan 2014-2023.
- Stortingsmelding 33. (2016-2017). Nasjonal transportplan 2018-2029.
- Stufflebeam, D. L. (1971). The relevance of the CIPP evaluation model for educational accountability. Paper presented at the Annual meeting of the American Association of School Administrators, Atlantic City, N.J.
- Sverdrup, S. (2002). Evaluering. Faser, design og gjennomføring. (2 ed.): Fagbokforlaget Vigmostad & Bjørke AS.
- Sverdrup, S. (2013). Følgforskning som en nyere tendens i norsk evaluering: Hva er det, og hvordan kan det gjennomføres? In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering. Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad & Bjørke.
- Tempo. (2008). Handbook on evaluation. Best practice. Version 3.
- Thorenfeldt, U. K. & Bertelsen, D. (2014). Cost-benefit Analysis for an ITS-solution based on Change in Consumers' Surplus for the Users. *Transportation Research Precedia Vol 1 (1)*, 101-108.
- Thorenfeldt, U. K., Bertelsen, D. E., & Øvstedal, L. R. (2011). Evaluering av enkelte kollektivtrafikktiltak i Trondheim vinteren 2011. SINTEF-rapport A20460. Trondheim.
- Thune-Larsen, H., Veisten, K., Rødseth, K. L., & Klæboe, R. (2014). Marginale eksterne kostnader ved vegtrafikk - med reviderte ulykkeskostnader TØI-rapport: 1307/2014, Transportøkonomisk institutt.
- Tornes, K. (2013). Om verdier og kriterier i evaluering. In A. Halvorsen, E. L. Madsen, & N. Jentoft (Eds.), *Evaluering. Tradisjoner, praksis, mangfold*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Torvatn, H. (1993). Use of evaluations of Norwegian technology transfer programs. (Doctoral thesis), The Norwegian Institute of Technology (NTH), Trondheim. (1993:20)
- Transportetatene. (2016). Grunnlagsdokument Nasjonal transportplan 2018-2029.
- Tufte, P. A. (2013). Forståelser av kausalitetsbegrepet i samfunnsvitenskapene. *Tidsskrift for samfunnsforskning, Universitetsforlaget, 53(3)*, 341-355.
- U.S. Department of Transportation. (2015). Beyond Traffic: The Smart City Challenge. Connected automation for greatest benefits. Information session #2, Dec 17 2015: U.S. Department of Transport.
- Urban, J. B., Hargraves, M., & Trochim, W. M. (2014). Evolutionary Evaluation: Implication for evaluator, researchers, practitioners, funders and the evidence-based program mandate. *Evaluation and Program Planning, 45*, 127-139.
- Vaa, T. (2006) Intelligente transportsystemer (ITS): En oversikt over effekter på atferd og ulykker. Rapport TØI 845/2006: Transportøkonomisk institutt.
- Vaa, T. (2007). Modelling driver behaviour on basis of emotions and feelings - intelligent transport systems and behavioural adaptations. In Cacciabue (Ed.) *Modelling driver behaviour in an autonomous environments: Critical issues in driver interactions with intelligent transport systems*
- Vaa, T., Penttinen, M., & Spyropoulou, I. (2007). Intelligent transport systems and effects on road traffic accidents: State of the art. *Intelligent Transport Systems, IET (1)*, 81-88.
- van der Knaap, P. (2004). Theory-based Evaluation and Learning: Possibilities and Challenges. *Evaluation, 10(1)*, 16-34.
- Van Der Laan, J. D., Heino, A., & De Waard, D. (1997). A simple procedure for the assessment of acceptance of advanced transport telematics. *Transportation Research Part C: Emerging Technologies, 5(1)*, 1-
- Vareide, K., Kobro, L., & Storm, H. (2013) Programteori for attraktivitet. Sammenendragsrapport. Telemarkforskning.
- Venkatesh, V., & Davis, F. D. (2000). A Theoretical Extension of the Technology Acceptance Model: Four Longitudinal Field Studies. *Management Science, 46(2)*, 186-204.
- Venkatesh, V., Morris, M. G., Davis, G. B. & Davis, F. D. (2003). User Acceptance of Information Technology: Toward a Unified View. *MIS Quarterly 27*: 425-478.
- Venkatesh, V., Thong, J. Y. L., Xu, X. (2012). Consumer Acceptance and Use of Information Technology: Extending the Unified Theory of Acceptance and Use of Technology. *MIS Quarterly 36*: 157-178.
- Venkatesh, V., Thong, J. Y. L., Xu, X. (2016). Unified Theory of Acceptance and Use of Technology: A Synthesis and the Road Ahead. *Journal of the Association for Information Systems 17(5)*, 328-376.
- Vlassenroot, S., Brookhuis, K., Marchau, V., & Witlox, F. (2010). Towards defining a unified concept for the acceptability of Intelligent Transport Systems (ITS): A conceptual analysis based on the case of Intelligent Speed Adaptation (Vol. 13).
- Volden, G. H., & Samset, K. (2017) Statlige investeringsiltak under lupen. Erfaring med evalueringer av de 20 første KS-prosjektene. Concept rapport nr. 52. NTNU.
- Volvo Cars. (2016). Drive Me, the worlds most ambitious and advanced public autonomous driving experiment, starts today [Press release, 09.09.2016]
- Volvo Cars. (2017). Swedish families help Volvo Cars develop autonomous drive cars [Press release, 12.12.2017]
- Vågane, L. (2009) Flere i bil? Status og potensial for endring av bilbelegget i Norge. TØI-rapport 1050/2009. Oslo: Transportøkonomisk institutt.
- Walton, M. (2014). Applying complexity theory: A review to inform evaluation design. *Evaluation and Program Planning, 45*, 119-126.

- Weber, M., Andreescu, L., Cuhls, K., Dragomir, B., Gheorghiu, R., Giesecke, S., . . . Sessa, C. (2018). *Transitions on the Horizon: Perspectives for the European Union's future research and innovation policies*. ISBN: 978-92-79-81266-8. European Commission.
- Weiss, C. (1972). *Evaluation Research: Methods for assessing program effectiveness*. Englewood Cliffs N.J.: Prentice Hall.
- Weiss, C. (1998). *Evaluation: Methods for Studying Programs and Policies* (2 ed.): Prentice Hall.
- Weiss, C. (Ed.) (1995). *Nothing as practical as good theory; Exploring Theory-Based Evaluation for Comprehensive Community Initiatives for Children and Families: The Aspen Institute*.
- WHO. (2018). *Global Status Report on Road Safety 2018*.
- Wood, S. P., Chang, J., Healy, T., & Wood, J. (2012). The potential regulatory challenges of increasingly autonomous motor vehicles. *Santa Clara Law Review*, 52(4).
- Work, D. B., Tossavainen, O. P., Jacobson, Q., & Bayen, A. M. (2009). Lagrangian sensing: traffic estimation with mobile devices. Paper presented at the American Control Conference, St. Louis, MO.
- World Economic Forum. (2015). *Deep shift - Technology tipping points and societal impact*. Survey report, September 2015. World Economic Forum.
- World Economic Forum. (2016). *World Economic Forum Annual Meeting 2016. Mastering the forth industrial revolution*. Davos-Klosters, Sveits.
- Yin, R. K. (2009). *Case study research design and methods* (4th ed). Thousand Oaks: Calif Sage Publications.
- Youker, B. W. (2013). Goal-free evaluation: A potential model for the evaluation of social work programs *Social Work Research*, 37(4).
- Zhang, L. A., Haas, C., & Tighe, S. L. (2008). Evaluation of the Economic Feasibility of Weigh-In-Motion in Canada. Paper presented at the 2008 Annual Conference of the Transportation Association of Canada, Toronto Ontario.
- Zlocki, A., Fahrenkrog, F., & Lu, M. (Eds.). (2016). *History and deployment of ITS (Intelligent Transport Systems) technologies and an outline of the book chapters* (Vol. Transportation Series 7). London: Institution of Engineering and Technology (IET).

Internettreferanser:

Artikkel i Teknisk Ukeblad 23. april 2016: <https://www.tu.no/artikler/denne-selvkjorende-bussen-skal-testes-i-norge-i-ar/346640>

Gartner Hype Cycle 2015:

https://blogs.gartner.com/smarterwithgartner/files/2015/10/EmergingTech_Graphic.png Oppslag 02.11.20.

Gartner Hype Cycle 2019:

https://blogs.gartner.com/smarterwithgartner/files/2019/08/CTMKT_741609_CTMKT_for_Emerging_Tech_Hype_Cycle_LargerText-1-1024x974.png Oppslag 02.11.20.

Washington State Department of Transportation: <http://www.wsdot.wa.gov/CommercialVehicle/CVISN/technology.htm>

Oppslag i juni 2016 og juni 2020.