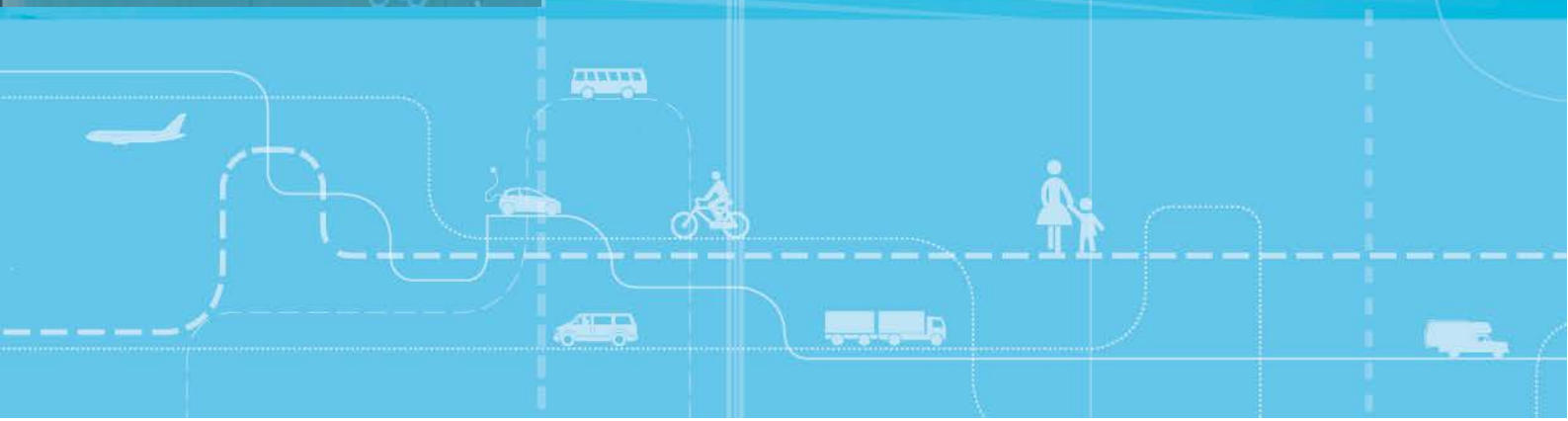


Trafikksikkerhetshåndboken: Bakgrunn om ulykker, risiko og metaanalyse



Trafikksikkerhetshåndboken: Bakgrunn om ulykker, risiko og metaanalyse

Alena Katharina Høye

Rune Elvik

Transportøkonomisk institutt (TØI) har opphavsrett til hele rapporten og dens enkelte deler. Innholdet kan brukes som underlagsmateriale. Når rapporten siteres eller omtales, skal TØI oppgis som kilde med navn og rapportnummer. Rapporten kan ikke endres. Ved eventuell annen bruk må forhåndssamtykke fra TØI innhentes. For øvrig gjelder [åndsverklovens](#) bestemmelser.

Tittel: Trafikksikkerhetshåndboken: Bakgrunn om ulykker, risiko og metaanalyse

Forfattere: Alena Katharina Høye, Rune Elvik

Dato: 11.2019

TØI-rapport 1692/2019

Sider: 71

ISBN elektronisk: 978-82-480-2225-1

ISSN: 0808-1190

Finansieringskilder: Vegdirektoratet
Samferdselsdepartementet

Prosjekt: 1175 –
Trafikksikkerhetshåndboken

Prosjektleder: Alena Katharina Høye

Kvalitetsansvarlig: Rune Elvik

Fagfelt: 21 – Trafikksikkerhet på veg

Emneord: Trafikksikkerhet; Meta-analyse;
Ulykkesrisiko

Title: Handbook of Road Safety Measures: Introductory part – crashes, risk, and meta-analysis

Author: Alena Katharina Høye, Rune Elvik

Date: 11.2019

TØI Report: 1692/2019

Pages: 71

ISBN Electronic: 978-82-480-2225-1

ISSN: 0808-1190

Financed by: Norwegian Public Roads Administration
Ministry of Transport and Communications

Project: 1175 – Handbook of Road Safety Measures

Project Manager: Alena Katharina Høye

Quality Manager: Rune Elvik

Research Area: 21 - Safety and crash countermeasures

Keywords: Road Safety; Meta-analysis; Systematic review; Crash risk

Sammendrag:

Trafikksikkerhetshåndboken gir en oversikt over kunnskap om virkningene av ulike trafikksikkerhetstiltak på antall ulykker. Trafikksikkerhetshåndboken bygger på forskning og undersøkelser om virkninger av trafikksikkerhetstiltak utført over hele verden og gir, så langt det er mulig, tallmessige anslag om virkninger på antall ulykker, drepte og skadde. Denne rapporten inneholder introduksjonsdelen til Trafikksikkerhetshåndboken. En kort versjon finnes i webversjonen av Trafikksikkerhetshåndboken på tsh.toi.no. Rapporten presenterer bakgrunnsinformasjon om trafikksikkerhet og trafikksikkerhetsforskning, generell informasjon om ulykker og risiko i trafikken, både i Norge og internasjonalt, samt om ulykkeskostnader og nyttekostnadsanalyser. Den inneholder også en innføring i metaanalyse med blant annet retningslinjer for systematiske litteratursøk og de viktigste statistiske metodene for log-odds metaanalyse og sensitivitetsanalyser.

Summary:

The Handbook of Road Safety Measures summarizes results from empirical studies that have investigated safety effects of different types of road safety measures. As far as possible, results are summarized by meta-analysis. The present report contains the introductory part of the Handbook. A short version is published in the online version of the Handbook at tsh.toi.no. The report presents some background information, including a summary of vision zero, general information about crashes and crash risk in Norway and internationally and an overview of crash costs and cost-benefit analysis. It also contains an introduction to meta-analysis, including guidelines for systematic reviews and the basics of the most important statistical methods in log odds meta-analysis and analysis of potential biases.

Language of report: Norwegian

Transportøkonomisk Institutt
Gaustadalleen 21, 0349 Oslo
Telefon 22 57 38 00 - www.toi.no

Institute of Transport Economics
Gaustadalleen 21, 0349 Oslo, Norway
Telefon 22 57 38 00 - www.toi.no

Forord

Trafikksikkerhetskåndboken gir en oversikt over aktuell kunnskap om virkninger av 148 ulike trafikksikkerhetstiltak. Den er et uunnvårlig hjelpemiddel for alle som arbeider med trafikksikkerhet, enten det er politisk, administrativt eller i forskning. Hovedfokuset i boken er hvor effektive tiltakene er for å redusere ulykker eller skader ved ulykker. Ulykkesstudiene som boken bygger på, er så langt som mulig oppsummert ved hjelp av metaanalyse.

Arbeidet med Trafikksikkerhetskåndboken ble påbegynt i 1978 på Transportøkonomisk institutt etter initiativ av Trond O. Pedersen som faglig hovedansvarlig for den første utgaven. Arbeidet var en del av «Program for trafikksikkerhetsforskning» som var finansiert av Samferdselsdepartementet. Siden 1980 og fram til i dag er Trafikksikkerhetskåndboken finansiert av Vegdirektoratet og Samferdselsdepartementet. Den første versjonen av Trafikksikkerhetskåndboken ble publisert i 1982. Rune Elvik har siden 1980 bidratt til Trafikksikkerhetskåndboken og han har stått for en stor andel av utviklingen av analysemetoder og presentasjon av tiltakene. Siden 2006 har Alena Katharina Høye vært prosjektleder.

Denne rapporten inneholder en lang versjon av innledningskapitlet i Trafikksikkerhetskåndboken med leseveiledning og bakgrunnsinformasjon om trafikksikkerhet og trafikksikkerhetsforskning, generell informasjon om ulykker og risiko i trafikken, både i Norge og internasjonalt, samt om ulykkeskostnader og nyttekostnadsanalyser. Rapporten inneholder også en innføring i metaanalyse. En kortere versjon av denne rapporten er publisert i online-versjonen av Trafikksikkerhetskåndboken på <https://tsh.toi.no>.

Rapporten er basert på den innledende delen til Trafikksikkerhetskåndboken som er skrevet av Rune Elvik og Truls Vaa i 1997 og oppdatert i 2009 av Elvik et al. (2009A). Den aktuelle rapporten er i hovedsak redigert av Alena Katharina Høye og Rune Elvik. Kapitlene 2 og 5 om ulykker og risiko i vegtrafikken samt ulykkeskostnader og nyttekostnadsanalyse er oppdatert med nyere tall. Kapittel 4 om opplegget for arbeidet med Trafikksikkerhetskåndboken (systematisk litteratursøk og metaanalyse) er i stor grad skrevet nytt i 2019. Siri Hegna Berge og Ole Johansson har kommentert tidligere versjoner av rapporten og Rikke Ingebrigtsen har bidratt til noen statistiske spørsmål. Rapporten er kvalitetssikret av avdelingsleder Trine Dale. Oppdragsgivernes kontaktpersoner har vært Arild Ragnøy og Anne Mette Bjerkan, begge fra Vegdirektoratet.

Oslo, november 2019

Transportøkonomisk institutt

Gunnar Lindberg
Direktor

Trine Dale
Avdelingsleder

Innhold

Sammendrag

Summary

1	Introduksjon	1
1.1	Utvikling av Trafikksikkerhetsboken.....	1
1.1	Trafikksikkerhetsbokens innhold og oppbygging.....	2
1.2	Nullvisjonen for trafikksikkerhet.....	3
1.3	Forskningens rolle i trafikksikkerhetspolitikken.....	4
2	Ulykker og risiko i vegtrafikken	6
2.1	Rapportering og underrapportering av trafikkulykker.....	6
2.2	Klassifikasjon av skadegrader.....	8
2.3	Antall trafikkulykker i Norge og utvikling over tid.....	9
2.4	Risiko i vegtrafikken i Norge sammenlignet med andre land.....	11
2.5	Faktorer som påvirker antall trafikkulykker og skadegraden i ulykker.....	13
2.6	Trafikanter trygghet og utrygghet: Et problem med mange sider.....	15
3	Et faglig grunnsyn	17
3.1	Ulykkesårsaker eller risikofaktorer?.....	17
3.2	Ulykker og risikotall som mål på trafikksikkerhet.....	19
3.3	Trafikkulykker som selvregulerende problem: Teorier om risikolikevekt og atferdstilpasning.....	21
3.4	Tolkning av p-verdier, statistisk signifikans og konfidensintervaller.....	23
4	Opplegg for arbeidet med Trafikksikkerhetsboken	26
4.1	Systematisk litteratursøking.....	26
4.2	Oppsummering av resultater med metaanalyse.....	30
5	Nyttekostnadsanalyse og ulykkeskostnader	57
5.1	Samfunnsøkonomiske analyser og nyttekostnadsanalyser.....	57
5.2	Ulykkeskostnader.....	59
6	Referanser	63

Sammendrag

Trafikksikkerhetshåndboken: Bakgrunn om ulykker, risiko og metaanalyse

TØI rapport 1692/2019

Forfattere: Alena Katharina Høye, Rune Elvik

Oslo 2019 71 sider

Formålet med Trafikksikkerhetshåndboken er å gi en oversikt over kunnskap om virkningene av ulike trafikksikkerhetstiltak på antall ulykker. Trafikksikkerhetshåndboken bygger på empiriske studier av trafikksikkerhetstiltak som er utført over hele verden. Den gir, så langt det er mulig, tallmessige anslag om virkninger på antall ulykker, drepte og skadde. Kvaliteten på de empiriske studiene varierer mellom ulike tiltak, noe boken også gir opplysninger om. I tillegg gis korte oversikter over tiltakenes virkninger på framkommelighet og miljøforhold, samt om kostnader og nytte-kostnadsforhold ved tiltakene. Denne rapporten gir bakgrunnsinformasjon om ulykker og risiko i vegtrafikken, hvordan analysene i Trafikksikkerhetshåndboken er gjennomført, hvordan man kan vurdere hvor pålitelig informasjonen er og hvordan informasjonen kan brukes.

Bakgrunn og leseveiledning for Trafikksikkerhetshåndboken

» Trafikksikkerhetshåndbokens innhold og oppbygging

Denne rapporten inneholder del 1 av Trafikksikkerhetshåndboken med bakgrunnsinformasjon og leseveiledning. Trafikksikkerhetshåndbokens hoveddel (del 2) inneholder de følgende 10 kapitlene:

1. Vegutforming og vegutstyr
2. Drift og vedlikehold av veger
3. Trafikkregulering
4. Kjøretøyteknikk og personlig verneutstyr
5. Kjøretøykontroll og verkstedgodkjenning
6. Krav til førere, føreropplæring og yrkeskjøring
7. Trafikantopplæring og informasjon
8. Kontroll og sanksjoner
9. Førstehjelp og medisinsk behandling
10. Overordnede virkemidler

Hvert kapittel er delt inn i underkapitler som beskriver et trafikksikkerhetstiltak (eller en gruppe tiltak). Hvert tiltakskapittel er delt inn i faste avsnitt som beskrives kort i følgende:

- **Problem og formål:** Det første avsnittet beskriver trafikksikkerhetsproblemet som tiltaket er ment å løse eller hvilke typer ulykker tiltaket kan påvirke. Formålet for tiltakene er ofte bedring av trafikksikkerheten, men tiltak kan også ha andre hovedformål som for eksempel bedring av framkommeligheten.
- **Beskrivelse av tiltaket:** Avsnittet beskriver utformingen og bruken av tiltaket, både i Norge og internasjonalt. Hovedfokuset er på tiltaket eller varianter av tiltaket som er brukt i de empiriske ulykkesstudiene som er oppsummert under «Virkning på ulykkene». For detaljert informasjon om utforming og bruk av tiltak i Norge henvises til Statens vegvesens håndbøker og andre relevante kilder.

- **Virkning på ulykkene:** Dette er det mest sentrale avsnittet. Det beskriver tiltakets virkning på ulykker (ev. skadegraden i ulykker), så langt som mulig som prosentvis endring (med konfidensintervall) og basert på systematiske litteraturstudier og metaanalyser av empiriske studier. Alternativt kan virkningen være estimert basert på en kvalitativ oppsummering av ulykkesstudier eller andre typer studier, for eksempel studier av virkningen på føreratferd.
- **Virkning på framkommelighet og miljø:** Disse avsnittene inneholder korte beskrivelser av hvordan tiltaket påvirker reisetiden eller reisemulighetene, samt miljøeffekter.
- **Kostnader og nytte-kostnadsvurderinger:** Disse avsnittene er avhengige av at det foreligger tilstrekkelig informasjon, både om kostnadene og virkningene på ulykker, samt ev. andre virkninger (for eksempel på reisetid). Slik informasjon er ikke tilgjengelig for alle tiltak. Nytte-kostnadsanalyser fra andre land er i noen tilfeller brukt, men resultatene lar seg ikke alltid uten videre overføre til Norge.
- **Formelt ansvar og saksgang:** Hvem som tar eller kan ta initiativ til et tiltak, hvilke formelle krav og hvilken saksgang som gjelder ved innføring av tiltaket og hvem som har ansvar for gjennomføring av tiltaket. Beskrivelsen viser hovedmønsteret, for detaljer henvises til Statens vegvesens håndbøker og ev. andre relevante kilder.

» Unngå misforståelser

- **Trafikksikkerhetsboken beskriver ikke utelukkende tiltak slik som de brukes i Norge.** De fleste tiltak som er beskrevet i Trafikksikkerhetsboken, finnes i mange ulike varianter og det finnes ofte forskjeller mellom (varianter av) tiltak som er undersøkt i empiriske ulykkesstudier og dem som brukes i Norge. Resultatene fra disse studiene kan likevel være relevante for bruken av tiltak i Norge, og det blir vurdert i hvilken grad resultatene kan overføres til norske forhold.
- **Trafikksikkerhetsboken skal/kan ikke erstatte håndbøker.** I beskrivelsene av tiltakene henvises det ofte til Statens vegvesens håndbøker. Innholdet i disse kan være kort oppsummert, men for detaljer er det henvist til håndbøkene.
- **Man kan ikke forvente at tiltak alltid og overalt vil ha den effekten som er funnet i enkelte empiriske studier eller i metaanalyse.** Bl.a. kan virkningen avhenge av hvilken variant av tiltaket som er undersøkt i de aktuelle empiriske studiene, i hvilken kontekst, hvordan og i hvilket omfang tiltaket brukes, samt gjennomføringen av studiene. Virkninger av tiltak kan også endre seg over tid.
- **Trafikksikkerhetsboken sier ikke noe om hvilke tiltak som «bør» brukes.** Trafikksikkerhetsboken, og trafikksikkerhetsforskning generelt, kan kun gi informasjon om hvilke sikkerhetsmessige effekter man kan forvente av tiltak. Den kan ikke gi noen oppskrift på hvilken trafikksikkerhetspolitikk man skal føre og hvilke tiltak man «bør» satse eller ikke satse på. Bl.a. kan det være andre mål å ta hensyn til, slik som framkommelighet eller personvern.

Ulykker og risiko i vegtrafikken

» Rapporterte (og ikke rapporterte) trafikkulykker

Den viktigste kilden til opplysninger om trafikkulykker med personskade i Norge er **politirapporterte personskadeulykker**. Dette er ulykker hvor minst ett kjøretøy (i hovedsak motorkjøretøy eller sykkel) er innblandet og hvor minst én person er skadd. Fallulykker med fotgjengere er følgelig ikke trafikkulykker etter denne definisjonen og dermed ikke del av politirapportert ulykkesstatistikk.

Ikke alle trafikkulykker med personskade er rapportert i offisiell ulykkesstatistikk. Ulykker som er mest **underrapportert**, er ulykker med mindre alvorlige skader, ulykker med myke trafikanter og eneulykker.

For alle rapporteringspliktige trafikkulykker i Norge sett under ett, er rapporteringsgraden estimert til 17%. Andelen er lavere for lett skadde (15%) enn for hardt skadde (under 37%) og drepte (tilnærmet 100%). For meget alvorlig skadde er andelen ikke rapporterte skader trolig svært liten (11-14%), men dette er veldig usikkert. For ulykker uten motorkjøretøy innblandet med lettere personskade er andelen ikke rapporterte skader tilnærmet null.

Ulykker med kun **materielle skader** er ikke rapporteringspliktige. Det finnes likevel registre over slike ulykker og mange studier som har undersøkt virkningen av trafikksikkerhetstiltak på antall ulykker er basert på slike registre fra ulike land.

» Risiko i trafikken i Norge sammenlignet med andre land

Norge har et lavt risikonivå i trafikken sammenlignet med andre land som har noenlunde samme antall registrerte motorkjøretøy per innbygger som Norge. Denne ledende stillingen har Norge hatt i mange år, også da antallet drepte i trafikken i Norge var mye høyere enn det er i dag.

Blant faktorene som bidrar til den lave ulykkesrisikoen i Norge i forhold til andre land, er at Norge har lave fartsgrenser og en lav promillegrense, og bruken av bilbelter er høyere enn i mange andre land. På den andre siden har Norge få motorveger og en meget varierende standard på det øvrige vegnett. I tillegg har ikke Norge vært et foregangsland når det gjelder sikkerhetskrav til kjøretøy.

På et mer generelt grunnlag kan man si at Norge i relativt stor grad baserer trafikksikkerhetsarbeidet på empirisk forskning, noe som gjør at det er forholdsvis stor sjanse for at det implementeres tiltak som har vist seg å være effektive.

» Faktorer som påvirker antall trafikkulykker og deres alvorlighetsgrad

Antallet skadde personer i trafikken er bestemt av tre generelle faktorer:

- **Trafikkmengde:** Økende trafikkmengde fører som regel til et større antall ulykker. Risikoen for den enkelte trafikant er derimot som regel lavere ved høyere trafikkmengder.
- **Ulykkesrisiko:** Med ulykkesrisiko menes sannsynligheten for å bli innblandet i en ulykke per kilometer man ferdes i trafikken. Faktorer som påvirker ulykkesrisikoen, er bl.a. type kjøretøy, type veg og vegutforming, fysiske miljøfaktorer (som for eksempel lysforhold) og egenskaper ved trafikantene (som for eksempel alder, kjønn, alkoholpåvirkning, trøtthet og mobilbruk).

- **Skaderisiko:** Med skaderisiko menes sannsynligheten for å bli skadet, gitt at man er blitt innblandet i en trafikkulykke. Faktorer som påvirker skaderisikoen i ulykker, er bl.a. fart, type kjøretøy, egenskaper ved trafikantene (især alder), bruk av personlig verneutstyr (bilbelter, hjelm), samt førstehjelp og medisinsk behandling. Skadenes konsekvenser for de skaddes livskvalitet avhenger i tillegg av hvor vellykket behandlingen av skadene er.

» Ulykker og risikotall som mål på trafikksikkerhet

Det finnes flere måter å definere trafikksikkerhet på med utgangspunkt i ulykkes- eller skadetall. De viktigste er beskrevet i det følgende.

Forventet ulykkestall: Det forventede ulykkes- (eller skade-) tall er det gjennomsnittlige antall ulykker (eller skadde) per tidsenhet *i det lange løp* ved uendret trafikkmengde og risiko. Evalueringer av trafikksikkerhetstiltak er ideelt sett basert på endringer i det forventede ulykkestallet. Det registrerte ulykkestallet er ikke nødvendigvis et godt mål på forventet ulykkestall. Små ulykkestall har som regel stor variasjon over tid. Når et tiltak settes inn på en veg som i en periode (for eksempel det siste året) har hatt uvanlig mange ulykker, vil antall ulykker sannsynligvis være lavere året etter, også uten tiltak. Dette er en såkalt regresjonseffekt. Hvis man i en før-etter studie finner en nedgang i det registrerte antall ulykker på for eksempel 40%, er det mulig at det forventede ulykkestallet bare har gått ned med for eksempel 10% (eller ikke i det hele tatt), mens resten av nedgangen skyldes regresjonseffekten. Man vil følgelig ikke kunne forvente en ulykkesnedgang på 40% på andre veger som har et gjennomsnittlig eller lavere antall ulykker. Det finnes statistiske metoder for å kontrollere for regresjonseffekter og resultater fra studier som har brukt slike metoder kan tolkes som effekter på forventede ulykkes-/skadetall.

Ulykkesrisiko: Ulykkesrisiko, eller trafikkenes systemrisiko, er definert som antall ulykker eller skadde per eksponeringsenhet. Eksponeringen måles ofte som million kjøretøy- eller personkilometer. Når ulykkesrisiko benyttes som avhengig variabel i ulykkesstudier, forutsetter man implisitt en lineær sammenheng mellom eksponering og antall ulykker, dvs. at antall ulykker endrer seg lineært med eksponeringen. Denne forutsetningen er imidlertid ikke riktig.

Ved økende antall kjøretøy- eller personkilometer er den prosentvise økningen av antall ulykker eller skadde som regel mindre enn den prosentvise økningen av antall kjøretøy- eller personkilometer. Hvis alt annet er likt, har for eksempel veger med høyere trafikkmengde som regel lavere ulykkesrisiko enn veger med lavere trafikkmengde, syklistene har lavere ulykkesrisiko i områder med mange syklistene enn i områder med få syklistene og førere som kjører mye, har i gjennomsnitt lave ulykkesrisiko enn førere som kjører lite. Dersom man i empiriske ulykkesstudier ikke tar hensyn til slike sammenhenger, kan resultatene være misvisende for tiltak som påvirker trafikkmengden i tillegg til ulykker.

Befolkningens helserisiko: Antall drepte eller skadde per 100.000 innbyggere per år. Helserisikoen som mål på trafikksikkerhet avhenger bl.a. av kjøretøybestanden, hvor mye som kjøres og fordelingen av ulike trafikantgrupper. Helserisiko sier dermed lite om hvor farlig det er å ferdes i trafikken. Land som har lav helserisiko kan ha høy ulykkesrisiko og omvendt.

Kvaliteten på undersøkelser om virkninger av trafikksikkerhetstiltak på ulykkene

Det viktigste formålet med studier av virkninger av trafikksikkerhetstiltak er å finne svaret på spørsmålet om tiltaket fører til et lavere forventet ulykkestall, lavere risiko og/eller lavere skadegrad i ulykker enn man ellers ville ha hatt, samt hvor stor virkningen er, ev. under ulike betingelser. Dette avsnittet beskriver hvilke metodekrav studier bør oppfylle for å kunne besvare disse spørsmålene.

» **Validitet: Metodekrav og grunnlag for årsaksslutninger**

For å gi grunnlag for årsaksslutninger må en undersøkelse gi valide resultater, dvs. at det er god grunn til å tro at de viser den sanne virkningen av et tiltak. Det kan skilles mellom fire former for validitet. Disse vurderes for alle ulykkesstudiene som inngår i analysene i Trafikksikkerhetskåndboken.

- **Statistisk validitet:** Nøyaktighet, feilfrihet og representativitet i resultatene. Resultater fra større undersøkelser (basert på mange ulykker) har som regel mindre tilfeldige feil enn mindre undersøkelser. Systematiske feil kan oppstå bl.a. som følge av underrapportering av ulykker.
- **Teoretisk validitet:** Samsvar mellom det en undersøkelse tar sikte på å måle og det som faktisk måles, samt at resultatet kan forklares ut fra en teori som støttes av resultatet. De fleste trafikksikkerhetsstudier bygger ikke på en eksplisitt teoretisk bakgrunn. Et problem med slike studier er at man ofte kan finne «forklaringer» på de aller fleste resultater, slik at det ikke er mulig å verken bekrefte eller avkrefte antakelser av hvordan et tiltak virker.
- **Intern validitet:** Holdbarheten av slutninger om årsakssammenheng mellom en årsaksfaktor og virkningen av denne. Indikasjoner på intern validitet er bl.a. at det er en statistisk sammenheng mellom tiltak og virkning, at retningen av sammenhengen er entydig (dvs. at man kan forutsette at tiltaket påvirker antall ulykker og ikke omvendt), at virkningen ikke kan forklares med manglende kontroll for forstyrrende variabler og for noen tiltak også at man finner større effekt av større «doser» av tiltaket. Typiske feilkilder i undersøkelser om virkninger av trafikksikkerhetstiltak som reduserer den interne validiteten, er bl.a. regresjonseffekter, endogenitet (når tiltak settes inn der det er mange ulykker; antall ulykker kan da være forhøyet selv med tiltak) og ulykkesmigrasjon (når ulykker skjer på andre steder istedenfor på de stedene hvor tiltak er satt inn).
- **Ekstern validitet:** Resultatenes generaliserbarhet. Resultater har høy ekstern validitet dersom de er stabile i tid og rom og på tvers av ulike undersøkelsesmetoder. Eventuelle forskjeller i resultater mellom undersøkelser bør kunne forklares ut fra kjente egenskaper ved metodene, landene eller forholdene på den tiden undersøkelsen ble utført. Den eksterne validiteten er redusert når disse er spesifikke for den konteksten undersøkelsen utføres i.

» **Hvordan er kravene til gode undersøkelser brukt i Trafikksikkerhetsåndboken?**

For alle undersøkelser som inngår i metaanalysene i Trafikksikkerhetsåndboken, er den metodiske kvaliteten vurdert og det er undersøkt om den metodiske kvaliteten påvirker resultatene. Når dette er tilfelle, er kun de metodisk beste undersøkelsene brukt i metaanalysen. For tiltak hvor det bare foreligger metodisk svake undersøkelser, presenteres de resultater som foreligger, men det gjøres oppmerksom på metodesvakhetene. I tillegg presenteres ofte resultater av studier som har undersøkt tiltakets virkninger på for eksempel føreratferd eller konflikter som indirekte mål på virkningen på ulykker.

Det optimale hadde vært eksperimentelle studier hvor det er et tilfeldig utvalg av enheter som får og ikke får tiltaket som skal evalueres. Slike studier er det som regel ikke mulig å gjennomføre i trafikksikkerhetsforskningen når antall ulykker eller skader skal være den avhengige variabelen. Det finnes likevel mange brukbare undersøkelser, bl.a. før-etterundersøkelser med kontroll for regresjonseffekter, generell ulykkesutvikling og andre forstyrrende variabler eller gode multivariate analyser. Enkle før-etter eller med-uten undersøkelser uten kontroll for forstyrrende variabler anses som utilstrekkelige og er ikke benyttet i metaanalyser.

Opplegg for arbeidet med Trafikksikkerhetsåndboken

» **Litteratursøk**

Trafikksikkerhetsåndboken bygger på en omfattende gjennomgang av norske og utenlandske undersøkelser om virkninger av trafikksikkerhetstiltak på ulykker og skader. Disse undersøkelsene er funnet gjennom en systematisk litteratursøking av i hovedsak bibliografiske databaser (bl.a. ISI Web of Knowledge, TRID), andre databaser (bl.a. ScienceDirect, Taylor & Francis online, TRB, PubMed) og internett (Google Scholar). I tillegg gjennomgås referanselister av relevante publikasjoner og i noen tilfeller gjøres det i tillegg mer åpne internettsøk.

En formell beskrivelse av fremgangsmåten ved litteratursøk for metaanalyser finnes i form at den såkalte PRISMA-sjekklisen (Moher et al., 2009, 2015).

Det er lagt mest vekt på å finne undersøkelser som har tallfestet virkningen av tiltakene på antall ulykker, ulykkesrisiko, antall skader eller skaderisiko. Studier av virkninger på for eksempel konflikter eller trafikantatferd er i hovedsak benyttet når det ikke foreligger studier av virkninger på ulykker eller når slike studier gir uklare eller sprikende resultater.

» **Metaanalyse**

Resultatene av studier av virkninger på ulykker av ulike tiltak er så langt som mulig oppsummert med metaanalyser, dvs. at det er beregnet vektete gjennomsnitt av resultatene fra flere empiriske studier. Når det er gjort metaanalyser er de følgende faktorene undersøkt i analysene.

- **Prosentvis endring av antall ulykker/skader:** Hovedresultatene fra metaanalyse oppgis i Trafikksikkerhetsåndboken som estimert prosentvis endring i antall ulykker, skadde eller drepte.

- **Usikkerhet i virkning:** Den statistiske usikkerheten i virkningen oppgis som 95% konfidensintervall. Et konfidensintervall er intervallet i hvilket man vil finne 95% av resultatene dersom man hadde gjennomført den samme studien under de samme forutsetningene uendelig mange ganger, gitt at den sanne effekten er lik den man har funnet i den aktuelle studien (eller metaanalysen). Større konfidensintervaller betyr større usikkerhet. Usikkerheten er større når det er stor usikkerhet i resultatene fra de enkelte studiene, når det er få studier, eller når det er store forskjeller mellom resultatene fra de enkelte studiene. Hvorvidt et resultat er betydningsfullt, avhenger ikke bare av statistisk signifikans (hvorvidt konfidensintervallet inneholder null), men også av hvor stor effekten er.
- **Metodisk gode vs. dårlige studier:** Hvor gode de enkelte empiriske studiene er påvirker ikke den statistiske usikkerheten i virkningen. Derfor er metaanalysene så langt som mulig basert på de metodiske beste studiene, eller det er systematisk undersøkt hvorvidt metodiske egenskaper påvirker resultatene.
- **Publikasjonsskjevhet:** Empiriske studier som ligger til grunn for metaanalyse, kan representere et skjevt utvalg av alle gjennomførte (eller potensielle) studier. Dette skyldes at studier med uventede, uønskede eller ikke-signifikante resultater ofte ikke blir publisert. Hvorvidt resultatene fra en metaanalyse kan være påvirket av publikasjonsskjevhet (eller andre typer skjevhet) er ved alle metaanalysene systematisk vurdert
- **Regresjonseffekter og type ulykkestall:** Tiltak settes ofte inn på steder med mange ulykker. På slike steder vil antall ulykker ofte gå ned som følge av tilfeldig variasjon, selv om man ikke setter inn noen tiltak. Denne nedgangen er en såkalt regresjonseffekt. I alle metaanalysene er det vurdert hvorvidt resultatene kan være påvirket av regresjonseffekter. Resultater som kan være påvirket av regresjonseffekter, gjelder registrerte ulykkestall. Når resultatene trolig ikke er påvirket av regresjonseffekter, sier man at de gjelder forventede ulykkestall, dvs. det ulykkestallet som man trolig vil observere i det lange løp. Resultater som er påvirket av regresjonseffekter viser ofte langt større effekter enn man kan forvente på steder med et gjennomsnittlig eller lavere antall ulykker.
- **Moderatorvariabler:** Tiltak kan ha ulike virkninger i ulike varianter eller under ulike forhold. Slike varianter eller forhold kalles moderatorvariabler. I de fleste metaanalysene er det undersøkt om det finnes slike moderatorvariabler som påvirker resultatene. Når slike variabler finnes, er virkninger oppgitt for de ulike tiltaksvariantene eller forholdene. Typiske eksempler på moderatorvariabler er skadegrad, ulykkestype og metodiske svakheter ved studiene.

Trafikksikkerheshåndboken presenterer som regel kun en relativt kort oppsummering av de viktigste resultatene fra metaanalyser. Mer detaljerte beskrivelser av analysene finnes ofte, især for de større analysene som er basert på mange undersøkelser, i egne TØI-arbeidsdokumenter, TØI-rapporter og/eller artikler i vitenskapelige tidsskrifter.

Nyttekostnadsanalyse og ulykkeskostnader

Formålet med nyttekostnadsanalyser er å gi opplysninger om hvor kostnadseffektivt eller samfunnsøkonomisk lønnsomt et tiltak kan være. Nytte-kostnadsanalyser er presentert for tiltak hvor det foreligger tilstrekkelig med informasjon, både om kostnadene og virkningene på ulykker, samt ev. andre virkninger (for eksempel på reisetid). Slik informasjon er ikke tilgjengelig for alle tiltak. Nyttekostnadsanalyser som er gjort for norske forhold, er så langt som mulig, gjort i samsvar med retningslinjene beskrevet i Statens vegvesens håndbok V712 (2018).

Nyttekostnadsanalyser er en type samfunnsøkonomisk analyse hvor alle fordeler og ulemper ved et tiltak regnes om til kroneverdier. Et tiltak hvor den samlede nytten (fordelene) er større enn kostnadene (ulempene) er samfunnsøkonomisk lønnsomt. Den samfunnsøkonomiske lønnsomhet kan uttrykkes på ulike måter:

- Netto nåverdi: Differanse mellom nytte og kostnad
- Nytte-kostnadsforhold: Nytten delt på kostnadene; er denne større enn én er tiltaket samfunnsøkonomisk lønnsomt
- Netto-nytte per budsjettkrone: Nettonytte (summen av alle nyttekomponentene minus kostnadene) delt på budsjettkostnadene (Statens vegvesen, Håndbok V712, 2018). Når denne er større enn null er tiltaket samfunnsøkonomisk lønnsomt.

Faktorer som inngår i dagens nyttekostnadsanalyser av trafikksikkerhetstiltak er i hovedsak:

- Trafikant- og transportbruker nytte: Bl.a. distanseavhengige kjørekostnader, tidsbruk, helseeffekter, utrygghet
- Operatør nytte: Kostnader, brukerinntekter og overføringer for bl.a. kollektivselskap, bompengeselskap, ferjeselskap, parkeringsselskap
- Budsjettvirkning: Bl.a. investering, drift og vedlikehold, tilskudd til kollektivtrafikk, skatteinntekter
- Trafikkulykker: Kostnader ved personskade og materiellskadeulykker; den dominerende posten i ulykkeskostnadene er verdsettingen av velferdstap, dvs. trafikantenes teoretiske betalingsvillighet for redusert risiko
- Restverdi
- Skattekostnad
- Støy og luftforurensning.

Kvalitetssikring av Trafikksikkerhetsboken

For å sikre en god faglig kvalitet på arbeidet med Trafikksikkerhetsboken, gjennomføres det en intern korrekturlesning, dvs. at hvert kapittel blir lest og kommentert av minst én forsker på TØI som har ekspertise på det aktuelle fagområdet. I tillegg sendes mange kapitler til relevante personer i bl.a. Vegdirektoratet med fagkunnskap på det aktuelle området. Det er også utarbeidet et stort antall artikler i internasjonale vitenskapelige tidsskrifter med peer-review ordning på grunnlag av Trafikksikkerhetsboken.

Summary

Handbook of Road Safety Measures – Background information about crashes, risk, and meta-analysis

TOI Report 1692/2019

Authors: Alena Katharina Høy, Rune Elvik

Oslo 2019 71 pages Norwegian language

The Handbook of Road Safety Measures summarizes results from empirical studies that have investigated safety effects of different types of road safety measures. As far as possible, results are summarized by meta-analysis. The present report contains the introductory part of the Handbook. A short version is published in the online version of the Handbook at tsh.toi.no. The report presents some background information, including a summary of vision zero, general information about crashes and crash risk in Norway and internationally and an overview of crash costs and cost-benefit analysis. It also contains an introduction to meta-analysis, including guidelines for systematic reviews and the basics of the most important statistical methods in log odds meta-analysis and analysis of potential biases.

Background

» Contents and structure of the Handbook of Road Safety Measures

This report contains background information about the Handbook of Road Safety Measures. The Handbook of Road Safety Measures contains the following ten chapters, each containing subchapters about specific road safety measures:

1. Road design and road equipment
2. Road maintenance
3. Traffic control
4. Vehicle design and protective devices
5. Vehicle and garage inspection
6. Driver training and regulation of professional drivers
7. Public education and information
8. Enforcement and sanctions
9. Post accident care
10. General-purpose policy instruments

Each of the subchapters describing specific road safety measures contains the following sections:

Problem and objective. This section describes the road safety problem which the measure is designed to solve or reduce. A road safety problem can be described in terms of a high number of accidents, a high accident rate or a high proportion of serious injuries. However, not all road safety problems can be described exhaustively in numerical terms only. This applies, for example, to the feeling of insecurity that some road users experience. Some of the measures described in the Handbook have primarily other aims than improving road safety, such as reducing travel times or improving conditions for walking and cycling.

Description of the measure. This section gives information concerning the design of a road safety measure and its intended function. The main focus is on how a measure is used in those evaluation studies that are summarized in the section about crash effects. Detailed technical descriptions are not given.

Effect on accidents. This section deals with the effects on accidents, or on the severity of injury in accidents, which have been found in research. Effects are stated in terms of the percentage change of the number of accidents or injuries, based on systematic reviews and, whenever possible, meta-analysis. Confidence intervals and the most important sources of uncertainty are described as well. For measures where no studies have been found that quantify effects on road safety, qualitative reviews are conducted.

Effect on mobility. In addition to the effect on accidents and injuries, many road safety measures also have effects on mobility. These impacts are briefly described, but not in as great detail as safety effects.

Effects on the environment. Effects on the environment are briefly described. Such effects include traffic noise and air pollution in a wide sense of these terms. Major incursions into the landscape and changes in land use should also be regarded as important environmental effects.

Costs and cost-benefit analysis. For the majority of measures, information is given regarding the cost of the measure in Norway. Cost-benefit analyses are presented whenever available. However, the results of cost-benefit analyses depend strongly on the context to which they refer. Monetary valuations of impacts, which are a key element of cost-benefit analysis, vary substantially between countries. As a rule, one would therefore not expect the results of cost-benefit analyses made in one country to apply directly to another country. The context to which most of the analyses presented refer, is the current situation in Norway.

Crashes and risk in road transport

» Reporting and underreporting of crashes

Most studies of the effects of road safety measures on crashes are based on official crash statistics. Such crash statistics are for the most part quite reliable as regards the most serious crashes. However, certain types of crashes can be highly underreported, especially slight injury crashes, crashes with vulnerable road users, and single vehicle crashes.

In Norway, all crashes involving vehicles (mainly motor vehicles and bicycles) that involve personal injury, have to be reported to the police. It is estimated that 17% of all road traffic crashes actually are reported. The reporting rate is lowest for slight injuries (15%), higher for serious injuries (below 37%) and highest for fatal injuries (close to 100%). Pedestrian accidents without vehicles involved are not included in official crash statistics. Property damage only crashes are not included either. However, there are registers both in Norway and other countries that include such crashes.

» Risk in Norway compared to other countries

Crash risk in road traffic is relatively low in Norway compared to other countries with similar numbers of motor vehicles per inhabitant. Norway has had this leading position for many years, including earlier years when there were far more killed or seriously injured in road traffic than today.

Among the contributing factors to the low crash risk in Norway are low speed limits, strict drunk driving laws, and a relatively high use of seat belts. On the other hand, Norway has few motorways and a highly varying standard on the remaining road network. In Norway decisions concerning road safety are to a relatively large degree based on research. Thus, there are good chances for implemented measures to be effective.

» **Factors affecting crashes and crash severity**

Numbers of crashes and injuries depend on the following three general factors:

- **Traffic volume:** Increasing volumes are usually associated with increases and crash numbers. However, crash numbers (especially injury crashes) usually increase less than proportional with volume. In other words, crash risk for each individual road user is lower on average at higher volumes.
- **Crash risk:** Crash risk is the average number of crashes per travelled kilometer. Factors that affect crash risk include, amongst other things: The of vehicle, road type, road design, environmental factors (such as light and weather conditions), and factors related to the road user (such as age, gender, intoxication, fatigue, and distraction).
- **Injury risk:** Injury risk is the average number of injuries per crash. Factors that affect injury risk include, amongst other things: Speed, type of vehicle, factors related to the road user (especially age), use of protective equipment, and post-accident care.

» **Crashes and risk as measures of road safety**

Road safety can be defined in different ways in terms of crash or injury numbers. The most important approaches are described in the following.

Expected crash numbers: The expected number of crashes (or injuries) is the average number of crashes per unit of time that (theoretically) can be observed in the long run, assuming unchanged volume and risk. Road safety evaluations are ideally based on expected crash numbers. The observed number of crashes per unit of time (for example one year or four years) is not necessarily an appropriate estimator of the expected number of crashes because of random variation. For example, when a road safety measure is implemented at the end of a period with exceptionally high crash numbers (for example one or two years), crash numbers will most likely decrease during the next one or two years, even without any safety measure. This effect is called regression to the mean. There are statistical methods for controlling for regression to the mean.

Crash risk: Crash risk (or injury risk) is defined as the number of crashes (or injuries) per unit of exposure. Exposure is often defined in terms of millions of vehicle or person kilometers. When crash risk is used as the dependent variable in road safety evaluations, one assumes implicitly that the number of crashes changes proportionally with exposure, all else being equal. This assumption is not always correct. Increasing exposure is often related to a reduction in risk. For example, as the amount of cycling increases, risk has been found to decrease for each individual cyclist in a number of studies. This is the so-called safety-in-numbers effect. Motor vehicle crashes also tend to increase less than proportional with exposure. Ignoring such relationships may lead to biased results of road safety evaluations.

Health risk: Health risk as a measure of road safety is defined as the number of killed or injured persons in road traffic per 100.000 population per year. Health risk does not take into account any kind of exposure measure. It does therefore not say anything about how dangerous it is to participate in road traffic. For example, a highly motorized country can have a high health risk, despite low crash risk. Countries with very little road traffic may have low health risk, although those that are participating in road traffic have a very high risk.

The quality of studies of the effects of road safety measures

The primary goal of road safety evaluations is to find out if a measure can be expected to reduce the expected number of crashes or injuries or if it reduces crash or injury risk, compared to a situation without the measure in question. This section describes the requirements empirical studies should meet in order to be able to answer such questions.

» **Validity: Methodological quality and causal inference**

In short, study validity denotes the degree to which the results of the study approximate the truth. The criteria of study quality that have been applied to assess the road safety evaluation studies referred to in this book, are to a great extent based on the validity framework of Cook and Campbell (1979). According to this framework, the quality of a study can be assessed in terms of four types of validity that are described in the following.

Statistical conclusion validity refers to the accurateness and the representativeness of the data and the results of statistical analyses in a study. Study results are statistically valid if they cannot be attributed to randomness or bias of the measurements and if they are representative of a known population of units. The statistical conclusion validity of road safety evaluations can be assessed in terms of the following criteria:

- Sampling technique
- Sample size
- Reporting of statistical uncertainty in results
- Measurement errors
- Specification of accident or injury severity.

Theoretical validity, or construct validity, refers to the theoretical foundation and to the operational definition of theoretical concepts and propositions. A study is theoretically valid if the study measures what it intends to measure and if the results can be explained in terms of theory. Criteria for theoretical validity are:

- Identification of relevant concepts and variables (such as moderator, mediator, and confounding variables)
- Hypotheses describing the relationships between variables
- Knowledge of causal mechanism.

Internal validity refers to the inference of a causal relationship between treatment and effect. This aspect of study validity is very important because the objective of a road safety evaluation study is to determine the *effects* on safety of a road safety measure, not the coincidence of a safety measure and some safety indicator. The following criteria indicate that there is a causal relationship between a safety measure A and a safety indicator B:

- Statistical association between treatment and effect
- A clear direction of causality
- A dose-response pattern (where this is relevant)

- Specificity of the effect (an effect is found in the target group of the measure, not in other groups)
- Control of confounding factors (other safety measures or other changes that coincide in time with the safety measure under evaluation, regression to the mean, self-selection bias, crash migration, study effects)
- Control for endogeneity (if a safety measure is implemented at sites with above average crash numbers)
- Correct specification of the functional form of the relationship between an independent variable (usually a road safety measure) and the estimator of safety effect (relevant in multivariate studies)
- No bias attributable to collinearity or omitted variables (too many, highly interrelated, or too few predictor variables a multivariate model).

External validity refers to the generalizability of study results. A study has high external validity if its findings are valid for different settings than those in which the study was made. It is difficult to assess the external validity of a single study. External validity is best assessed by comparing the findings of studies that have been made in different settings. However, context specific effects of safety measures are not necessarily a methodological weakness, but rather a property of reality. Nevertheless, context specific effects reduce external validity. To some extent, high external validity can make up for the absence of a strong theoretical foundation for road safety evaluation studies. Results that have been replicated a large number of times, in many studies made in many countries, are more likely to show true effects than results reported by just a few studies in just a few countries.

» Assessment of methodological quality in the Handbook of Roads Safety Measures

All studies included in meta-analysis in the Handbook of Road Safety Measures are assessed with respect to methodological quality. Whenever possible, it is also investigated to what degree the results are affected by methodological weaknesses, for example by comparing results between methodologically strong versus weak studies. If methodological quality varies and if it is found to affect the results, only the methodologically best studies are selected for meta-analysis. Simple before-after or simple case-control studies (without any control for confounders) are generally not included in meta-analysis.

If only weak studies are available, the results are presented as they are, but potential effects of methodological weaknesses are discussed. Additionally, results from studies that have investigated effects on indirect measures of safety may be included in the presentation, such as behavior and conflict studies. Ideally, only results from experimental studies would be included in meta-analysis. However, such studies are seldom possible in the field of road safety.

Conducting systematic reviews for the Handbook of Road Safety Measures

» *Literature search*

The Handbook of Road Safety Measures is based on systematic reviews of international studies of the effects of road safety measures on crashes and injuries. The reviews are based on searches of scientific databases (such as ISI Web of Knowledge, TRID), other data bases (such as ScienceDirect, Taylor & Francis online, TRB, PubMed) and the internet (Google Scholar). The reviews are primarily based on the PRISMA check list by Moher et al. (2009, 2015).

The main focus is on studies that have reported quantitative results describing the effects of a safety measure on the number of crashes or injuries, or on crash or injury risk. Studies that have investigated effects on indirect safety indicators, such as conflicts or behavior, are mainly used for safety measures that have not been evaluated in crash studies. They may also be used when results from crash studies are highly inconclusive or contradictory.

» *Meta-analysis*

As far as possible, results from empirical crash studies are summarized by means of meta-analysis in the Handbook of Road Safety Measures. Meta-analysis implies that weighted averages are calculated for a measure of effect that is reported from several studies for a specific road safety measure. The following aspects of the results from individual empirical studies can be investigated by means of meta-analysis:

- ***Percentage change of the number of crashes or injuries:*** The main results from meta-analysis are reported as percentage changes of the number of crashes or injuries. Ideally, they refer to expected crash/injury numbers.
- ***Confidence intervals:*** The statistical uncertainty of the results is reported as 95% confidence intervals. A confidence interval is the interval which in 95 % of the cases would contain the true value of the effect if the same study had been conducted an infinite number of times under the same conditions, provided that the true effect is equal to the one found in the current study (or meta-analysis). Large confidence intervals imply large uncertainty. Large uncertainty in meta-analysis can be due to large uncertainty in the individual studies or large differences between the results of the individual studies. The practical significance of a result depends not only on statistical significance (whether the confidence interval does not “no effect”), but also on the size of the effect and its validity.
- ***Methodologically strong vs. weak studies:*** The statistical uncertainty (the size of the confidence intervals) is directly not affected by the methodological quality of the individual studies. Therefore, meta-analyses in the Handbook of Road Safety Measures are as far as possible based on the methodologically strongest studies. Whenever possible, it is also investigated systematically if the methodological quality of individual studies is related to the results, for example by subgroup-analysis or meta-regression.
- ***Publication bias:*** The individual studies that are included in meta-analysis may be a biased selection of all potential studies that exist, or may exist, for a given road safety measure. The reason is that studies with unexpected, undesired, or non-significant results often remain unpublished. Whenever possible, it is investigated systematically to what degree the results of meta-analysis may be affected by publication bias or other types of bias that may lead to similar effects.

- **Moderator variables:** Road safety measures may have different effects, depending on specific characteristics of the measure or circumstances under which it is implemented. In most meta-analyses in the Handbook of Road Safety Measures it is investigated systematically if there the results may be affected by moderator variables that can be defined on study level. If any such variables are identified, results are reported for the different levels of these variables. Typical study level moderator variables are crash type, crash severity, road type, and road user group (for example by age or gender).

The Handbook of Road Safety Measures presents for the most part only short summaries of the most important results from meta-analysis. More detailed results are described in working documents (unpublished, available on request), TØI-reports, and, as far as possible, in scientific journals.

Cost-benefit analysis

The aim of cost-benefit analysis is to provide information about the cost-effectiveness of a measure or if its benefits exceed the costs. In cost-benefit analysis benefits and costs are assigned monetary values. Cost-effectiveness can be expressed in terms of:

- Net present value: The difference between benefit and costs of the measure; it is positive if benefits are greater than costs
- Benefit-cost ratio: The benefit divided by the cost of the measure; if this ratio is larger than one, the measure is cost-effective.
- Net benefit per unit of public expenditure: The net present value (benefits minus costs) divided by public expenditures. If this ratio is greater than zero, the measure is cost-effective.

Factors included in cost-benefit analyses of road safety measures are mainly:

- Road user benefits: These include changes in travel time, vehicle operating costs, feeling of insecurity and public health
- Transport operator benefits: These are costs and benefits to public transport operators, toll scheme operators, ferry companies and parking companies
- Budgetary impacts: These include investments, maintenance and operation and income from taxes
- Road accidents: Costs of injury accidents and property damage accidents; the dominant item is willingness-to-pay for reduced risk of death or injury
- Opportunity cost of taxation (loss of efficiency)
- Noise and air pollution.

1 Introduksjon

Trafikksikkerhåndboken gir en systematisk oversikt over aktuell kunnskap om virkninger av ulike trafikksikkerhetstiltak. Den bygger på norsk og internasjonal forskning om virkninger av trafikksikkerhetstiltak og gir, så langt det er mulig, tallmessige anslag om virkninger på antall ulykker, drepte og skadde. Kvaliteten på disse opplysningene varierer mellom tiltakene. Boken gir derfor også opplysninger om usikkerheten i virkningene. I tillegg beskriver boken virkninger på framkommelighet og miljøforhold, samt kostnader og nytte-kostnadsforhold ved tiltakene. Trafikksikkerhåndboken henvender seg først og fremst til dem som arbeider med trafikksikkerhet, og skal også være til nytte for alle som ønsker å vite mer om effekten av trafikksikkerhetstiltak.

Denne rapporten gir relevant bakgrunnsinformasjon for Trafikksikkerhåndboken: Ulykker og risiko i vegtrafikken (kapittel 2), det faglige grunnsynet (kapittel 3) og opplegget for arbeidet med Trafikksikkerhåndboken (kapittel 4). Kapittel 4 beskriver bl.a. hvordan analysene er gjennomført, hvordan man kan vurdere hvor pålitelig informasjonen er og hvordan informasjonen i Trafikksikkerhåndboken kan brukes.

1.1 Utvikling av Trafikksikkerhåndboken

Arbeidet med Trafikksikkerhåndboken ble påbegynt i 1978 på Transportøkonomisk institutt etter initiativ av Trond O. Pedersen som faglig hovedansvarlig for den første utgaven. Arbeidet var en del av «Program for trafikksikkerhetsforskning» som var finansiert av Samferdselsdepartementet. Siden 1980 og fram til i dag er Trafikksikkerhåndboken finansiert av Vegdirektoratet og Samferdselsdepartementet.

I de første årene, fra 1978-1984, var Trond O. Pedersen prosjektleder for Trafikksikkerhåndboken, med Karin Berard-Andersen som prosjektmedarbeider. Rune Elvik har vært prosjektmedarbeider siden 1980, prosjektleder i 1984-2006, og han står fram til i dag for utviklingen av analysemetoder og presentasjon av tiltakene. Alena Katharina Høye har vært prosjektleder siden 2006 og bidrar i dag til en stor del av kapittelrevisjonene. Truls Vaa har bidratt til Trafikksikkerhåndboken siden 1988 og har sammen med Rune Elvik skrevet de fleste kapitlene fram til omtrent 2009. Egil Østvik bidro i 1986-1989, og Anne Borger Mysen i 1991-1997. I 2008-2018 har Michael W.J. Sørensen bidratt med en del kapitler, i tillegg til kvalitetssikring. Etter 2006 har også enkelte andre medarbeidere på TØI bidratt til Trafikksikkerhåndboken (i alfabetisk rekkefølge): Astrid H. Amundsen, Per-Andreas Langeland, Rolf Hagman, Ross O. Phillips, Sunniva F. Meyer og Tor-Olav Nævestad.



Figur 1: Trond O. Pedersen, Rune Elvik, Truls Vaa.

I 2011 har, i tillegg til Samferdselsdepartementet og Vegdirektoratet, også det svenske Trafikverket bidratt til finansieringen av arbeidet med Trafikksikkerhetshåndboken. I denne sammenhengen var flere forskere fra VTI involvert i revisjonen av noen kapitler: Anna Anund, Anna Vadeby, Gudrun Øberg, Jonas Jansson og Susanne Gustafsson.

Transportøkonomisk institutt utgav trafikksikkerhetshåndboken for første gang i 1982 (signert av daværende instituttsjef Dag Bjørnland). Den er senere kommet i reviderte utgaver på norsk i 1989 (Elvik et al., 1989), 1997 (Elvik et al., 1997) og i 2012 (Høye et al., 2012). Boken ble utgitt på finsk i 1993 og på russisk i 1997 og 2000. Boken ble utgitt på engelsk i 2004 (Elvik & Vaa, 2004) og 2009 (Elvik et al., 2009A), på spansk i 2006 og 2013 (Elvik et al., 2013) og på portugisisk i 2015 (Elvik et al., 2015). Boken er også tilgjengelig gratis på <http://tsh.toi.no/> hvor reviderte kapitler legges ut fortløpende.

1.1 Trafikksikkerhetshåndbokens innhold og oppbygging

Trafikksikkerhetshåndboken er delt inn i tre deler: Generell del (del 1 – denne rapporten), tiltak og konkrete virkemidler (del 2) og referanse- og oppslagsdel (del 3). Hver del kan leses for seg. Kapitlene i del 2 er også ment å kunne leses selvstendig.

Del 1 beskriver bokens formål og oppbygging, hvordan boken er blitt til, ulykker og risiko i vegtrafikken, et faglig grunnsyn, teorier og metoder i trafikksikkerhetsforskningen og bruk av boken. Dette er innholdet i den foreliggende rapporten.

Del 2 beskriver konkrete virkemidler og tiltak for å bedre trafikksikkerheten. Del 2 består av 10 kapitler, som igjen er delt inn i underkapitler. Hvert underkapittel beskriver et trafikksikkerhetstiltak eller en gruppe av trafikksikkerhetstiltak. De ti kapitlene i del 2 beskriver til sammen 148 trafikksikkerhetstiltak på følgende områder:

1. Vegutforming og vegutstyr
2. Drift og vedlikehold av veger
3. Trafikkregulering
4. Kjøretøyteknikk og personlig verneutstyr
5. Kjøretøykontroll og verkstedgodkjenning
6. Krav til førere, føreropplæring og yrkeskjøring
7. Trafikantopplæring og informasjon
8. Kontroll og sanksjoner
9. Førstehjelp og medisinsk behandling
10. Overordnede virkemidler

Hvert av kapitlene som beskriver et trafikksikkerhetstiltak i del 2 av boken er skrevet etter en felles disposisjon. Tiltakskapitlene er inndelt i faste avsnitt:

- **Problem og formål:** Beskrivelse av trafikksikkerhetsproblem tiltaket er ment å løse eller redusere. Et problem kan for eksempel være et høyt ulykkestall, høy ulykkesrisiko eller alvorlige skader i en trafikantgruppe, ulykkestype, vegtype mv. Formålet med tiltakene er ofte bedring av trafikksikkerheten, men de kan også ha andre hovedformål som for eksempel bedre fremkommelighet.
- **Beskrivelse av tiltaket:** Utforming og bruk av tiltak (ev. varianter av tiltaket), både i Norge og internasjonalt. Hovedfokuset er på tiltaket / varianter av tiltakene som er brukt i de empiriske ulykkestudiene som er oppsummert under «Virkning på ulykkene». For detaljert informasjon om utforming og bruk av tiltak i Norge henvises det til Statens vegvesens håndbøker og andre relevante kilder.
- **Virkning på ulykkene:** Virkningen på ulykker (ev. skadegraden i ulykker) oppgis så langt som mulig som prosentvis endring (med konfidensintervall), basert på empiriske studier. Når det foreligger resultater fra tilstrekkelig mange studier, er disse oppsummert med metaanalyse. Når det ikke foreligger studier av virkninger på ulykker, er virkningen omtalt på annen måte, for eksempel basert på virkninger på førerferd. I tillegg beskrives de viktigste kildene til slik usikkerhet for hvert tiltak.
- **Virkning på framkommelighet:** Hvordan tiltaket påvirker reisetiden eller reisemuligheter.
- **Virkninger på miljøforhold:** Hvilke miljøeffekter som kan forventes av tiltaket, bl.a. støy, luftforurensning, klimaeffekter mv. Også ev. virkninger på bl.a. utrygghet og sosiale forhold kan inngå i dette avsnittet.
- **Kostnader:** Tiltakskostnader kan være basert på opplysninger som er hentet fra offentlige budsjett og regnskap, forskningsrapporter, produsenter eller forhandlere. Gode kostnadstall er ikke alltid lette å finne.
- **Nytte-kostnadsvurderinger:** Hvorvidt tiltaket kan forventes å være samfunnsøkonomisk lønnsomt. Nytte-kostnadsvurderinger forutsetter at det foreligger informasjon om tiltakets nytte (nyttien omfatter alle effekter som har målbare og prissatte konsekvenser) og kostnader. Når slik informasjon ikke foreligger er det i noen tilfeller likevel gjort generelle vurderinger av hvorvidt tiltaket kan forventes å være lønnsomt. Alternativt beregnes det hvor mye tiltak kan koste uten å være ulønnsomt.
- **Formelt ansvar og saksgang:** Hvem som tar eller kan ta initiativ til et tiltak, hvilke formelle krav og hvilken saksgang som gjelder ved innføring av tiltaket og hvem som har ansvar for gjennomføring av tiltaket, herunder hvem som bærer utgiftene til et tiltak. Reglene om formelt ansvar og saksgang kan være innviklede for en del tiltak. Beskrivelsen viser hovedmønsteret. For mer detaljerte beskrivelser henvises til Statens vegvesens håndbøker og andre relevante kilder.

Del 3 inneholder forklaring av ord, symboler og forkortelser som brukes i boken.

1.2 Nullvisjonen for trafikksikkerhet

Nullvisjonen er grunnlaget for trafikksikkerhetsarbeidet i Norge (Samferdselsdepartementet, 2016-2017) og andre land:

«Regjeringens hovedmål for transportsikkerhet er å redusere transportulykkene i tråd med nullvisjonen. I planperioden vil regjeringen øke transportsikkerheten innen vegtrafikk, jernbanetransport, sjotransport og luftfart. Hovedmålet innebærer en visjon om at det ikke skal forekomme ulykker med drepte eller hardt skadde i transportsektoren. Innen 2030 skal antall drepte og hardt skadde i vegtrafikken reduseres til maksimalt 350, mens det høye sikkerhetsnivået innen øvrige transportformer skal opprettholdes og styrkes.» (Samferdselsdepartementet, 2016-2017, s. 14).

Ifølge nullvisjonen skal det ikke forekomme ulykker med drepte eller varig skadde i trafikken. Trafikksikkerhetsarbeidet i Norge og mange andre land tar utgangspunktet i nullvisjonen som først ble lansert i Sverige (Kommunikasjonsdepartementet, 1996; Samferdselsdepartementet, 2016-2017; Statens vegvesen, 2014; Tingvall, 1997; Tingvall & Haworth, 1999; Vägverket, 1996). Nullvisjonen er lansert som et langsiktig ideelt mål for et trafikksystem der kravet til trafikksikkerhet er dimensjonerende.

Nullvisjonen innebærer bl.a. at:

- Nye veger må utformes slik at de leder til sikker atferd og beskytter mot fatale konsekvenser av feilhandlinger
- Det må legges spesiell vekt på å prioritere tiltak rettet mot de alvorligste ulykkene
- Ansvar for trafikkulykker ligger hos dem som planlegger og driver trafikksystemet, mens trafikanter er ansvarlige for å følge reglene for bruk av trafikksystemet.

Det grunnleggende spørsmålet innenfor rammen av nullvisjonen er hvor stor fysiske krefter mennesker kan utsettes for ved ulykker før det oppstår alvorlige skader. Når man har fastslått dette, kan man utlede hvor høy fart som kan tillates og hvordan veger og kjøretøy må utformes for ikke å forårsake større skader ved ulykker enn den som ligger under det nivået som fører til livsvarige skader.

En vanlig innvending mot nullvisjonen er at den er urealistisk. Dette gjelder når man tolker nullvisjonens krav om at liv og helse aldri kan utbyttes mot andre goder (Tingvall, 1997, sitert etter Elvik, 1999) bokstavelig og som kun gjeldende for trafikksystemet. Med en slik tolkning vil det ikke være forsvarlig å bruke penger på annet enn trafikksikkerhet så lenge det forekommer drepte eller hardt skadde i trafikken. Men en mindre streng tolkning som kun gjelder å redde liv (på trafikksikkerhet og andre områder) ville det ikke vært forsvarlig å bruke penger på andre ting enn å redde liv.

Det ville implisere at dødeligheten på andre områder ville øke, trolig langt mer enn dødeligheten i trafikken ville gå ned (Elvik, 2017). Elvik (1999) viser i et hypotetisk regneeksempel at dersom man ville implementere Nullvisjonen fullt ut i Norge, ville man redusere det årlige antall trafikkdrepte med 210, men antall dødsfall som skyldes andre årsaker ville øke med 1355

Når man derimot tolker nullvisjonen som en visjon, kan den stimulere til utvikling av nye tiltak og til å sette større fokus på de mest effektive tiltakene mot de mest alvorlige ulykkene. De formelle planleggings- og prioriteringsmetodene alene stimulerer i liten grad til utvikling av nye tiltak.

1.3 Forskningens rolle i trafikksikkerhetspolitikken

Trafikksikkerhetsforskning er et hjelpemiddel som skal gjøre trafikksikkerhetspolitikken mer effektiv. Med trafikksikkerhetspolitikken menes her de beslutninger myndighetene tar om hvilke mål man skal oppnå, og hvilke tiltak det skal satses på for å bedre trafikksikkerheten.

Forskningens rolle som grunnlag for trafikksikkerhetspolitikken er blant annet å vurdere virkninger av trafikksikkerhetstiltak på antall ulykker og skader, samt å vurdere kunnskapenes holdbarhet ut fra den metodiske kvaliteten av undersøkelsene (ikke ut fra resultatene).

Trafikksikkerhåndboken gir informasjon om effekter av trafikksikkerhetstiltak som er funnet i empiriske studier, men ingen anbefalinger om bruk eller ikke-bruk av bestemte tiltak. Selv om et tiltak er effektivt, er det ikke alltid opplagt at det «bør» brukes i stort omfang, eller at tiltak som ikke reduserer ulykker, ikke «bør» brukes. Eksempelvis kan tiltak:

- Ha et begrenset bruksområde: Et eksempel er envegsregulering av gater. Disse har vist seg å redusere ulykker, men kan åpenbart ikke benyttes i mesteparten av vegnettet
- Være mindre lønnsomme enn alternative tiltak: For eksempel kan nedsettelse av fartsgrenser redusere antall ulykker. Likevel kan andre tiltak være mer lønnsomme, for eksempel flere fartskontroller.
- Ha andre effekter enn trafikksikkerhet som anses som viktigere: Eksempelvis har reversible kjørefelt ofte ingen ulykkesreducerende effekt, men slike kjørefelt kan likevel være avgjørende for å unngå ellers uløselige køproblemer.

I hvilken grad forskningsbasert kunnskap kan bidra til bedre trafikksikkerhet er undersøkt i Sverige (Elvik et al., 2009B). Det er anslått at antall trafikkdrepte i 2005, som var 440, ville ha vært mer enn dobbelt så høyt (921) uten de tiltakene som ble satt inn i perioden 1971-2004 på grunn av resultater fra svensk trafikksikkerhetsforskning.

Hvor går grensen mellom fag og politikk og når kan man si at forskere inntar en politisk rolle de ikke bør ha? Noen vil si at det ikke er noen skarp grense og at forskere i kraft av sin kunnskap må kunne gi råd om hvilke tiltak man bør satse på. Mer problematisk blir det hvis forskere anbefaler tiltak som er i strid med vedtatte politiske mål. En beregning av såkalt «optimale» fartsgrenser i 2017 (Elvik, 2017) gir et eksempel på dette. Optimale fartsgrenser fastsettes slik at summen av alle kostnader ved trafikken – ulykker, tid, kjøretøys driftskostnader, støy, utslipp og global oppvarming – blir lavest mulig. Beregningen viste at de optimale fartsgrenser i Norge er høyere enn dagens fartsgrenser, unntatt for fartsgrensene 80 og 100 km/t. Men hvis man øker fartsgrensene, vil farten øke og antall drepte og skadde vil gå opp. Dette strider mot et vedtatt politisk mål om å redusere antall drepte og hardt skadde. Det er ikke en forskningsoppgave å anbefale andre politiske mål enn dem politikerne har vedtatt. Derimot ligger det innenfor forskningens mandat å gi råd om hvordan de politiske mål best kan nås.

2 Ulykker og risiko i vegtrafikken

2.1 Rapportering og underrapportering av trafikkulykker

De fleste studier som har undersøkt virkninger av tiltak på ulykker/skader er basert på politirapporterte ulykker. Rapporteringspliktige trafikkulykker i Norge er alle ulykker med «betydelig personskade» hvor et kjøretøy er innblandet. Rapporteringsgraden varierer mellom ulike typer ulykker. Det er især mindre alvorlige ulykker og ulykker med fotgjengere og syklister som har en lav rapporteringsgrad. Fallulykker med fotgjengere er ikke rapporteringspliktige i Norge og dermed ikke representert i offisiell ulykkesstatistikk.

Den viktigste kilden til opplysninger om trafikkulykker med personskade i Norge er politirapporterte personskadeulykker. Plikten til å rapportere trafikkulykker med personskade til politiet fremgår av vegtrafikkloven, paragraf 12, tredje ledd: «Har trafikkuhell medført død eller skade på person og skaden ikke er ubetydelig, skal de som er innblandet i uhellet, sørge for at politiet snarest mulig blir underrettet om uhellet». Begrepet trafikkuhell omfatter ulykker der minst ett kjøretøy i bevegelse er innblandet. På grunnlag av vegtrafikklovens bestemmelser om rapporteringsplikt for trafikkulykker med personskade, kan det skilles mellom tre hovedgrupper av ulykker som skjer på offentlig trafikkområde:

- **Rapporteringspliktige trafikkulykker:** Alle ulykker med «betydelig personskade» hvor et kjøretøy er innblandet, er rapporteringspliktige. Betydelig personskade er ikke definert i vegtrafikkloven; et mulig kriterium er om den skadde søker medisinsk behandling. Kjøretøy omfatter motorkjøretøy og sykler (eneulykker med sykkel og kollisjoner mellom sykler eller mellom en sykkel og en fotgjenger, er følgelig rapporteringspliktige).
- **Ikke rapporteringspliktige trafikkulykker:** Ulykker med ubetydelig personskade er ikke rapporteringspliktige, selv om kjøretøy er innblandet.
- **Ikke trafikkulykker:** Ulykker uten kjøretøy innblandet er ikke definert som trafikkulykker; dette er i hovedsak fallulykker med fotgjengere.

Det er en vesentlig forskjell i rapporteringsgrad mellom rapporteringspliktige trafikkulykker der motorkjøretøy er innblandet og rapporteringspliktige trafikkulykker der motorkjøretøy ikke er innblandet (Hvoslef, 1996).

Underrapportering: Det er godt kjent at ikke alle trafikkulykker med personskade er rapportert i offisiell ulykkesstatistikk. Årsakene er bl.a. at lett skadde personer lar være å rapportere ulykken (for eksempel for å unngå bøter eller økte forsikringspremier), eller fordi skadegraden undervurderes eller feilklassifiseres av politiet (Patil et al., 2012).

Blant de viktigste implikasjonene av underrapportering er usikkerhet i de estimerte virkningene av sikkerhetstiltak (Hauer & Hakkert, 1988) og at effekten av tiltak på typer skader med liten rapporteringsgrad kan bli underestimert (Patil et al., 2012). Et konkret eksempel på hvordan rapporteringsgraden påvirker registrerte ulykkestall er de forholdsvis store endringene i antall skadde og drepte i 1977 (økning) og 1978-79 (nedgang) som delvis skyldes endringer i rapporteringsrutinene for trafikkulykker i 1977 og 1978 (Fridstrøm et al., 1993).

Faktorer som påvirker rapporteringsgraden: Rapporteringsgraden for trafikkulykker med personskade påvirkes av flere faktorer:

- **Skadegrad:** Rapporteringsgraden er som regel lavere for mindre alvorlige skader. Metaanalysen til Elvik og Mysen (1999) som er basert på 49 studier fra 13 land viser at rapporteringsgraden i gjennomsnitt er 95% for drepte, 70% for alvorlig skadde og 25% for lett skadde. Borger et al. (1995) viser at rapporteringsgraden er 2,6 ganger så høy for alvorlige skader som for lettere skader.
- **Trafikantgrupper og ulykketype:** Rapporteringsgraden er som regel lavest for myke trafikanter. Metaanalysen til Elvik og Mysen (1999) viste at rapporteringsgraden i gjennomsnitt er 61% for motorkjøretøy (uten motorsykler), 46% for motorsykler, 22% for sykler og 58% for fotgjengere.
- **Ulykketype:** Rapporteringsgraden er som regel lavere ved eneulykker enn ved kollisjoner (Borger et al., 1995; Elvik & Mysen, 1999).
- **Land:** Det er store forskjeller i rapporteringsgraden mellom ulike land (Elvik & Mysen, 1999; Hutchinson, 1984).

Rapporteringsgrad i Norge: Rapporteringsgraden for ulike typer trafikkulykker i Norge er undersøkt av Lund (2019), basert på politirapporterte personskadeulykker og informasjon fra ulike helseregistre (dødsårsaksregister, statistikkregisteret for uføretrygd, nasjonalt traumeregister, register over spesialisthelsetjenesten (Norsk pasientregister) og registeret over primærhelsetjenesten) i 2017. Resultater viser følgende for ulike skadegrader:

- **Antall drepte:** Det er stor overensstemmelse mellom politiregistrerte dødelige skader og dødsårsaksregister.
- **Antall hardt skadde (meget alvorlig eller alvorlig skadde):** Andelen av alle hardt skadde som er registrert av politiet er anslått til mindre enn 37%. For meget alvorlig skadde er andelen anslått til 11-14%.
- **Antall lett skadde:** Det er anslagsvis 15% av alle lett skadde som er registrert av politiet.
- **Alle trafikkskadde:** Andelen som er registrert av politiet er anslått til 17%.

En ulempe med den relativt store underreporteringen av meget alvorlig skadde er at dette gjør det umulig å undersøke hvordan disse har utviklet seg over tid. En norsk studie fra 2001 (Lund & Bjerkedal, 2001) viste at uføretrygderaten etter trafikkulykker hadde steget mer enn antall drepte har falt i løpet av årene 1988-1997. Med andre ord har det samlede antall drepte og uføretrygdede økt. At antall drepte går ned, kan forklares med bedre akuttmedisinske tjenester. Disse fører til at det er flere som overlever, men en del av disse kan likevel være så alvorlig skadd at de blir ufør. En tilsvarende studie anbefaler Lund (2019) å gjennomføre i Norge.

Basert på data som ble samlet inn på skadelegevakten i Oslo i 2014 har Elvik (2018) estimert rapporteringsgrader i **sykkelulykker**. I kollisjoner mellom sykkel og motorkjøretøy er rapporteringsgraden estimert til 21,5% for lettere skader og 90,0% for alvorlige skader (24,3% for alle skader). I eneulykker med sykkel er rapporteringsgraden estimert til 0,4% for lettere skader og 2,3% for alvorlige skader (0,5% for alle skader). Alle ulykkene sett under ett var rapporteringsgraden 6,5% for lettere skader og 30,2% for alvorlige skader (7,4% for alle skader).

To eldre studier som har undersøkt rapporteringsgraden for ulike ulykkestyper i Norge (Borger et al., 1995; Hvoslef, 1996) har anslått at rapporteringsgraden er på omtrent 33% når man ser alle rapporteringspliktige trafikkulykker under ett. Rapporteringsgraden varierer mye mellom ulike ulykkestyper. For eksempel er rapporteringsgraden for ulykker med personbil estimert til 53%, mens rapporteringsgraden for ulykker hvor ingen motorkjøretøy er innblandet, er langt lavere (2,5% for kollisjoner mellom sykler og 0,7% for enulykker med sykkel). Resultatene baseres på personskaderegisteret ved Statens institutt for folkehelse (nå: Folkehelseinstituttet) og offisiell ulykkesstatistikk. Personskaderegisteret, som eksisterte mellom 1990 og 2002, bygde på skaderegistrering ved fire sykehus og legevakter i ulike landsdeler.

2.2 Klassifisering av skadegrader

Det finnes ulike klassifiseringssystemer for skadegraden i ulykker. I Trafikksikkerhetshåndboken benyttes inndelingen i dødsulykker/drepte, personskadeulykker/personskader, materiellskadeulykker/materiellskader og uspesifisert skadegrad. Personskader (og -ulykker) deles noen ganger inn i hardt skadde og lett skadde; hardt skadde kan videre deles opp i alvorlig og meget alvorlig skadde.

For tiltak som primært virker på antallet ulykker eller skader, skilles det i Trafikksikkerhetshåndboken mellom følgende alvorlighetsgrader for ulykker:

- **Dødsulykker/drepte:** Ulykker der minst ett menneske har blitt drept; som drept i ulykke klassifiseres personer som dør innen 30 døgn etter ulykken av skadene vedkommende ble påført i ulykken.
- **Personskadeulykker/personskader:** Ulykker med personskade, som regel inkludert dødsulykker/drepte; kan videre deles inn i:
 - Hardt skadde (tilsvarende meget alvorlig eller alvorlig skadde i norsk ulykkesstatistikk)
 - Lett skadde.
- **Materiellskadeulykker/materiellskader:** Ulykker uten personskade / kun materielle skader.
- **Ulykker av uspesifisert skadegrad:** Disse omfatter i de fleste tilfeller en blanding av dødsulykker, øvrige personskadeulykker og materiellskadeulykker (eller tilsvarende skader), men med ukjente innbyrdes forholdstall. Som regel dominerer materiellskadeulykker og/eller lettere personskadeulykker.

Skadegraden ved personskader kan klassifiseres med hjelp av ulike skalaer. I Norge skilles mellom dødelige, meget alvorlige, alvorlige og lette skader; meget alvorlige og alvorlige skadde slås ofte sammen til hardt skadde. I USA er det vanlig å bruke de såkalte KABCO-skalaen (K: Fatality; A: Incapacitating injury; B: Evident injury; C: Possible injury; O: Property damage only).

Tre andre skalaer som ofte brukes er AIS, MAIS og ISS.

- **AIS (Abbreviated Injury Scale):** AIS-skalaen har syv verdier. AIS 6 er en dødelig skade, AIS 3, 4 og 5 er alvorlige skader, AIS 2 er en moderat skade, AIS 1 er en lettere skade, og AIS 0 er uskadd. Det beregnes én verdi for skader på hver av ni kroppsregioner.
- **MAIS (Maximum Abbreviated Injury Scale):** Verdien en skadd person får på MAIS-skalaen er den høyeste verdi hun / han har på AIS-skalaen.

- **ISS (Injury Severity Score):** ISS er derivert fra AIS skalaen og har verdier fra 1 til 75. De 9 kroppsregionene som er definert med AIS-skalaen er omgruppert til 6 kroppsregioner. ISS beregnes som sum av de kvadrerte verdiene for de tre kroppsregionene med de høyeste AIS verdiene. Hvis en kroppsregion har en skade som er klassifisert som AIS 6, er ISS-verdien 75.

Hvordan disse tre skalaene brukes, kan illustreres med følgende eksempel: Person A har en meget alvorlig hodeskade (AIS 5) og en lettere kneskade (AIS 2). MAIS-verdien er MAIS 5. Person B har samme kneskade (AIS 2) men ingen hodeskade, MAIS-verdien er MAIS 2. På ISS skalaen er verdien for person A $2^2 + 5^2 =$ ISS 29. For person B er ISS verdien $2^2 =$ ISS 4. Person C har samme skader som person A, men dør av hodeskaden. Hodeskaden vil da være klassifisert som AIS 6. Person C har MAIS 6 og ISS 75.

2.3 Antall trafikulykker i Norge og utvikling over tid

Det finnes ulike mål på trafikksikkerhet i ett land; de viktigste er antall drepte og skadde (eller antall ulykker), antall drepte/skadde per innbygger og antall drepte/skadde per million kjøretøykilometer. I Norge har antall drepte og skadde gått betydelig ned i løpet av de siste årtiene. For antall drepte og skadde per innbygger og per million kjøretøykilometere har nedgangen vært enda større fordi både befolkningen og – i enda større grad – trafikkmengden har økt.

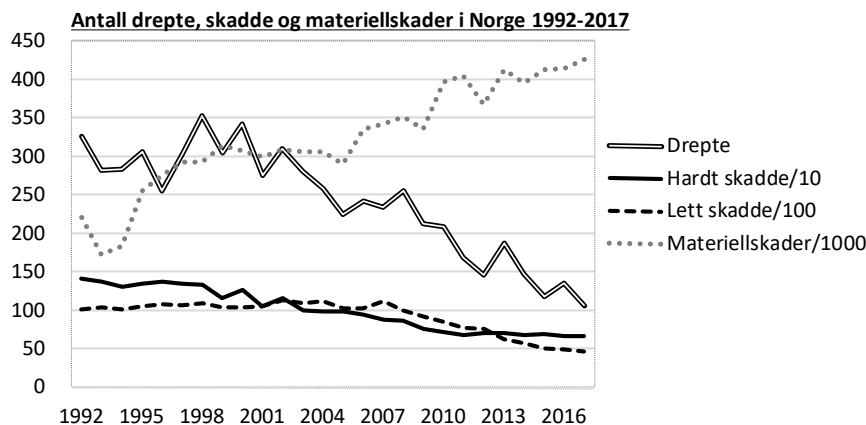
Hvert år er det flere hundre tusen skader i vegtrafikken, hvorav de aller fleste er materiellskader og dermed ikke rapporteringspliktige. I årene 2013-2017 har det i gjennomsnitt per år vært 418 000 materiellskader (rapportert til det skadeforsikringsbaserte registeret TRAST) og 6 063 personskader (politirapporterte personskadeulykker). Dvs. at 1,5% av alle skadene medfører personskade. Blant personskadene er de aller fleste lettere skader (5246 eller 87%). Andelen hardt skadde var 11% (678 skadde) og 2,3% av alle skadde ble drept (138 drepte). Alle antallene og andelene gjelder gjennomsnittet i årene 2012-2017.

I hvilken grad trafikksikkerheten er bedret i Norge etter 1970 avhenger av hvordan man definerer trafikksikkerhet. Generelt finnes tre definisjoner:

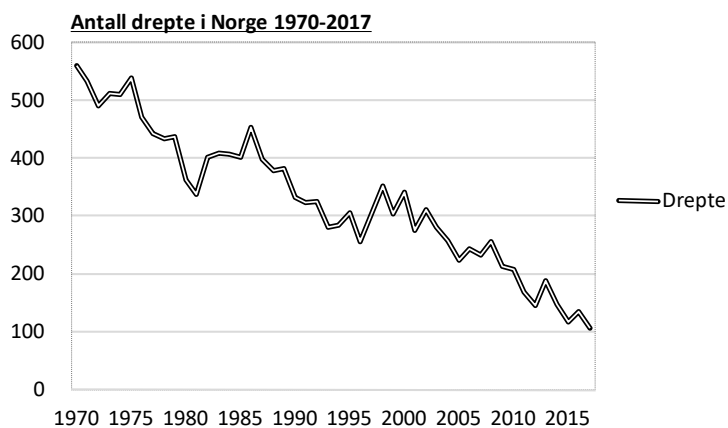
- **Antall ulykker, drepte og skadde:** Gjennomsnittlig antall ulykker og skadde eller drepte personer per år.
- **Befolkningens helseisiko:** Antall drepte eller skadde per 100.000 innbyggere per år.
- **Trafikkens systemisiko:** Antall drepte eller skadde per million kjøretøykilometer (ulykkesisiko).

Antall person- og materiellskader

Utviklingen av antallet drepte, hardt skadde, lett skadde og materiellskader i vegtrafikkulykker i Norge er vist i figur 2 og figur 3. Fram til ca. 1970 økte antallet skadde personer i politirapporterte trafikulykker i Norge. Antall drepte var på sitt høyeste i 1970, med 560 drepte.



Figur 2: Antall drepte, hardt skadde og lettere skadde i politirapporterte trafikkuulykker i Norge 1992-2017 (SSB) og antall materiellskader fra 1992-2017 (TRAST).



Figur 3: Antall drepte i politirapporterte trafikkuulykker i Norge 1970-2017 (SSB).

Antallet drepte kan variere en god del fra år til år som følge av tilfeldigheter, men i et langsiktig perspektiv er det likevel ingen tvil om at antallet drepte i trafikken i Norge har vært synkende. Det samme gjelder også for antall personskadeulykker, mens antall materiellskader har økt over tid. Den prosentvise endringen fra 1992 til 2017 er -67% for drepte, -53% for hardt skadde, -55% for lett skadde og +93% for materiellskader.

Helserisiko: Drepte og skadde per 100.000 innbyggere

Befolkningen i Norge har økt med 37% fra 1970 til 2017 (fra 3,86 mill. til 5,296 mill.) og med 23% fra 1992 (4,274 mill.) til 2017 (basert på tall fra SSB). Utviklingen av antall drepte og skadde per 1000 innbyggere ligner på utviklingen av de absolutte antallene drepte og skadde, men den prosentvise nedgangen er noe større på grunn av befolkningsveksten. Den prosentvise endringen fra 1992 til 2017 er -73% for drepte, -62% for hardt skadde og -63% for lett skadde.

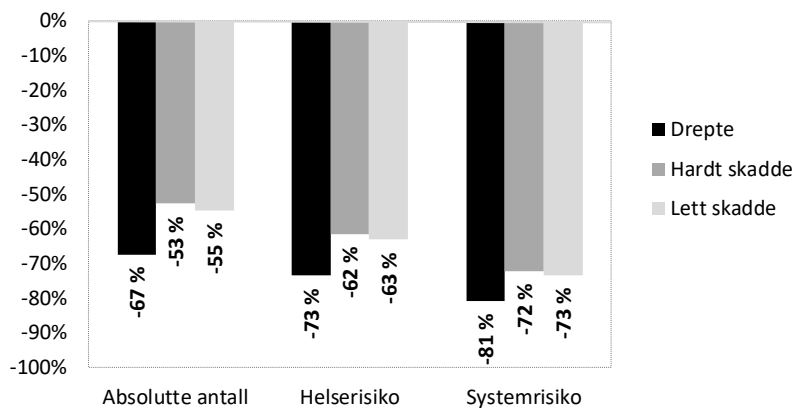
Helserisiko er lite egnet som mål på hvor sikkert det er å ferdes i trafikken eller for å sammenligne trafikksikkerheten mellom ulike land. Helserisiko i trafikken avhenger av hvordan og hvor mye befolkningen reiser, risikoen per kjørt kilometer og medisinsk behandling av skader. Land med få motorkjøretøy kan ha lav helserisiko i vegtrafikken, selv om risikoen per kjøretøy eller per kilometer er svært høy. Land med mange motorkjøretøy kan ha høy helserisiko, men lav systemrisiko. Vurderingen av helserisikoen vil trolig også avhenge av hvor mye trafikken bidrar sammenlignet med andre dødsårsaker.

Trafikkens systemrisiko: Drepte og skadde per million kjøretøykilometer

Det totale antall kjøretøykilometer i Norge har fra 1992 til 2017 økt med 70% eller omtrent 2% per år. Dette er basert på Farstad et al. (2018) og omfatter all person- og godstransport i Norge. Utviklingen av trafikkens systemrisiko ligner på utviklingen av de absolutte antallene drepte og skadde og utviklingen av helserisikoen, men den prosentvise nedgangen er større fordi det totale antall kjøretøykilometer har økt mer enn befolkningen. Den prosentvise endringen av risikoen fra 1992 til 2017 er -81% for drepte, -72% for hardt skadde og -73% for lett skadde.

Trafikkens systemrisiko er et bedre mål på trafikksikkerheten enn både absolutte antall skadde/drepte og antall skadde/drepte per innbygger. Dette fordi den viser hvor høy risikoen er for dem som ferdes i trafikken.

Figur 4 viser en oversikt over de prosentvise endringene av absolutte antall drepte, hardt skadde og lettere skadde, helserisikoen og systemrisikoen som er beskrevet ovenfor.



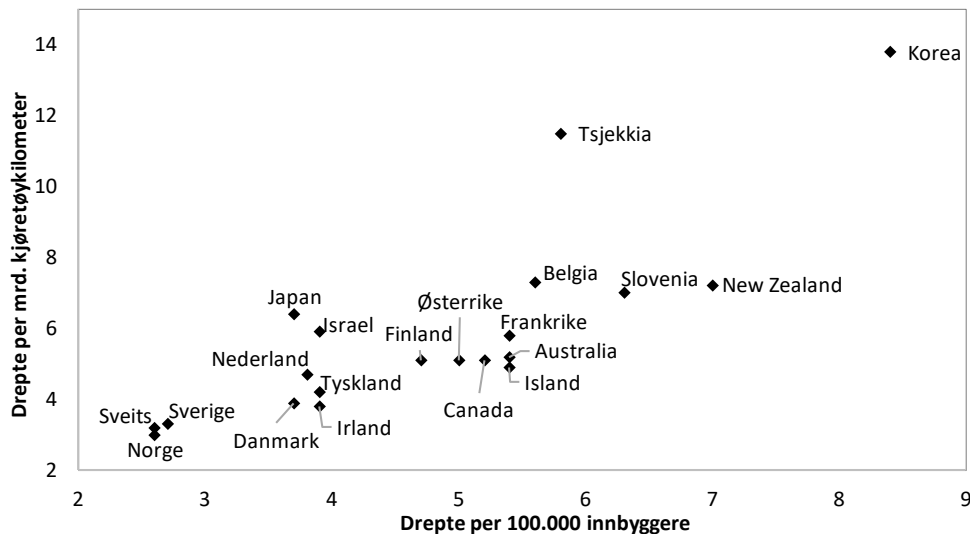
Figur 4: Endringene av absolutte antall drepte, hardt skadde og lett skadde, helserisikoen og systemrisikoen i Norge i 1992-2017.

Som figuren viser har nedgangen av antall drepte vært større enn nedgangen av antall hardt og lett skadde. Nedgangen var størst for systemrisikoen, mindre for helserisikoen og minst for absolutte antall. Hvordan risikoen i vegtrafikken har forandret seg for enkelte trafikantgrupper er beskrevet i rapportene «Risiko i vegtrafikken» (Bjørnskau, 2000, 2003, 2008, 2011, 2015).

2.4 Risiko i vegtrafikken i Norge sammenlignet med andre land

Norge har i mange år vært ett av landene med færrest drepte i trafikken, både per 100.000 innbyggere og per million kjøretøykilometer.

Norge har et lavt risikonivå i vegtrafikken sammenlignet med andre land som har noenlunde det samme antall personbiler per innbygger som Norge. På grunnlag av den internasjonale IRTAD-databasen viser figur 5 helse- og systemrisikoen i trafikken i ulike land.



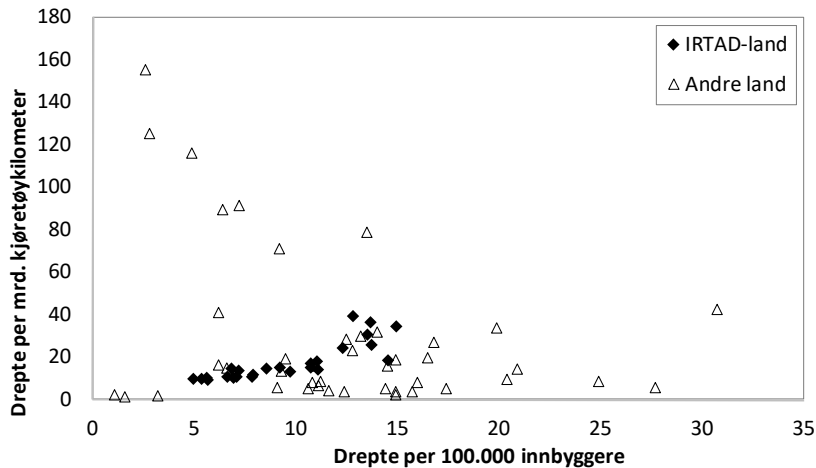
Figur 5: Helse- og systemrisiko (drepte per 100.000 innbyggere) og trafikkenes systemrisiko (drepte per mrd. kjøretøykilometer) i 20 land i 2016 (IRTAD).

I 2016 var de tre land med lavest antall drepte, både per innbygger og per kjøretøykilometer, Norge, Sveits og Sverige. Norge har i mange år vært blant landene med lavest antall drepte per innbygger og kjøretøykilometer, også da antallet drepte i trafikken i Norge var mye høyere enn det er i dag (OECD, 1994). Også Sverige, Sveits og Nederland har lenge vært blant de «beste» på trafikksikkerhet.

Blant forklaringene på den relativt lave risikoen i vegtrafikken i Norge er at Norge alltid har hatt fartsgrenser og at disse er relativt lave sammenlignet med fartsgrensene i andre land (Elvik, 1995A). Norge innførte også tidlig en fast promillegrense på 0,5 (0,2 siden 2001) og har strenge sanksjoner for promillekjøring.

USA har høyere antall drepte per innbygger enn alle andre land i figuren (11,6; ikke vist i figuren da dette ville gjøre den svært vanskelig å lese), mens antall drepte per kjøretøykilometer er på nivå med for eksempel Belgia og New Zealand (7,3). Antall kjøretøy per innbygger er langt høyere i USA enn i andre land.

Figur 5 viser for IRTAD-land (land som er medlem i OECD) at det er en positiv sammenheng mellom helse- og systemrisiko, dvs. at land med høyere helse- og systemrisiko i gjennomsnitt også har en høyere systemrisiko enn land med lav helse- og systemrisiko. En slik sammenheng finner man imidlertid ikke i alle land. Figur 6 viser sammenhengen mellom helse- og systemrisiko i IRTAD-landene (landene som inngår i figur 5) og en rekke andre land. Figuren er basert på eldre data (2009) og resultatene for IRTAD-landene avviker derfor noe fra figur 5.



Figur 6: Helseisiko i trafikken i IRTAD land og andre land (Elvik et al., 2009A).

Figur 6 viser en tendens til at lavere helseisiko medfører høyere systemrisiko i ikke-IRTAD land. Især blant landene med lav helseisiko er det stor variasjon i systemrisikoen. Tallene kan imidlertid være vanskelige å sammenligne. I vestlige, høyt motoriserte land regner man med at alle drepte i trafikken kommer med i offentlig statistikk. Men dette gjelder ikke nødvendigvis alle land i verden. Dessuten varierer definisjonen av drepte i trafikken. Mange utviklingsland regner bare med drepte på stedet, ikke drepte innen 30 dager, som de fleste vestlige land.

2.5 Faktorer som påvirker antall trafikkuulykker og skadegraden i ulykker

Antall ulykker og skadegraden i ulykkene kan generelt sett påvirkes av de tre overordnede faktorene trafikkmengde, ulykkesrisiko og skaderisiko. Trafikksikkerhetstiltak kan redusere antall ulykker/skader ved å påvirke en eller flere av disse, med ulike grader av frivillighet. Trafikkmengden henger som regel ikke lineært sammen med ulykkesrisiko, slik at ulykkesrisikoen som regel synker med økende trafikkmengde (safety-in-numbers).

Antallet skadde personer i trafikken er bestemt av tre hovedgrupper av faktorer: Trafikkmengde, ulykkesrisiko (trafikkens systemrisiko) og skaderisiko. Generelt sett kan derfor antallet skadde personer i trafikken reduseres ved å redusere trafikkmengden, ulykkesrisikoen eller skadegraden i ulykker. Den faktoren som i mange studier har vist seg å forklare den største andelen av variasjonen i antall ulykker, er trafikkmengden.

Trafikkmengde

Økende trafikkmengde medfører som regel flere ulykker. Antall ulykker øker imidlertid ikke lineært med trafikkmengden. Når trafikkmengden øker med én prosent øker antall ulykker som regel med mindre enn én prosent, dvs. at ulykkesrisikoen synker med økende trafikkmengde (Høye, 2016A; Elvik et al., 2013). Dette er den såkalte safety-in-numbers effekten. Det finnes imidlertid også en hazard-in-numbers effekt som Elvik (2013B) viser for fotgjengerulykker i gangfelt (økende risiko med økende antall fotgjengere). Elvik & Goel (2019) viser imidlertid at de aller fleste studier som har undersøkt sammenhengen mellom trafikkmengde og ulykker har funnet en safety-in-numbers effekt. Effekten finnes for alle trafikantgrupper, men den er i gjennomsnitt noe større for fotgjengerulykker enn for andre ulykker.

Mulige årsaker til safety-in-numbers kan være at økende trafikk fører til at farten går ned, at trafikantene skjerper oppmerksomheten og at det er mindre risiko for at førere blir uoppmerksomme som følge av manglende avveksling. Videre medfører stor trafikkmengde (nær kapasitetsgrensen) lavere fart og konsekvensene av ulykker er som regel mindre alvorlige ved lavere fart. Veger med stor trafikkmengde har ofte også bedre standard enn veger med lite trafikk.

Mekanismer som kan bidra til safety-in-numbers blant fotgjengere og syklister er undersøkt av Fyhri et al. (2017). Resultatene tyder på at syklister og bilister utvikler bedre interaksjonsmønstre om sommeren når det er mange som sykler enn om våren når det fortsatt er relativt få som sykler. Denne effekten blir imidlertid motvirket av at det om sommeren er flere uerfarne syklister som har høyere risiko enn de mer erfarne syklister som også sykler om vinteren.

Tiltak som påvirker trafikkmengden kan i ulik grad medføre inngrep i folks valgfrihet. Slike tiltak omfatter bl.a. informasjonstiltak som skal få folk til å gå og sykle mer, tiltak som gjør det mer behagelig å sykle og mindre behagelig å kjøre bil, bompenger/vegprising, økte drivstoffavgifter, parkeringsrestriksjoner og kjøreforbud.

Ulykkesrisiko

Ulykkesrisiko, eller trafikdens systemrisiko, er definert som antall ulykker eller skadde per eksponeringsenhet (som regel per kjørt kilometer). En lang rekke risikofaktorer påvirker ulykkesrisikoen i trafikken og mange av disse kan påvirkes av trafikksikkerhetstiltak. De kan for eksempel være knyttet til:

- **Reisemåte/kjøretøytype:** Ulykkes- og skaderisikoen varierer mellom ulike trafikantgrupper og kjøretøytyper. Hvordan disse har utviklet seg over tid, er beskrevet i Bjørnskau (2015).
- **Vegsystemet:** Det er store forskjeller mellom ulykkesrisikoen på ulike typer veger; bl.a. har motorveger typisk betydelig lavere risiko enn andre veger. Gjennomsnittlige risikotall og skadekostnader for ulike vegtyper i Norge er bl.a. beregnet av Høye (2014, 2016A), Høye og Nævestad (2019), Erke og Elvik (2006) samt i flere eldre studier som også har beregnet risikotall for bl.a. kryss, tunneler og broer (Elvik & Muskaug, 1994; Hvoslef, 1995; Wold, 1995; Amundsen & Raner, 1997; Amundsen & Engebretsen, 2008; Sakshaug & Johannessen, 2005).
- **Fysiske miljøfaktorer:** Det finnes mange fysiske miljøfaktorer som bl.a. lysforhold og føreforhold som påvirker ulykkesrisikoen. Bjørnskau (2015) beskriver også bl.a. variasjonen i ulykkesrisiko med ukedag og tid på døgnet.
- **Trafikantene:** Forskjeller i ulykkesrisikoen mellom menn og kvinner i ulike aldre er bl.a. beskrevet av Bjørnskau (2015).

Tiltak som reduserer ulykkesrisikoen omfatter både «passive» tiltak som ikke krever noen atferdsendringer (for eksempel vegbelysning), og tiltak som krever atferdsendringer med ulike grader av frivillighet. Sistnevnte omfatter for eksempel fartskampanjer som helt frivillig tiltak, fartsgrenser og fartshumper som mindre frivillig tiltak, og tvingende ISA som ikke-frivillig tiltak.

Skaderisiko

Med skaderisiko menes risikoen for at man blir skadet, gitt at man er blitt innblandet i en trafikkulykke. Skadenes konsekvenser for de skaddes livskvalitet avhenger blant annet av hvor alvorlige skadene er (skadegraden) og hvor vellykket behandlingen av skadene er, i og utenfor medisinske institusjoner. Skadereduserende tiltak kan være «passive» (for eksempel forbedret kollisjonssikkerhet) eller kreve en handling fra trafikantens side, som det å bruke hjelm eller bilbelte.

Faktorer som påvirker skaderisikoen i ulykker er bl.a.:

- **Kjøretøyets masse:** Større masse medfører som regel bedre beskyttelse for dem som befinner seg i kjøretøyet, samt økt skaderisiko for andre personer (Høye, 2019). Det betyr bl.a. at lastebiler har stort skadepotensiale for andre trafikanter, mens lastebilførere langt sjeldnere enn andre trafikanter blir drept eller skadd i ulykker. På den andre enden av skalaen er syklistene og fotgjengere som i de fleste ulykkene er den part som får de mest alvorlige skadene.
- **Farten i ulykkesøyeblikket:** Høyere fart medfører, hvis alt annet er likt, mer alvorlige skader. Med økende fart øker skaderisikoen langt mer enn proporsjonalt med farten, mest for de mest alvorlige skadene. Dette er relativt godt dokumentert (bl.a. Elvik, 2009, 2013A; Elvik et al., 2019). Fart er også avgjørende for skadeomfanget i kollisjoner mellom motorkjøretøy og fotgjengere (Krøyer, 2015).
- **Egenskaper ved trafikanten:** Det er især alderen som er avgjørende for sannsynligheten for alvorlige skader. Eldre har i gjennomsnitt høyere skaderisiko og mer alvorlige skader enn yngre (Bjørnskau, 2015).
- **Bruk av personlig verneutstyr:** Dette omfatter bl.a. sykkelhjelmer, motorsykkelhjelmer, verneutstyr for motorsyklister og bilbelter.
- **Førstehjelp og medisinsk behandling:** Ved alvorlige skader kan det være avgjørende for utfallet hvorvidt den skadde får raskt medisinsk hjelp, samt hvordan vedkommende behandles på skadestedet, transporttiden til sykehus, type sykehus og behandlingen vedkommende får på sykehuset.
- **Andre faktorer:** Andre relevante faktorer er bl.a. ulykkestype (møteulykker er blant de mest alvorlige ulykkene med motorkjøretøy), hvor man sitter i bilen (midtsetet bak er som regel den sikreste plassen) og vegens omgivelser (bratt fjellterreng er det farligste å kjøre utfor i).

2.6 Trafikanter trygghet og utrygghet: Et problem med mange sider

Den subjektive opplevelsen av risiko (trygghet) henger ikke alltid sammen med den faktiske risikoen. Ulykkesbelastede steder er ikke nødvendigvis de samme stedene som de hvor folk flest føler seg utrygge, mens steder som gir en følelse av utrygghet ikke nødvendigvis er spesielt ulykkesbelastet. For ulike transportmidler derimot finner man en større sammenheng mellom opplevd og faktisk risiko.

Med trygghet menes folks følelse av sikkerhet, det vil si hvordan man subjektivt opplever risikoen for ulykker eller ubehagelige hendelser. Trygghet i trafikken avhenger av mange ulike faktorer (i tillegg til faktisk ulykkesrisiko) som bl.a. frivillighet i eksponering for risikoen, grad av personlig kontroll over den, katastrofepotensialet ved ulykker, hvor kjent eller tilvent risikoen er, samt personegenskaper (bl.a. spenningsøkning som henger sammen med å oppsøke eller skape risikofylte situasjoner, Vaa et al., 2002; 2016). Trygghet kan oppleves både på egne og andres vegne (omsorgsangst; Køltzow, 1986). Ulempene som oppstår for samfunnet pga. motorisert trafikk i form av redusert gange og sykling (i stor grad som følge av utrygghet) er minst like store som ulempene i form av luftforurensning og minst dobbelt så stor som ulempene i form av støy ifølge Sælensminde (2004). På den annen side kan falsk trygghet i trafikken også være et problem når dette fører til mindre forsiktig atferd og dermed flere ulykker.

Når man sammenligner hvordan risikoen ved ulike transportmidler oppleves, finner man som regel at kollektive transportmidler oppleves som tryggere enn private transportmidler. Blant private transportmidler oppleves bilen som tryggest og motorsykkel som minst trygg, mens gange og sykling ligger i mellom.

Selv om den faktiske risikoen ofte blir under- eller overestimert, er det i en rekke undersøkelser fra forskjellige land funnet statistisk pålitelige sammenhenger mellom den subjektive og den objektive risikoen ved ulike transportmidler (Elvik & Bjørnskau, 2005). Derimot ble det ikke funnet noen sammenheng mellom subjektiv og objektiv risiko på ulike steder (Hvoslef, 1980; Johansson & Naeslund, 1986; Vaa, 1991). At ulykkene ikke nødvendigvis hoper seg mest opp der hvor folk tror at det er farligst, kan delvis forklares med atferdstilpasning, dvs. at mange er mindre forsiktige der det virker trygt (falsk trygghet) og mer forsiktige der det virker minst trygt.

3 Et faglig grunnsyn

Trafikksikkerhåndboken er skrevet ut fra et faglig grunnsyn som omfatter oppfatninger om blant annet hva som er god og mindre god trafikksikkerhetsforskning, og hvordan man best kan få innsikt i hvorfor ulykker skjer og hvordan ulykker best kan forhindres. Et faglig grunnsyn kan ikke begrunnes fullt ut i den forstand at man en gang for alle kan vise at det er riktig eller galt, men noen oppfatninger er likevel bedre begrunnet enn andre. I dette kapitlet legger vi fram vårt syn på:

- **Ulykkesårsaker og risikofaktorer:** Hvorfor skjer trafikkulykker?
- **Ulykker og risikotall som mål på trafikksikkerheten:** Hva er det beste målet på trafikksikkerhet?
- **Trafikkulykker som selvregulerende problem:** Hvorfor endrer trafikantene atferd som følge av enkelte trafikksikkerhetstiltak?
- **Tolkning av p-verdier, statistisk signifikans og konfidensintervaller:** Hva sier disse om hvor pålitelig, generaliserbar og praktisk relevant resultater fra empiriske studier er?

Spørsmålet om hvordan man best kan undersøke virkningene av trafikksikkerhetstiltak og hvilke metodekrav som bør stilles til slike undersøkelser krever en mer inngående drøfting. Denne drøftingen er gjennomført i større grad i kapittel 4.

3.1 Ulykkesårsaker eller risikofaktorer?

Ved å undersøke risikofaktorer for ulykker er det mulig å estimere både ulykkestall og effekter av tiltak, uten at det er nødvendig å ta stilling til ansvar eller skyld. Førere kan ofte betraktes som de skyldige i ulykker. Ansvaret for ulykker kan likevel ligge i faktorer utenfor føreren, som for eksempel utformingen av vegsystemet, samt i organisatoriske faktorer som påvirker slike faktorer.

En utbredt oppfatning fram til ca. 1960 var at man ikke kunne bekjempe trafikkulykker effektivt uten å kjenne «trafikkulykkenes egentlige årsaker». En slik oppfatning kom bl.a. til uttrykk i den første stortingsmeldingen om trafikksikkerhet i Norge (Justisdepartementet, St meld 83, 1961-62, Om tiltak for å fremme trafikksikkerheten), hvor det heter:

«En grundig planlegging av tiltakene for å hindre trafikkulykker er av stor betydning dersom en skal oppnå gode resultater. Skal planleggingen bli effektiv, er det nødvendig å kjenne og analysere de problemer i trafikken som tiltakene kan rettes mot. Et slikt planleggingsarbeid er det i dag ikke mulig å gjennomføre helt tilfredsstillende. En har ennå ikke tilstrekkelig kjennskap til trafikkulykkenes egentlige årsaker og følgelig heller ikke hvilke botemidler som er de beste. Som regel er det et kompleks av årsaker som virker sammen ved trafikkulykker og gjør det vanskelig å vurdere betydningen av de enkelte årsakelementer.»

Det finnes imidlertid svært ulike oppfatninger av hva man kan betrakte som en årsak, det er ikke alltid mulig å finne konkrete «årsaker», og det å identifisere «årsaker» til ulykker er ikke nødvendigvis tilstrekkelig for å kunne forhindre framtidige ulykker. Dette betyr likevel ikke nødvendigvis at det er håpløst å finne effektive tiltak mot trafikkulykker.

Et statistisk årsaksbegrep - risikofaktorer. Med årsaken til en hendelse, mener man ofte en faktor eller hendelse som har hatt en effekt (for eksempel en ulykke), som ellers ikke ville ha skjedd (Elvik, 2004). De fleste ulykker er et resultat av et sett av medvirkende faktorer som til sammen var tilstrekkelige til å utløse den. I trafikksikkerhetsforskningen tas det utgangspunkt i et statistisk årsaksbegrep: *Risikofaktorer*. Risikofaktorer er faktorer som påvirker sannsynligheten for en ulykke eller skade og er verken nødvendige eller tilstrekkelige betingelser for ulykker. Kunnskap om risikofaktorer gjør det mulig å forutsi ulykkestall eller å estimere effekter av tiltak, uten at det er nødvendig å ta stilling til verken årsaker eller skyld.

Skyld, ansvar og årsak er ikke det samme. I offentlig debatt om årsaker til trafikkuulykker hevdes det av og til at «80-90% av ulykkene skyldes menneskelig svikt». Slike utsagn kan tilsynelatende underbygges med resultater fra ulykkesanalyser. Ulykkesanalysegruppene (UAG) til Statens vegvesen har for eksempel funnet at det er store andeler av dødsulykkene hvor førerrelaterte faktorer har bidratt til at ulykken skjedde og til skadeomfanget. De mest vanlige førerrelaterte faktorene (med andel av dødsulykkene hvor de ble funnet) er: Høy fart etter forholdene (32% av ulykkene), manglende informasjonsinnhenting (28%), kjøring i ruspåvirket tilstand (22%), feil beslutning/avgjørelse (16%), godt over fartsgrensen (15%) og trøtthet (14%). Faktorene som har bidratt til skadeomfanget i flest ulykker er manglende bruk av bilbelte (30%) og kritisk treffpunkt (23%). Andelene er basert på ulykkene i 2005-2015.

Allerede i 1987 kommenterer Grime (1987) følgende om britiske dybdestudier av ulykker: *«Menneskelige faktorer var til stede ved omlag 95% av ulykkene. Det er kanskje ikke overraskende, siden trafikantene er innblandet i alle trafikkuulykker og det nesten alltid er mulig å tenke seg noe trafikantene kunne ha gjort for å unngå ulykken. Men, når vi tenker på tiltak, så er ikke de mest effektive tiltakene nødvendigvis knyttet til den dominerende utløsende faktor ved ulykker, men kan ligge på et annet område.»*

Med andre ord må man også spørre hvorfor feilhandlingen ble begått. Mye av forklaringen på feilhandlinger i trafikken kan ligge i utformingen av trafikksystemet. Bl.a. har det vist seg at vegutformingen i stor grad påvirker sannsynligheten for at førere begår feilhandlinger og for at disse fører til ulykker.

Dette betyr at man må skille mellom skyld og ansvar. Selv om vegtrafikkloven eller trafikkreglene er brutt av minst en av de involverte i de fleste ulykkene (dvs. at minst en av de involverte kan holdes rettslig ansvarlig), kan man ikke uten videre konkludere at de viktigste risikofaktorene bak ulykkene er knyttet til egenskaper ved førerne.

Dette synet er sentralt i nullvisjonen for trafikksikkerhet, som sier at ansvaret for trafikkuulykker ligger hos dem som planlegger og driver trafikksystemet, mens trafikantene er ansvarlige for å følge reglene for bruk av trafikksystemet (jf. avsnitt 1.2).

Ulykkesanalysegrupper og faktorer som medvirker til ulykker. Etter større ulykker med fly, skip og tog er det vanlig å nedsette ulykkeskommisjoner som undersøker mulige årsaker til ulykkene. I Norge har alle dødsulykker i vegtrafikken blitt undersøkt av ulykkesanalysegrupper (UAG) siden 2005. UAG er organisert på region- og distriktsnivå, jobber etter retningslinjer fra Statens vegvesen og har kompetanse på veg, kjøretøy og trafikant. Både Statens vegvesen og politiet har i tillegg varslingsplikt til Statens havarikommisjon for transport, som har en egen seksjon for etterforskning av vegtrafikkuulykker siden 2005. Havarikommisjonen skal primært varsles etter ulykker i tunnel og ulykker som involverer tunge kjøretøy eller transport av farlig gods.

UAG utarbeider ulykkesrapporter som oppsummerer hvilke faktorer som antas å ha medvirket til at ulykken skjedde og til skadeomfanget. Her unngår man bevisst begrep som «årsak» og «skyld». Slike medvirkende faktorer omfatter både faktorer ved de innblandede trafikantene, vegrelaterte faktorer, miljøfaktorer, faktorer knyttet til redningsarbeid, samt generelle organisatoriske faktorer som kan ha bidratt til de andre faktorene. En grunnleggende begrensning ved dybdestudier av ulykker er at slike studier mangler en kontrollgruppe og at man derfor ikke uten videre kan si om for eksempel bestemte feilhandlinger er overrepresentert i ulykker i forhold til i trafikken ellers.

3.2 Ulykker og risikotall som mål på trafikksikkerhet

Antall ulykker (eller skadde/drepte personer) kan være et relevant mål på trafikksikkerheten, for eksempel når man undersøker endringen i ulykkestall som følge av at man setter inn ett tiltak. Det kan være nyttig å skille mellom registrerte og forventede ulykkestall. Forventede ulykkestall er det antallet som man kan forvente i det lange løp ved uendret risiko, trafikkmengde etc. og er dermed i mindre grad påvirket av tilfeldige svingninger enn registrerte ulykkestall. Når risikotall brukes som mål på trafikksikkerheten, dvs. at man setter antall ulykker i forhold til et eksponeringsmål (for eksempel trafikkmengde eller befolkningstall), er det viktig å ta hensyn til sammenhengen mellom eksponeringsmålet og antall ulykker. Med økende eksponering øker som regel også antall ulykker, men antall ulykker øker mindre enn proporsjonalt med økningen av eksponeringen.

Registrerte og forventede ulykkestall

Trafikksikkerhet kan defineres ut fra det registrerte eller forventede ulykkes- eller skadetall eller som risiko (trafikkens systemrisiko eller befolkningens helserisiko; jf. avsnitt 2.3). Alle definisjonene medfører en del tolkningsproblemer. Dette avsnittet drøfter noen av disse problemene.

Med **registrert ulykkestall** menes antall ulykker (eller antall skadde eller drepte personer) som er registrert på en veg eller i et område i en gitt tidsperiode. Det kan være stor tilfeldig variasjon i registrerte ulykkestall, især når det er få ulykker. Dette kan være et problem i evalueringstudier, især når tiltak settes inn på spesielt ulykkesbelastede steder. Når ulykkestallet i en før-etter-studie er uvanlig høyt i førperioden, kan man forvente at antall ulykker vil gå ned i etterperioden, selv om man ikke setter inn noe tiltak.

Derfor er det ofte mer hensiktsmessig å bruke det **forventede ulykkestallet**. Med forventet ulykkestall menes det gjennomsnittlige antall ulykker i det lange løp ved uendret trafikkmengde og uendret risikonivå. Ved å beregne det forventede ulykkestallet i førperioden av før-etter undersøkelser unngår man å overestimere effekten av tiltak som settes inn på spesielt ulykkesbelastede steder.

Forventede ulykkestall er det ikke mulig å måle da man teoretisk ville måtte ha en uendelig tidsperiode hvor alle faktorer som kan påvirke antall ulykker, er uendret. Med den såkalte empiriske Bayes (EB) metoden er det mulig å beregne det forventede ulykkestallet som et vektet gjennomsnitt av det observerte ulykkestallet og det **«normale» ulykkestallet**. Det normale ulykkestallet er estimert ut fra den generelle sammenhengen mellom vegegenskaper og antall ulykker. Ved beregningen av det forventede ulykkestallet får det normale ulykkestallet større vekt jo lavere antallet er. Dette fordi det er en større andel tilfeldig variasjon i små ulykkestall enn i store ulykkestall.

Som regel utvikler man multivariate ulykkesmodeller som predikerer normale antall ulykker som en funksjon av trafikkmengden og ulike vegegenskaper. Slike modeller er ofte Poisson eller negative binomial-modeller, som har følgende form:

$$\text{Forventet antall ulykker} = \alpha * V^\beta * e^{\sum X_i \beta_i} = \alpha * e^{\ln(V) * \beta_0 + \sum X_i \beta_i}$$

V er et mål på eksponering (volume, dvs. trafikkmengde), β_i er koeffisienten for trafikkmengden, X_i er andre prediktorvariabler og β_i ($i = 1, 2, \dots$) er koeffisientene for de andre prediktorvariablene. For en mer detaljert beskrivelse av multivariate ulykkesmodeller henvises til Gaudry og Lassarre (2000).

Små ulykkestall er også et problem når man forsøker å tolke *utviklingen over tid*. Hvis man for eksempel observerer en økning av antall drepte i trafikken fra 145 til 187 (+29% - dette var antall drepte i Norge i 2012 og 2013), kan man ikke konkludere at antall drepte er i ferd med å øke (fra 2013 til 2014 gikk antall drepte ned igjen med 21% til 147 og har fortsatt å g ned i årene etterpå). Når det er så stor tilfeldig variasjon er det også lite hensiktsmessig å prøve å finne forklaringer for endringer som ble observert fra det ene året til det neste. Det er mer hensiktsmessig å se på utviklingen over lengre tid og om det finnes en sammenheng med kjente risikofaktorer som kan forklare deler av utviklingen (jf. Høye et al. 2014).

For å beregne *konfidensintervaller* for ulykkestall, kan man legge til grunn at tilfeldig variasjon i ulykkestall som regel følger en Poisson-fordeling. Varians og gjennomsnitt er identiske i en Poisson-fordeling. Et 95%-konfidensintervall kan følgelig beregnes som antall ulykker pluss/minus kvadratroten av antall ulykker multiplisert med 1,96. For eksempel er 95%-konfidensintervallet for 135 drepte (antall drepte i Norge i 2016):

$$145 \pm 1,96 * \sqrt{145} = 145 \pm 24$$

Konfidensintervallet er følgelig fra 121 til 168. Dermed har økningen av antall drepte i Norge fra 2013 til 2014 vært statistisk signifikant, men likevel ikke i tråd med den langsiktige trenden. Betydningen av statistisk signifikans er nærmere diskutert i avsnitt 3.4.

Sammenhengen mellom eksponering og ulykkesrisiko

Ulykkesrisiko er som regel definert som antall ulykker (eller skadde eller drepte) per kjøretøykilometer eller transportarbeid (person- eller tonnkilometer). Dersom man benytter denne definisjonen av risiko i evalueringsstudier, forutsetter man implisitt en lineær sammenheng mellom eksponering og antall ulykker, dvs. at antall ulykker endrer seg lineært med eksponeringen.

Denne forutsetningen er ikke riktig (Hauer, 1995; Elvik et al., 2009C). De fleste risikomål som er definert per kjøretøykilometer eller per personkilometer har en betydelig ikke-linearitet. Dvs. at risikoen som regel henger sammen med eksponeringen, slik at økende eksponering i gjennomsnitt medfører synkende risiko. To eksempler på hvilke problemer dette kan medføre, er beskrevet i følgende.

Eksempel 1 - Safety in numbers: Studier som undersøker virkningen av å bedre sykkelinfrastruktur, finner ofte at tiltak fører til økt antall syklist og lavere ulykkesrisiko for syklist (Elvik & Goel, 2019). Som følge av den såkalte «safety-in-numbers» effekten kan det økte antall syklist i seg selv forventes å føre til lavere ulykkesrisiko. Det vil følgelig være feil å tilskrive hele reduksjonen av ulykkesrisikoen det nye tiltaket. Dette betyr selvfølgelig verken at bedre sykkelinfrastruktur ikke har noen effekt i seg selv, eller at virkningen som følge av økt sykling er irrelevant.

Eksempel 2 - Sammenheng mellom kjøremengde og ulykkesrisiko: Empiriske studier viser ofte at kvinner har høyere ulykkesrisiko enn menn (bl.a. Bjørnskau, 2015; Forsyth et al., 1995). Forskjellen kan forklares med at menn i gjennomsnitt kjører mer enn kvinner og at økende årlig kjørelengde medfører lavere risiko. Når man sammenligner kvinner og menn med samme årlige kjørelengde, finner man ingen forskjell i gjennomsnittlig ulykkesrisiko.

3.3 Trafikkulykker som selvregulerende problem: Teorier om risikolikevekt og atferdstilpasning

Trafikksikkerhetstiltak har som regel som formål å påvirke konkrete risikofaktorer som man antar henger sammen med ulykkesrisikoen. Trafikksikkerhetstiltak kan i tillegg ha utilsiktede effekter på andre risikofaktorer, især førernes atferd. Slike atferdseffekter kan helt eller delvis oppveie virkningen av de risikofaktorer tiltaket er ment å virke gjennom. Når slike atferdseffekter kan oppstå og hvordan de kan påvirke effekten av tiltak, beskrives i teorien om atferdstilpasning. Eldre teorier antok at atferdseffekter av tiltak skyldes at førere kompensere for redusert risiko med å endre atferden slik at risikoen er den samme som uten tiltak.

Teorien om risikolikevekt - en generell teori til forklaring av ulykker?

En generell teori til forklaring av ulykker som lenge har vært diskutert, er teorien om risikolikevekt (risk homeostasis; Wilde, 1982, 1986, 1988, 1994). Ifølge risikolikevektsteorien tilpasser trafikantene hele tiden sin atferd på grunnlag av en sammenligning av opplevd og ønsket risiko. Hvis det er et avvik, tilpasses atferden slik at avviket forsvinner. Eksempler på slike tilpasninger, er at de fleste senker farten når det snør og vegen er glatt eller øker farten på belyste veger. Ut fra samme resonnement ble det i forbindelse med innføringen av standarder for bilsikkerhet i USA på 60- og 70-tallet, betvilt at økt bilsikkerhet kunne føre til økt sikkerhet fordi førerne vil kompensere med høyere fart (Hedlund, 2000).

Hvilket risikonivå trafikantene ønsker, avhenger ifølge risikolikevektsteorien av hvordan de vurderer fordeler og ulemper ved ulik atferd. Her kan det være store individuelle forskjeller. Den opplevde risikoen avhenger ifølge teorien både av det faktiske ulykkestall og av trafikantenes evne til å oppfatte farer i trafikken. Hvordan trafikantene skal få kjennskap til faktisk ulykkesrisiko på ulike veger, er uklart. Siden teorien predikerer at en reduksjon av den opplevde risikoen fører til at trafikantene tilpasser atferden slik at risikoen kommer tilbake til ønsket nivå, slutter Wilde at tiltak som reduserer den faktiske risikoen, ikke kan ha noen effekt med mindre det ønskede risikonivået også senkes.

De viktigste svakhetene i teorien lar seg sammenfatte som følgende (Bjørnskau & Fosser, 1996; Fosser et al., 1996; Hedlund, 2000; Sagberg et al., 1997; Sivak & Tsimhoni, 2008):

- I sin mest ekstreme form predikerer teorien at trafikksikkerhetstiltak aldri vil ha noen effekt. Dette er beviselig feil, det finnes mange tiltak som har vist seg å ha en effekt, og trafikksikkerheten er blitt betydelig bedre i løpet av de siste årtiene, ikke minst på grunn av bedre bilsikkerhet. Det er også feil at ulike typer veg har samme risiko, for eksempel har motorveger langt lavere risiko enn andre veger.
- Det er uklart hvorfor trafikanter skal være motivert til å holde risikoen på samme nivå. Teorien påstår at trafikanter, når risikoen går ned, endrer atferd for å få risikoen tilbake til et høyere nivå.

- I en «mildere» variant kan teorien forklare enhver effekt av sikkerhetstiltak: Redusert antall ulykker kan forklares med at ønsket risikonivå er senket, uendret ulykkestall kan forklares med atferdstilpasningen, økt ulykkestall kan forklares med at atferdstilpasningen har mer enn kompensert for tiltaket uten at den faktiske risikoøkningen ble oppfattet.
- Teorien kan ikke predikere trafikanter atferdstilpasning.
- Teorien sier ingenting om hvordan «ønsket risikonivå» skal måles og hvordan dette best kan påvirkes. Dette blir derfor en udefinert sekkepost hvor alt man ikke klarer å måle kan plasseres.
- Ønsket risikonivå er i teorien uavhengig og upåvirket av faktisk risiko, noe som neppe er en rimelig antakelse.

Det er likevel enighet om at Wilde, gjennom teorien om risikolikevekt, har pekt på vesentlige forhold av betydning for trafikksikkerhet: Især at ikke alle trafikksikkerhetstiltak virker slik man skulle vente fordi folk kan endre atferd, samt at folks ønsker om å unngå ulykker har stor betydning både for deres atferd i trafikken og for muligheter for å gjennomføre tiltak.

Teori om atferdstilpasning

Den generelle logikken i teorier om atferdstilpasning er at trafikksikkerhetstiltak kan påvirke både risikofaktorer tiltaket er ment å virke gjennom, og andre risikofaktorer. Andre risikofaktorer omfatter især førerens atferd, for eksempel fart. Effekten på andre risikofaktorer kan helt eller delvis oppveie virkningen av de risikofaktorer tiltaket er ment å virke gjennom.

Et eksempel er vegbelysning som er ment å øke oppdagelsesavstanden i mørke. Den virkningen vegbelysning ville ha på ulykkene dersom bare oppdagelsesavstanden ble påvirket, kalles *ingeniøreffekten* av tiltaket. Hvis vegbelysning også fører til at trafikantene kjører fortere og blir mindre oppmerksomme (*atferdseffekten*), kan dette redusere virkningen. *Nettoeffekten* av tiltak er bestemt både av ingeniøreffekten og atferdseffekten, samt retningen og styrken i disse effektene.

Faktorer som påvirker atferdstilpasning er bl.a. følgende (Bjørnskau, 1994B; Elvik, 2004; Hedlund, 2000):

- **Synlige tiltak** fører til mer atferdstilpasning enn mindre synlige tiltak. Eksempelvis antas at feltskiftevarsler er mer utsatt for atferdstilpasning enn ettergivende rattstamme. På samme måte kan tiltak som gjør at man føler seg sikrere, tenkes å føre til mer atferdstilpasning enn tiltak som ikke gjør at man føler seg sikrere.
- **Tiltak som reduserer ulykkesrisikoen**, er mer utsatt for atferdstilpasning enn tiltak som reduserer skadegraden ved ulykker. Som eksempel antas at ABS-bremser er mer utsatt for atferdstilpasning enn kollisjonspuler.
- **Tilpasning av atferden (uten tiltak)** til den risikofaktor tiltaket er ment å påvirke, fører til at tiltak er mer utsatt for atferdstilpasning. For eksempel antas periodisk kontroll av lette kjøretøy å være mer utsatt for atferdstilpasning enn vegbelysning, fordi trafikantene kompenserer for tekniske feil og mangler, men de tilpasser ikke atferden til mørke på en slik måte at risikoøkningen i mørke forsvinner.
- **Større ingeniøreffekter** medfører større sannsynlighet for atferdstilpasning. Det er for eksempel mer sannsynlig at forbedring av bilers hovedlykter (som forutsettes brukt som kjørelens) fører til atferdstilpasning ved mørkekjøring enn ved kjøring i dagslys.

- **Muligheten for å oppnå økt nytte** med atferdstilpasning medfører mer atferdstilpasning. Et eksempel er at å kjøre fort på belyste veger gir økt nytte for de aller fleste. Å lese avis (som et ekstremt eksempel) i en bil med feltskiftevarsler og frontkollisjonsvarling, kan også oppleves som noe som gir økt nytte. Det er derimot vanskelig å tenke seg en atferdstilpasning til bomanlegg ved planoverganger mellom veg og jernbane som vil øke trafikantens nytte.

For mange av tiltakene som beskrives i Trafikksikkerhåndboken, er det undersøkt hvorvidt atferdstilpasning kan forekomme. I en del tilfeller finnes det belegg for at atferdstilpasning kan ha ført til at virkningen er mindre enn den kunne ha vært ut fra ingeniøreffekten (for eksempel vegbelysning, reasfaltering). For mange tiltak, især nyere kjøretøytiltak, er virkningene på ulykker hittil i liten grad dokumentert empirisk, men i hovedsak estimert ut fra tiltakenes virkninger på (ønskede og uønskede) effekter på atferd. Ut fra slike beregninger er det imidlertid svært vanskelig eller umulig å estimere virkningen på ulykker da det som regel ikke er mulig å dokumentere alle effekter på atferd og fordi sammenhengen mellom atferd og ulykker i for liten grad er kjent.

3.4 Tolkning av p-verdier, statistisk signifikans og konfidensintervaller

Statistisk signifikans og konfidensintervaller sier noe om den statistiske usikkerheten ved effekter. De sier derimot ingenting om hvorvidt resultatene er pålitelige eller generaliserbare. Konfidensintervaller kan brukes for å diskutere resultatenes praktiske relevans, dvs. hvilken praktisk betydning ulike effekter innenfor konfidensintervallet kan ha. Dette gjelder både den sammenlagte effekten og effekter på de ytterste grensene til konfidensintervallet.

Det er vanlig praksis å rapportere statistisk signifikans og konfidensintervaller for resultater av empiriske studier og metaanalyser. Statistisk signifikans og konfidensintervaller tolkes og brukes ofte på en misvisende måte. Dette avsnittet oppsummerer de vanligste feilene og gir en oppsummering av den statistiske bakgrunnen og anbefalinger for bruken av statistisk signifikans og konfidensintervaller (kort oppsummert i figur 7).

Begrepet **statistisk signifikans** er nært knyttet til statistisk hypotesetesting. I en vitenskapelig studie ønsker man kanskje å undersøke om det er forskjell mellom to grupper som får ulik behandling eller undersøke hvilke faktorer som øker risikoen for ulykker. Nullhypotesen som settes opp, er gjerne at tiltaket eller behandlingen *ikke* har noen effekt eller at en faktor *ikke* påvirker ulykkesrisikoen. Et resultat betegnes som statistisk signifikant hvis det er lite trolig at det kunne inntreffe rent tilfeldig hvis nullhypotesen er sann. P-verdien er definert som sannsynligheten for det man har observert (eller noe mer ekstremt) gitt at nullhypotesen er sann. Det har blitt vanlig praksis å si at en effekt med p-verdi lavere enn 0,05 er statistisk signifikant.

En veldig forenklet men (dessverre) vanlig tolkning av statistisk signifikans er at en signifikant effekt viser at «det finnes en effekt», mens en effekt som ikke er signifikant, må tolkes som «ingen effekt» (Amrhein et al., 2019).

Når en effekt er signifikant på ($p < 0,05$), vil **konfidensintervallet** ikke inneholde null. Konfidensintervallet beregnes som en funksjon av effekten og standardavviket til effekten. Et konfidensintervall (som er basert på $p = 0,05$) viser, rent hypotetisk, intervallet i hvilket 95% av effektene som estimeres i hypotetiske empiriske studier (som alle er nøyaktig like den aktuelle studien), vil ligge *dersom* den effekten man har funnet i den aktuelle studien, er den «sanne» effekten.

Konfidensintervaller avhenger av effektens standardavvik. Det betyr at små konfidensintervaller indikerer liten statistisk usikkerhet, mens store intervaller indikerer stor usikkerhet. Konfidensintervallet for et resultat (en sammenlagt effekt) av metaanalyse vil være liten når undersøkelsene som inngår i effektestimater baseres på mange ulykker og når alle undersøkelsene har funnet lignende effekter. Store konfidensintervaller kan i metaanalyse bety to ting: Stor statistisk usikkerhet i de enkelte resultatene, eller heterogene resultater.

En vanlig men feil tolkning av konfidensintervaller er at disse viser i hvilket intervall den sanne effekten ligger med 95% sannsynlighet. Det finnes imidlertid ingen sannsynlighet for den sanne effekten (den er som den er). Tolkningen at et konfidensintervall som inneholder null, betyr at det ikke finnes noen effekt (at den sanne effekten er lik null), er like feil som at et ikke-signifikant resultat betyr «ingen effekt».

En annen feiltolkning er at et resultat med et lite konfidensintervall er pålitelig. Konfidensintervaller er uavhengige av den metodiske kvaliteten på studiene. En effekt med et lite konfidensintervall er derfor ikke nødvendigvis pålitelig. Hvis de enkelte studiene er metodisk svake, kan effekten likevel være misvisende, uansett hvor lite konfidensintervallet er. Et resultat med et stort konfidensintervall kan være basert på en pålitelig studie, men veldig store konfidensintervaller kan likevel gjøre at resultatet er vanskelig å tolke.

Da p-verdier ble utviklet i 1920 av Ronald Fisher, var formålet ikke å ta beslutninger om det finnes en effekt eller ikke, men å finne ut hvorvidt et resultat kan være verdt å bli sett nærmere på, som en del av en større forskningsprosess (Nuzzo, 2014). P-verdier tolkes ofte slik at de viser sannsynligheten for at et resultat kun er en følge av tilfeldigheter. Denne tolkningen er *feil*. En p-verdi viser sannsynligheten for at man får et visst resultat under gitte forutsetninger; for eksempel:

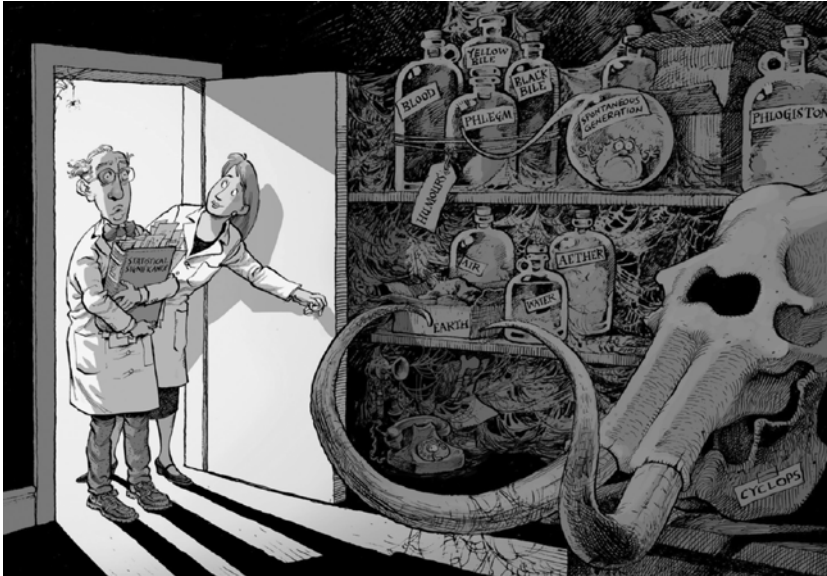
Dersom virkningen av tiltak X ikke har noen effekt på antall ulykker (effekten er $\pm 0\%$) og man gjennomfører den aktuelle studien et uendelig antall ganger, vil andelen ganger man får en effekt som er like stor eller større enn i den aktuelle studien, være lik p.

For å gjenta: P-verdien sier *ingenting* om sannsynligheten for å få et visst resultat (for eksempel $\pm 0\%$) eller for at den «sanne» effekten har en viss størrelse. P-verdier sier heller ikke noe om hvor sannsynlig det er at man vil replisere et visst resultat i en senere studie (Nuzzo, 2014).

Den relativt utbredte bruken av statistisk signifikans som kriterium for «effekt vs. ingen effekt» medfører en rekke problemer og ulemper (Amrhein et al., 2019; Ziliak & McCloskey, 2008):

- Hvis to effekter er like store, men kun den ene av dem er statistisk signifikant, er effektene ifølge signifikanstesten inkompatible. Dersom en effekt A er liten men statistisk signifikant, og en effekt B er *større* enn A men ikke-signifikant, måtte man konkludere at effekt B er *mindre* (nemlig en «ikke-effekt») enn A (en «effekt»). En slik tolkning ville være feil bruk av hypotesetesten og konklusjonene vil være misvisende eller feil.
- Effekter som er basert på svært store antall (av for eksempel ulykker) kan være statistisk signifikante selv om de er så små at de ikke er av noen som helst praktisk betydning. Motsatt kan svært store effekter være ikke-signifikante men likevel av stor praktisk betydning.
- Konfidensintervaller som inneholder null, kan også inneholde store effekter av stor praktisk betydning. Tolkningen av et konfidensintervall som inneholder null som «ingen effekt» kan derfor føre til konklusjoner som har lite mer virkeligheten å gjøre.

- Effekter som ligger rett utenfor intervallet, er veldig like effekter som ligger rett innenfor intervallet og ligger omtrent like langt fra intervallets midtpunkt og motsatt ende. Det blir dermed feil å generalisere tolkninger for alle effekter innenfor vs. alle potensielle effektene utenfor intervallet.
- Bruken av 95% intervaller er et vilkårlig kriterium og det finnes ingen faglig begrunnelse. Man kunne like godt ha brukt for eksempel 93,72%.



Figur 7: Anbefaling for bruk av statistisk signifikans (Amrhein et al., 2019).

American Statistical Association (ASA, 2016; Wasserstein & Lazar, 2016) har gitt ut en veileder om tolkning av p-verdier og statistisk signifikans som inneholder de følgende seks prinsippene:

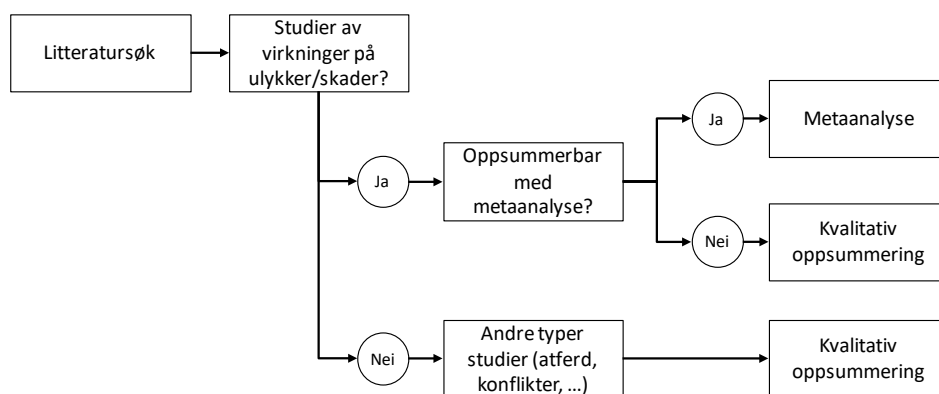
- (1) P-verdier kan indikere i hvilken grad data er kompatible med en spesifisert statistisk modell.
- (2) P-verdier måler *ikke* sannsynligheten for at den undersøkte hypotesen er sann eller at dataene kun er et resultat av tilfeldigheter.
- (3) Vitenskapelige konklusjoner og beslutninger om bruk av tiltak bør *ikke* utelukkende være basert på om en p-verdi ligger over eller under en viss grense.
- (4) Korrekte slutninger forutsetter full rapportering og transparens.
- (5) En p-verdi og statistisk signifikans måler *ikke* størrelsen på en effekt eller dens betydning.
- (6) En p-verdi i seg selv er *ikke* et godt mål på forklaringsverdien av en modell eller en hypotese.

Som **alternativ** til statistisk signifikans som et «ja/nei» kriterium, anbefaler Amrhein et al. (2019) og Nuzzo (2014) å rapportere både effektens størrelse og konfidensintervaller og å diskutere den praktiske betydningen av både effektestimater og de ytre grensene til konfidensintervallet.

For å oppsummere kan konfidensintervaller benyttes for å diskutere størrelsen og praktisk relevans til et effektestimater og effekter innenfor konfidensintervallet, især ved de ytterste grensene. De kan ikke brukes for å slå fast om det finnes eller ikke finnes en effekt. I Trafikksikkerhåndboken rapporteres konfidensintervaller for alle sammenlagte effekter som er beregnet med metaanalyse og det er lagt vekt på en diskusjon av den praktiske relevansen og generaliserbarheten til resultatene. P-verdier rapporteres ikke i Trafikksikkerhåndboken (det benyttes $p=0,05$ ved beregning av konfidensintervaller).

4 Opplegg for arbeidet med Trafikksikkerhåndboken

For alle kapitlene i Trafikksikkerhåndboken gjøres et systematisk litteratursøk for å finne relevante studier. Hovedfokuset er på studier som har undersøkt virkningen på ulykker, antall skadde eller drepte personer eller skadegraden i ulykker. Fremgangsmåten er skjematisk illustrert i figur 8.



Figur 8: Skjematisk fremgangsmåte for litteratursøk ved oppdatering av tiltaks kapitler i Trafikksikkerhåndboken.

Når det er mulig, er resultatene oppsummert med **metaanalyse** (se avsnitt 4.2). Forutsetninger for dette er beskrevet nedenfor (avsnitt 4.2.2). Kort sagt er det nødvendig at det for et visst antall studier er mulig å beregne et felles effektmål samt usikkerheten til effekten. Effektmålet er som regel den prosentvise endringen av antall ulykker eller skader med tiltak sammenlignet med uten tiltak.

Når det ikke er mulig å oppsummere resultatene med metaanalyse, er det gjort en **kvalitativ oppsummering** av resultatene, ofte supplert av tabelloversikter over tilgjengelige resultater fra de enkelte studiene.

Dersom det finnes ingen, svært få eller kun metodisk svake studier av virkninger på ulykker, skadde/drepte eller skadegrad, er også andre typer studier tatt med. Det kan for eksempel være studier som har undersøkt virkningen på trafikantatferd, konflikter, eller andre indirekte mål på ulykkers- eller skaderisiko. Slike studier kan kun indirekte si noe om mulige virkninger på antall ulykker. Det finnes noen atferdsindikatorer som har en relativt godt kjent sammenheng med antall ulykker (for eksempel fart), men for mange studier er sammenhengen i liten grad kjent, og det kan være flere faktorer som påvirker ulykker enn de som er undersøkt.

4.1 Systematisk litteratursøking

For alle tiltak som er beskrevet i Trafikksikkerhåndboken, er det gjort systematiske litteratursøk for å finne flest mulig empiriske studier som har undersøkt tiltakenes effekt på ulykker/skader.

Trafikksikkerhåndboken bygger på en omfattende gjennomgang av norske og utenlandske undersøkelser om virkninger av trafikksikkerhetstiltak på ulykker og skader. Slike undersøkelser blir funnet gjennom systematiske litteratursøk som beskrevet i de følgende avsnittene. Systematiske litteratursøk er et viktig første skritt i metaanalyse og bidrar til å unngå systematiske feil i utvalg av studier som inngår i metaanalysen (Egger et al., 2001).

Trinn i systematiske litteratursøk

Relevante trinn i systematiske litteratursøk er beskrevet i PRISMA-sjekklisten (Moher et al., 2009, 2015). Med hjelp av denne sjekklisten kan planlagte litteraturstudier registreres i databaser som for eksempel PROSPERO¹ (noe som i økende grad gjøres for bl.a. Cochrane-reviews, men er mest relevant for medisinske studier).

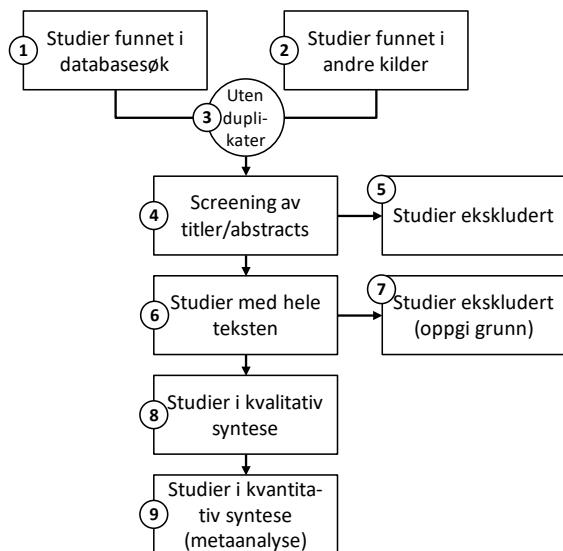
Informasjon om litteratursøket som bør dokumenteres ifølge PRISMA-sjekklisten er:

- (1) Bakgrunn og konkrete spørsmål som skal besvares
- (2) Kriterier for valg av studier, for eksempel typer forsøksopplegg, publikasjonsår, type publikasjon
- (3) Informasjonskilder for litteratursøket og publikasjonsår
- (4) Søkestrategi; denne skal helst beskrives slik at søket lar seg replisere
- (5) Søkeresultatene:
 - Beskrivelse av prosessen for å velge studier, gjerne supplert av et flow-diagram (se avsnittet under) og metoden for å dokumentere prosessen
 - Beskrivelse av metoden for å hente ut data fra de utvalgte studiene
- (6) Variablene for hvilke informasjon skal dokumenteres
- (7) Resultatvariabler for hvilke informasjon skal dokumenteres
- (8) Metoder for å vurdere skjevheter i enkelte studier (for resultater, hele studien eller begge deler), samt hvordan denne informasjonen vil bli brukt.

Utvalg av studier

Hvordan studiene velges ut for videre analyse (metaanalyse eller annen analyse) kan beskrives med hjelp av et flytdiagram. Figur 9 viser et eksempel på et slikt diagram etter Moher et al. (2009).

¹ <https://www.crd.york.ac.uk/prospero/>



Figur 9: Flytdiagram for utvalg av studier etter Moher et al. (2009).

De enkelte trinnene er beskrevet i det følgende. Noen publikasjoner har et krav om at antall studier oppgis for hver av boksene i figuren.

Trinn 1-3 Søk i databaser og andre kilder: Informasjonskilder for litteratursøk i forbindelse med revisjon av Trafikksikkerhetsåndboken er i hovedsak:

- Databasesøk: ISI Web of knowledge, TRID, ScienceDirect (<https://www.sciencedirect.com>), Taylor & Francis (<https://www.tandfonline.com>), TRB (<http://www.trb.org>), PubMed (<https://www.ncbi.nlm.nih.gov/pubmed>)
- Andre kilder: Tidligere utgaver av Trafikksikkerhetsåndboken, internett (i hovedsak Google Scholar).

Søkeord til database- og internettsøk defineres for hvert tiltak. For alle tiltak er ord som beskriver tiltaket kombinert med minst ett av «crash», «accident», «injury» og «fatality».

Supplerende søk: Mens databasesøk er den klassiske tilnærmingen som brukes i de fleste publiserte metaanalysene, kan supplerende søk med andre, mer åpne metoder ha flere fordeler:

- Man kan finne et større antall relevante studier
- Det kan være systematiske forskjeller mellom studier som man finner i databaser og andre kilder, noe som vil medføre skjevhet i resultatene fra metaanalyse dersom man utelater studier funnet med andre metoder.

Cooper et al. (2018) viser at man med hjelp av supplerende søk som regel kan øke antall studier som kan inkluderes i metaanalyse, uten at kvaliteten på studiene er dårligere enn dersom man kun benytter klassisk databasesøk. Det supplerende søket i denne studien omfattet bl.a. internettsøk, kontakt med eksperter, søking i referanselister, og «highly focused» søk i databaser. I en annen studie ble kun 31 av totalt 41 studier inkludert i metaanalyse funnet i databasesøk, mens de resterende ti studiene ble funnet i supplerende søk (Papaioannou et al., 2010). I studien til Helmer et al. (2001) ble 30% av alle relevante studiene funnet i supplerende søk. De mest effektive supplerende søkemetodene i denne studien var søk i spesialiserte databaser, skanning av referanselister og søk for hånd i relevante tidsskrifter.

Etter vår erfaring kan relativt åpne internettsøk (Google, Google Scholar) ikke bare føre til at man finner flere studier, man kan også finne nye søkeord som man ikke hadde tenkt på på forhånd. Dette kan bidra til at man finner relevante studier som man ikke hadde funnet ved utelukkende å gjøre forhåndsdefinerte databasesøk. Videre kan slike søk være mer egnet for å finne studier på andre språk enn engelsk (især norsk og tysk).

Systematiske forskjeller mellom studier som man finner i databaser og andre kilder kan oppstå når «uønskede» eller ikke-signifikante resultater blir refusert av vitenskapelige tidsskrifter. Slike studier forblir ofte upublisert, men kan noen ganger være publisert i former som er mindre tilgjengelige ved formaliserte databasesøk, for eksempel som rapporter, populærvitenskapelige artikler, eller i andre språk enn engelsk (Guyatt et al., 2011C).

Trinn 4-5 screening av titler/abstracts: I dette trinnet velges studier som ble funnet i litteratursøket (trinn 1-3) som kan tenkes å være relevante for syntesen. Trinnet kan deles opp i to (todelingen er ikke del av PRISMA-retningslinjene):

- Screening av *titler*: Titler hvor det er veldig lite trolig at studien inneholder relevant informasjon, sorteres ut, alle andre går videre til screening av abstracts.
- Screening av *abstracts*: Abstracts gir som regel et bedre grunnlag for å vurdere relevansen. Her er det følgelig mulig å være noe mer restriktiv enn ved screeningen av titlene.

I litteratursøk for Trafikksikkerhåndboken er det i dette trinnet viktig å definere kriteriene for utvalg av studiene, avhengig om det vil være mulig å gjøre en metaanalyse:

- Skal det gjøres en metaanalyse må studien handle om det aktuelle tiltaket, virkningen må være undersøkt for antall ulykker, skadde eller drepte personer eller skadegraden i ulykker, og det må være mulig å beregne et effektestimat og usikkerheten til dette (det siste er som regel ikke mulig å vurdere ut fra bare tittel og abstract).
- Skal det ikke gjøres noen metaanalyse, kan kriteriene defineres mindre strengt og i løpet av prosessen kan det vurderes hvorvidt kun ulykkesstudier skal være med eller om også andre typer studier er relevante.

Trinn 6-7 screening av originalstudiene: For alle studiene som har gått videre fra forrige trinn, skaffes hele publikasjonen. For studier som ikke viser seg å være relevante, oppgis grunnen (for eksempel at tiltaket er ikke relevant i sammenhengen eller at publikasjonen ikke inneholder informasjon for å beregne effektestimat eller statistisk vekt). Studier som oppfyller alle kriteriene, går videre til syntesen.

Trinn 8 kvalitativ syntese: Dette trinnet er ved revisjonen av Trafikksikkerhåndboken kun relevant dersom det ikke er funnet relevante studier av virkninger på ulykker som kan oppsummeres med metaanalyse.

Trinn 9 kvantitativ syntese (metaanalyse): I dette trinnet blir resultatene fra alle relevante studiene oppsummert med metaanalyse. Dette er beskrevet nærmere i neste avsnitt.

4.2 Oppsummering av resultater med metaanalyse

Så langt som mulig er studier som har undersøkt effekten av tiltak på ulykker/skader, oppsummert med bruk av metaanalyse i Trafikksikkerhetshåndboken. Metaanalyse er en metode for å beregne en sammenlagt effekt basert på resultatene fra flere sammenlignbare enkeltstudier. I motsetning til en enkel oppstilling av resultater fra enkeltstudiene, er det med metaanalyse i en viss grad mulig å oppdage og korrigere for skjevheter og systematiske feil i resultatene for enkeltstudier.

Metaanalyse er en metode for å beregne vektete gjennomsnitt av effektestimater fra enkelte empiriske studier. Den mest vanlige typen effektestimater i Trafikksikkerhetshåndbok er en prosentvis endring av antall ulykker. Effektestimaterne blir vektet med et mål på den statistiske usikkerheten ved estimatene. Noe forenklet sagt er usikkerheten mindre og vekten større for effektestimater som er basert på mange ulykker enn for effektestimater som er basert på få ulykker.

Egger et al. (2001) siterer følgende om metaanalyse fra en artikkel i Wall Street Journal i 1992:

Meta-analysis begins with scientific studies, usually performed by academics or government agencies, and sometimes incomplete or disputed. The data from the studies are then run through computer models of bewildering complexity, which produce results of implausible precision.

Metoden for å beregne metaanalyse er beskrevet i de følgende avsnittene. At resultatene er av «usannsynlig presisjon» finnes det i dag metoder for å unngå; disse er også beskrevet nedenfor. Bl.a. er det viktig å vurdere systematiske feilkilder både i enkeltstudiene og i utvalget av studier som inngår i metaanalysen, og å skille mellom statistisk signifikans og praktisk betydning og generaliserbarhet.

Retningslinjene i PRISMA-sjekklisten inneholder følgende trinn for syntese (metaanalyse er en form for syntese) etter avsluttet litteratursøk (Moher et al., 2015):

- (1) Kriterier og valg av metode (kvantitativ/kvalitativ syntese)
- (2) Beregning av sammenlagte effekter og vurdering av konsistens
- (3) Tilleggsanalyser (bl.a. sensitivitetsanalyser, subgruppeanalyser, metaregresjon)
- (4) Analyse av skjevheter (bl.a. publikasjonsskjevhet)
- (5) Vurdering av resultatenes pålitelighet («strength» i artikkelen til Moher et al., 2015).

Disse trinnene må ikke nødvendigvis gjennomføres i denne rekkefølgen. Elvik (2005B) foreslår å gjøre en eksplorativ analyse i et første trinn hvor man studerer effektestimaterne i traktogrammer for å vurdere hvorvidt det er meningsfullt å beregne en sammenlagt effekt, om datamaterialet bør deles inn i subgrupper eller om man bør avstå helt fra å beregne sammenlagte effekter. Elvik (2005B) foreslår å bruke traktogrammer i denne eksplorative analysen, men det er også mulig å benytte andre typer diagram (som forest plot) eller en metaregresjonsanalyse.

4.2.1 Kriterier for kvantitativ oppsummering av resultatene med metaanalyse

Resultater fra enkeltstudier som har undersøkt effekten av tiltak på ulykker/skader, kan oppsummeres med metaanalyse når det foreligger informasjon om størrelsen på effekten og den statistiske usikkerheten til denne.

Metaanalyser kan utføres når:

- Det finnes flere resultater (fra mest mulig sammenlignbare studier) som lar seg uttrykke som et felles effektmål (i Trafikksikkerhetskåndboken som regel en prosentvis endring av antall ulykker eller skader)
- Det er mulig å beregne den statistiske usikkerheten som er knyttet til de estimerte effektene (for eksempel basert på antall ulykker som inngår i undersøkelsen).

At studiene bør være mest mulig sammenlignbare innebærer at de har undersøkt effekten av det samme tiltaket på det samme effektmålet. Teknisk betyr det at resultatene fra enkeltstudiene ikke bør være for heterogene; dette er beskrevet i avsnitt 4.2.3.

Opplegget for metaanalyser av undersøkelser om virkninger av trafikksikkerhetstiltak er nærmere beskrevet i egne rapporter (Christensen, 2003; Elvik, 2005C; Borenstein et al. 2009). I denne rapporten beskrives bare hovedtrekkene.

Trafikksikkerhetskåndboken presenterer som regel kun en relativt kort oppsummering av de viktigste resultatene fra metaanalyse. Mer detaljerte beskrivelser av analysene finnes ofte, især for de større analysene som er basert på mange undersøkelser, i egne TØI-rapporter og/eller artikler i vitenskapelige tidsskrifter.

4.2.2 Beregning av sammenlagte effekter: Log-odds metode av metaanalyse

Log-odds metoden av metaanalyse er benyttet i Trafikksikkerhetskåndboken for å beregne sammenlagte effekter fra enkeltstudier i prosentvis endring av antall ulykker/skader, samt 95%-konfidensintervaller rundt effektene.

Enheten i en metaanalyse er et effektestimert. I Trafikksikkerhetskåndboken er et effektestimert et tallmessig anslag på endringer i ulykestall, ulykkesrisiko, skadetall, skadegrad, eller skaderisiko som kan tilskrives det undersøkte tiltaket. I metaanalyse regnes det ut en **sammenlagt effekt**, dvs. et veid gjennomsnitt av effektestimertene.

Det mest vanlige effektmålet i undersøkelser av virkninger av sikkerhetstiltak på antall ulykker, skadde eller drepte er et **oddsforhold**, dvs. den prosentvise endringen av antall ulykker/skadde/drepte. Derfor blir metaanalysene som presenteres i Trafikksikkerhetskåndboken beregnet med log-odds metoden (Fleiss, 1981; Shadish & Haddock, 1994). Effektestimert og sammenlagte effekter beregnes i flere trinn, som er beskrevet mer detaljert i de følgende avsnitt.

En av de viktigste avgjørelsene i metaanalyse er hvilke effektestimert som skal inngå i beregningen av sammenlagte effekter og hvordan disse eventuelt skal grupperes for å beregne flere sammenlagte effekter (for eksempel for ulike skadegrader og ulykkestyper). Dette vurderes i de følgende trinnene av metaanalyse (avsnittene 4.2.3 til 4.2.6).

(1) Beregning av effektestimater: Oddsforhold

Fra hver undersøkelse beregnes ett eller flere anslag på effekten av det aktuelle tiltaket (effektestimater), som regel i form av oddsforhold som sier noe om hvor stort antall ulykker eller ulykkesrisikoen er med et tiltak i forhold til antall ulykker eller ulykkesrisikoen uten tiltaket. Et oddsforhold på for eksempel 0,87 betyr at sjansen² for å ha en ulykke med tiltaket er 0,87 ganger så stor som sjansen uten tiltak, eller at risikoen er 13% lavere med tiltaket enn uten tiltaket. Når oddsforholdet er 1,13 er sjansen for en ulykke 1,13 ganger så stor, eller 13% større, med tiltaket enn uten tiltaket. Et oddsforhold har den generelle formen:

$$\text{Oddsforhold} = \frac{\text{Ulykker i forsøksgr. med tiltak} / \text{Ulykker i forsøksgr. uten tiltak}}{\text{Ulykker i kontrollgr. med tiltak} / \text{Ulykker i kontrollgr. uten tiltak}}$$

Forsøks- og kontrollgruppe er, ideelt sett mest mulig like, bortsett fra at tiltaket settes inn kun i forsøksgruppen. I før-etterstudier er «Forsøksgruppe med tiltak» og «Forsøksgruppe uten tiltak» synonym med «Før tiltaket er satt inn» og «Etter at tiltaket er satt inn». I før-etterstudier uten kontrollgruppe kan «Kontrollgruppe» i formelen over være et eksponeringsmål. I noen undersøkelser er oddsforhold beregnet basert på resultater av multivariate modeller. Effektestimaterne kan da tolkes på samme måte som vanlige oddsforhold.

(2) Beregning av statistiske vekter

For hvert effektestimater beregnes en statistisk vekt som brukes i beregningen av vektete gjennomsnitt. Vekten øker med synkende usikkerhet knyttet til effektestimaterne. Dermed får effektestimater som er basert på mange ulykker som regel større vekter enn effektestimater som er basert på få ulykker.

Beregningen av statistiske vekter som er beskrevet i dette avsnittet gjelder metaanalyse med en såkalt felleeffekts (FE) modell. Hva dette innebærer og alternative måter å beregne statistiske vekter på, er beskrevet i neste avsnitt (under nr. (3)).

Den statistiske vekten er invers til variansen av den naturlige logaritmen til effektestimaterne. Dette fordi oddsforhold ikke er symmetriske og dermed ikke normalfordelte (oddsforhold kan for eksempel ikke bli mindre enn 0, men de kan teoretisk bli uendelig store). Variansen er et meningsfylt mål kun når man kan forutsette en normalfordeling.

I en før-etterstudie med kontrollgruppe beregnes variansen til logaritmen til et oddsforhold som:

$$\text{Varians} = \sum_i \frac{1}{\text{Antall}_i}$$

Hvor Antall_i er antall ulykker/skader/drepte i gruppe i (forsøks- / kontrollgruppe) før og etter at tiltaket er satt inn. Den statistiske vekten er:

$$\text{Statistisk vekt} = \frac{1}{\text{Varians}}$$

² Ordet sjanse er benyttet istedenfor risiko fordi oddsforhold er forhold mellom to odds (beste oversettelse for odds er sjanse) hvor hver av disse beregnes som forhold mellom antall ulykker/skader med vs. uten tiltak; risiko beregnes som antall ulykker/skader per enhet av et eksponeringsmål (for eksempel antall ulykker per mill. kjøretøykilometer). Ved terningkast beregnes oddsen (sjansen) for å få for eksempel en sekser som antall seksere delt på summen av antall enere til antall femmere, mens «risikoen» for å få en sekser beregnes som antall seksere per totalt antall terningkast.

Når antall ulykker som inngår i et resultat, ikke er oppgitt kan den statistiske vekten beregnes ved hjelp av konfidensintervallet (dersom dette er oppgitt eller lar seg beregne for eksempel med hjelp av standardavvik, t- eller p-verdi). Dette gjelder også for effektestimater som er beregnet basert på statistiske regresjonsmodeller. Variansen til slike effektestimater avhenger ikke bare av antall ulykker som inngår i modellen, men også av hvor godt variabelen og regresjonsmodellen predikerer ulykkesrisiko.

Vekten beregnes i slike tilfeller som funksjon av effektestimater og den nedre (eller øvre) grensen av et 95% konfidensintervall:

$$\text{Statistisk vekt} = \left(\frac{1,96}{\ln(\text{Effekt}) - \ln(KI_{\text{Ned}})} \right)^2$$

Hvor Effekt er effektestimater (oddsforhold) KI_{Ned} er nedre grense av konfidensintervallet (se nedenfor under punkt (5)).

(3) Valg av modell

Metaanalysen kan utføres med to forskjellige modeller:

Felles effekt (fixed effects, FE): Denne modellen forutsetter at alle resultater som inngår i analysen representerer den samme (felles) effekten. Denne forutsetningen kan testes ved å beregne heterogeniteten i effektestimater: Er denne veldig liten (eller ikke statistisk signifikant), kan man forutsette at effektestimater representerer en felles effekt (strengt tatt kan man forutsette dette kun når det ikke finnes heterogenitet).

Når det finnes systematisk varians i effektestimater (heterogenitet), vil man med en FE modell underestimere usikkerheten, dvs. at konfidensintervallet vil være for liten. De statistiske vektene beregnes i FE modellen som beskrevet i det forrige avsnittet (under nr. (2)).

Variabel effekt (random effects, RE): Denne modellen kan og bør benyttes når det er (statistisk signifikant) heterogenitet i resultatene, dvs. når det finnes systematisk variasjon i effektestimater. Slik systematisk variasjon kan for eksempel skyldes at det er systematiske forskjeller mellom undersøkelsene.

De statistiske vektene beregnes i RE modellen som en funksjon av vekten i en FE modell og et mål på heterogeniteten. Beregningen av statistiske vektene med RE modellen er nærmere beskrevet i Christensen (2003). Forskjeller mellom store og små statistiske vektene i RE modellen er mindre, jo mer heterogenitet det er i resultatene. Store studier får mindre vekt med en RE modell enn med FE modell når det er heterogenitet. Sammenlignet med resultatene fra en FE modell, har resultatene fra en RE modell:

- Større konfidensintervaller: Dette er en fordel fordi man unngår å underestimere konfidensintervallens størrelse.
- Forskjellige sammenlagte effekter: Dette kan være en ulempe dersom det er en systematisk skjevhet i resultatene, for eksempel publikasjonsskjevhet. Den sammenlagte effekten vil da trekkes nærmere resultatene fra de minste studiene, noe som ville forsterke effekten av skjevheten (Sterne et al., 2011; Thompson & Pocock, 1991).

Ingen av de to modellene kan ta hensyn til kvaliteten på de enkelte studiene (Thompson & Pocock, 1991). Hvis for eksempel en av studiene er veldig stor og metodisk veldig dårlig, vil denne likevel få en større statistisk vekt (større i FE enn i RE metaanalyse) enn de øvrige studiene og dermed har relativt stor innflytelse på den sammenlagte effekten. For å undersøke effekten av og eventuelt ta hensyn til metodisk kvalitet, kan man for eksempel lage subgruppeanalyser, eller man kan definere visse forsøksopplegg som generelt uegnet (dette gjelder især enkle før-etter eller med-uten studier uten kontrollgruppe eller annen form for kontroll for forstyrrende variabler).

Når det *ikke er heterogenitet* i effektestimaterne, vil resultatene fra en FE og RE modell være identiske. Forskjellene mellom resultatene fra FE og RE modellen blir større, jo større heterogeniteten er. De aller fleste metaanalyser som er presentert i Trafikksikkerhåndboken, er derfor beregnet med RE metoden. FE modellen er kun brukt når det er for få effektestimater for å kunne beregne statistiske vekter med RE modellen.

(4) Beregning av vektet gjennomsnitt

Sammenlagte effekter beregnes i log-odds metaanalyse ved å beregne det vektete gjennomsnittet av logaritmene av oddsforholdene:

$$\text{Sammenlagt effekt} = e^{\left(\frac{\sum_i \text{vekt}_i \cdot \ln(\text{effekt}_i)}{\sum_i \text{vekt}_i}\right)}$$

Hvor effekt_i er oddsforhold i og vekt_i er den statistiske vekten (beregnet med FE eller RE modellen) til oddsforhold i .

(5) Beregning av konfidensintervall og prediksjonsintervall

Konfidensintervaller rundt de sammenlagte effektene viser hvor stor statistisk usikkerhet som er knyttet til resultatet. Som regel beregnes 95% konfidensintervaller, dvs. at intervallet i 95% av alle tilfellene vil inneholde den sanne virkningen av et tiltak. Hvis «ingen virkning» ligger utenfor konfidensintervallet sier man at effekten er statistisk signifikant.

Den øvre grensen for konfidensintervallet beregnes som:

$$KI_{\text{øvre}} = e^{\left(\frac{\sum_i \text{vekt}_i \cdot \ln(\text{effekt}_i)}{\sum_i \text{vekt}_i} + \frac{1,96}{\sqrt{\sum_i \text{vekt}_i}}\right)}$$

Den nedre grensen beregnes tilsvarende, med en minus (istedenfor pluss) mellom de to brøkene i stor parentes.

I tillegg til konfidensintervall anbefales ofte å beregne såkalte **prediksjonsintervaller** i RE-metaanalyse (Graham & Moran, 2012; Higgins et al., 2009; Riley et al., 2011).

Prediksjonsintervaller kan beregnes for sammenlagte effekter som er basert på minst tre effektestimater (Riley et al., 2011).

I motsetning til konfidensintervaller som kun viser usikkerheten i den sammenlagte effekten, tar prediksjonsintervaller hensyn både usikkerheten og variasjonen av de underliggende effektene mellom studiene (Langan et al., 2012; Riley et al., 2011). Når det er stor heterogenitet i effektestimaterne, kan prediksjonsintervallene bli svært store (Graham & Moran, 2012). Det samme gjelder prediksjonsintervaller som er basert på få studier (Riley et al., 2011).

Tolkningen av prediksjonsintervaller er at intervallene viser i hvilket intervall resultater av framtidige studier kan forventes å ligge (Graham & Moran, 2012).

Higgins et al. (2009) viser med hjelp av ett eksempel at den samme sammenlagte effekten med samme konfidensintervall kan ha vidt forskjellige tolkninger, avhengig av fordelingen av de enkelte effektestimater. I eksempelet er den ene sammenlagte effekten basert på åtte effektestimater uten statistisk signifikant heterogenitet, mens den andre er basert på 80 effektestimater med stor heterogenitet. Til tross for at konfidensintervallene er like store for de to sammenlagte effektene, kan den andre ikke tolkes slik at effekter av framtidige studier kan forventes å ligge innenfor konfidensintervallet. Her vil et prediksjonsintervall gi et langt mer realistisk inntrykk av hvor store effekter som kan forventes i potensielle framtidige studier.

4.2.3 Vurderingen av heterogenitet og konsistens

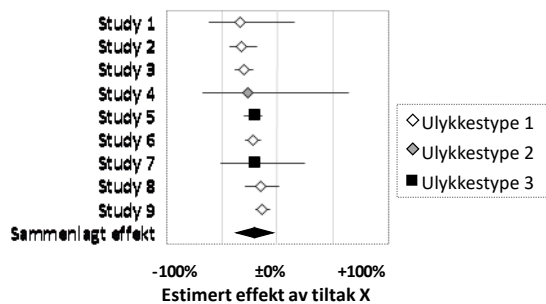
Når enkeltresultatene som inngår i en metaanalyse, er heterogene (ikke konsistente), tyder dette på at det ikke finnes én felles underliggende effekt, men at enkeltresultatene representerer ulike «sanne» effekter, for eksempel av ulike typer tiltak, under ulike forhold, eller for ulike typer ulykker eller skader. Man kan ta hensyn til heterogenitet ved å benytte en random effects (RE) metaanalyse, da vil man bl.a. unngå å beregne for små konfidensintervall. Man kan også prøve å forklare heterogenitet ved å undersøke effekten av mulige moderatorvariabler. I noen tilfeller kan den beste løsningen være å la være å beregne en sammenlagt effekt.

Informasjon om konsistens i effektestimater som inngår i en metaanalyse er viktig for å kunne avgjøre hvorvidt resultatene kan generaliseres (Higgins et al., 2003; Guyatt et al., 2011G). Med konsistens menes fravær av heterogenitet, dvs. hvorvidt effektestimater kan antas å ha en felles underliggende effekt. Konsistens kan vurderes ut fra (Guyatt et al., 2011G):

- Hvor like de enkelte effektestimater er
- I hvilken grad konfidensintervallene overlapper
- Statistiske kriterier som tester av heterogenitet og I^2 .

Forest plot

I presentasjonen av resultater fra metaanalyse er det vanlig at man innledningsvis presenterer resultatene fra de enkelte studiene i en såkalt forest plot (Anzures-Cabrera & Higgins, 2010). En forest plot viser alle effektestimater med konfidensintervallene som vist i eksempelet i figur 10. I eksempelet er resultatene sortert etter størrelse på effektene; resultatene kan også sorteres for eksempel alfabetisk etter forfatter eller kronologisk. Effektestimater i eksempelet vises med ulike figurer, avhengig av hvilken type skade de gjelder, men det er også mulig å la størrelsen på symbolene variere med variansen (større symboler for effektestimater med mindre varians). Videre er det vanlig å vise den sammenlagte effekten nederst i diagrammet (konfidensintervallet til den sammenlagte effekten i eksempelet er så liten at det ikke vises).



Figur 10: Forest plot, eksempel.

I en empirisk studie hvor forskere med mye erfaring i metaanalyse vurderer graden av heterogenitet i ulike typer diagram (Bax et al., 2008), var forest plots (og diagrammer som viser standardiserte residualer) de diagrammene hvor det var størst sammenheng mellom forskernes vurderinger av heterogenitet og faktisk heterogenitet. Diagrammene var basert på simulerte metaanalyse med ulike grader av heterogenitet.

Fordelingen til effektestimaterne og statistiske vekter: Traktdiagrammer

Fordelingen til effektestimaterne kan si noe om hvorvidt man kan anta at det finnes én felles effekt som ligger til grunn i alle studiene og om det kan være systematiske skjevheter i effektestimaterne. Heterogenitet og skjevheter kan også undersøkes med statistiske metoder som er nærmere forklart nedenfor.

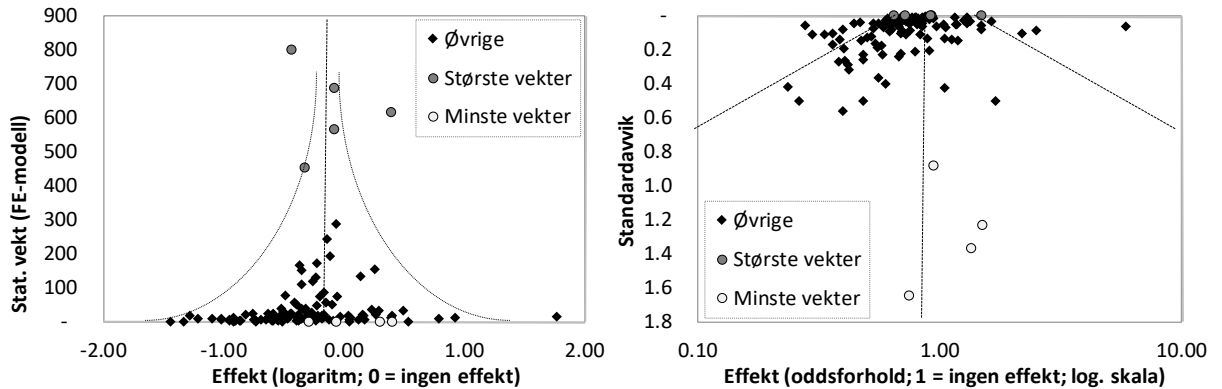
Med fordeling menes sammenhengen mellom effektestimatenes størrelse og statistiske vekter eller standardavvik som vist i eksemplet i figur 11 i to traktdiagrammer (engelsk funnel plots). De to diagrammene i figuren er basert på det samme (fiktive) datasettet. Til venstre vises et klassisk traktdiagram med den statistiske vekten på Y-aksen, slik som det har vært brukt tidligere i Trafikksikkerhetshåndboken (Elvik et al., 2009A). Til høyre i figuren vises et traktdiagram etter Sterne et al. (2011) med standardavviket på Y-aksen. Traktdiagrammer som er basert på andre mål en standardavvik eller statistisk vekt (for eksempel antall enheter i studiene eller varians), anbefales ikke (Peters et al., 2008; Sterne & Egger, 2001).

Dersom alle effektestimaterne har en felles «underliggende» fordeling, kan man forvente at fordelingen ligner en trakt, dvs. at:

- Fordelingen er symmetrisk
- Effektestimaterne med de største vektene ligger i midten av fordelingen, mens effektestimaterne på utsiden av fordelingen (langs X-aksen) har stadig mindre statistiske vekter (dvs. større standardavvik).

De to typene traktdiagram i figur 11 gir forskjellig informasjon om fordelingene:

Statistisk vekt på Y-akse	Standardavvik på Y-akse (omvendt)
<ul style="list-style-type: none"> ▪ Spredningen vises i hovedsak blant effektestimaterne med de <u>største</u> vektene, mens de med de minste vektene ligger relativt tett pakket på bunnen ▪ Konfidensintervallinjene er buet ▪ Anbefales av Sterne & Egger (2001) i metaanalyse når enkeltstudiene har veldig ulik størrelse 	<ul style="list-style-type: none"> ▪ Spredningen vises i hovedsak blant effektestimaterne med de <u>minste</u> vektene; det er disse som er mest relevante i vurderingen av publikasjonsskjevhet ▪ Konfidensintervallinjene er rette ▪ Anbefales av Sterne & Egger (2001) og Sterne et al. (2011)



Figur 11: Resultatfordelingsdiagram for et fiktivt tiltak med sammenlagt effekt (stiplet loddrett linje); klassisk trakttdiagram (t.v. med trakt) og trakttdiagram etter Sterne et al. (2005; t.h.).

«Trakt»-konturene i de to diagrammene i figur 11 viser områdene som de fleste effektestimaterne kan forventes å ligge innenfor ved en symmetrisk fordeling. Traktene tegnes alltid slik at de er symmetriske. I trakttdiagrammer etter Sterne et al. (2005) skal traktene tegnes slik at 95% av effektestimaterne ligger innenfor trakten under forutsetning av at fordelingen er symmetrisk. I begge figurene kan traktene hjelpe med å vurdere (visuelt) hvorvidt fordelingene er symmetriske.

Ved å studere slike trakttdiagrammer kan man danne seg en oppfatning om hvor informativt et veid gjennomsnittresultat kan være. Trakttdiagrammer gir følgende informasjon om egenskaper ved fordelingen av resultater, som er nærmere beskrevet i det følgende:

- Modalitet og spredning
- Symmetri
- Avvikfølsomhet.

Modalitet og spredning: Med modalitet menes hvor mange «humper» eller topper fordelingen av resultater har. Generelt sett kan man finne følgende typer fordeling etter modalitet og spredning:

- **Unimodale fordelinger:** Slike fordelinger (som for eksempel den i figur 11) tyder på at effektestimaterne kan ha en felles underliggende fordeling. Som regel vil det være meningsfylt å beregne en sammenlagt effekt. Likevel kan slike fordelinger være heterogene (mer om dette nedenfor).
- **Bimodale fordelinger:** Fordelinger med to «humper» kan oppstå når effektestimaterne kommer fra to underliggende fordelinger med forskjellige midtpunkter. I slike tilfeller kan man forsøke å dele opp datamaterialet i to unimodale fordelinger (og følgelig også beregne to sammenlagte effekter).
- **Fordelinger uten midtpunkt:** Slike fordelinger kan skyldes at det ikke finnes noen felles underliggende fordeling, at det finnes flere underliggende fordelinger eller at man har for få datapunkter. I de første to tilfellene vil det ikke være meningsfylt å beregne en sammenlagt effekt. Ved få datapunkter kan en sammenlagt effekt være vanskelig å tolke og prediksjonsintervallet kan være veldig stor, med mindre alle effektestimaterne viser omtrent det samme.

Symmetri: Symmetrien kan vurderes visuelt ved hjelp av traktogrammer som vist i figur 11 og med hjelp av statistiske metoder. Også forskjeller mellom det uvektede gjennomsnittet og den sammenlagte effekten kan tyde på at fordelingen er usymmetrisk. Fordelinger som er usymmetriske, kan være påvirket av systematiske skjevheter som for eksempel publikasjonsskjevhet. Slike skjevheter og statistiske metoder for å vurdere (og ev. korrigere for) skjevheter er beskrevet nedenfor i avsnitt 4.2.5.

Avvikfølsomhet: Med avvikfølsomhet (engelsk outlier bias) menes at det finnes enkelte effektestimater, især med store statistiske vekter, som avviker mye fra de øvrige. Hadde det i det venstre traktogrammet i figur 11 vært et effektestimater med en logaritmisk effekt på 1,5 og en statistisk vekt på 900, ville denne effekten ha vært klart avvikende fra de øvrige effektene. Hvordan man kan teste avvikfølsomhet er beskrevet nedenfor i avsnitt 4.2.4.

Statistisk test av heterogenitet og konsistens

En forest plot kan gi en første indikasjon om hvor heterogene effektestimaterne er. Heterogeniteten kan testes statistisk bl.a. med Cochran's Q statistisk (Higgins et al., 2003) eller med I²-statistikken (Borenstein et al., 2017; Moher et al., 2015) som begge er beskrevet i det følgende. Heterogeniteten kan testes både for alle effektestimaterne samlet og for grupper av effektestimater (gruppert for eksempel etter skadegrad og ulykkestype).

Cochran's Q: Den mest vanlige testen for heterogenitet er statistikken Cochran's Q. Denne beregnes som sum av de kvadrerte avvikene mellom de enkelte effektestimaterne og den sammenlagte effekten (FE) hvor effektestimaterne er vektet med de samme vektene som i metaanalysen. P-verdiene er basert på en Chi-kvadrat fordeling med k-1 frihetsgrader hvor k er antall effektestimater (Cochran, 1954; Higgins et al., 2003). Når Cochran's Q er statistisk signifikant, tolkes dette som regel slik at det er heterogenitet i effektestimaterne. For tester av heterogenitet gjelder imidlertid de samme prinsippene som for andre signifikanstester (se avsnitt 3.4). Disse støtter ikke en strikt tolkning av «heterogenitet eller ikke heterogenitet».

Den største ulempen med Cochran's Q-test er at p-verdien ikke kan tolkes som en indikator av hvor stor heterogeniteten er eller om den er praktisk relevant, bl.a. fordi den avhenger av antall effektestimater (Higgins et al., 2003):

- Når det er få studier har testen lite styrke, dvs. at et ikke-signifikant resultat ikke nødvendigvis kan tolkes som fravær av heterogenitet.
- Testen har derimot stor styrke når det er mange studier, slik at et statistisk signifikant resultat ikke nødvendigvis betyr at heterogeniteten er av praktisk betydning.

Når man gjør subgruppeanalyser (hvor man beregner nye sammenlagte effekter basert på grupper av effektestimater), kan ikke-signifikante verdier av Cochran's Q ikke nødvendigvis tolkes slik at heterogeniteten er mindre i en subgruppe enn for alle effektestimaterne. Den ikke-signifikante verdien kan også være et resultat av at det er færre effektestimater.

I²: Som et alternativ til testen for heterogenitet med Cochran's Q har Higgins et al. (2003) utviklet statistikken I². I² beskriver andelen av variansen i effektestimaterne som skyldes heterogenitet (og ikke tilfeldig variasjon), dvs. hvilken andel av variansen som ville bli igjen dersom alle studiene viste den «sanne» underliggende effekten for de respektive studiene.

Når de enkelte effektestimaterne i en metaanalyse vises i en forest plot, kan I² brukes for å vurdere i hvilken grad variansen i effektestimaterne skyldes forskjeller i de «sanne» underliggende effekter i de enkelte studiene (Borenstein et al., 2017).

I^2 beregnes som en funksjon av Cochrans' Q og antall frihetsgrader (antall effektestimater minus én):

$$I^2 = 100\% * \frac{Q - df}{Q}$$

Negative verdier settes lik 0%. Higgins et al. (2003) foreslår følgende tolkning av verdier på I^2 :

- 25%: Lite heterogenitet
- 50%: Middels heterogenitet
- 75%: Stor heterogenitet.

Lite heterogenitet kan tolkes som at alle studiene har den samme eller lignende underliggende effekter. Når det ikke er heterogenitet ($I^2 = 0$), har alle studiene den samme underliggende effekt og en FE modell for metaanalyse vil være hensiktsmessig (Riley et al., 2011). Stor heterogenitet kan tolkes slik at det er store forskjeller mellom de underliggende effektene i de enkelte studiene. Når det er heterogenitet i resultatene, vil en RE-modell være mer hensiktsmessig enn en FE modell.

I motsetning til en relativt vanlig oppfatning sier I^2 *ingenting* om i hvilket intervall effektene av det undersøkte tiltaket varierer; dette intervallet må beregnes som **prediksjonsintervall** (Borenstein et al., 2017; se avsnitt 4.2.2 under punkt (5)).

Fordeler med I^2 (framfor Cochrans' Q) er at verdiene kan sammenlignes mellom metaanalyser som er basert på ulike antall effektestimater eller på ulike typer av resultatvariabler. Derfor kan I^2 benyttes bl.a. for å sammenligne heterogeniteten mellom den sammenlagte effekten som er basert på alle effektestimaterne og sammenlagte effekter som er beregnet for subgrupper.

Ta hensyn til eller forklare heterogenitet?

Det finnes flere muligheter for å håndtere resultater som er heterogene, som ikke er unimodale eller som har så stor spredning at beregning av sammenlagte effekter kan være tvilsom (Christensen, 2003; Higgins et al., 2009):

- **Ikke beregne noen sammenlagt effekt:** Når det er stor heterogenitet som det ikke er mulig å forklare, kan det være den beste løsningen å ikke beregne en sammenlagt effekt. Indikasjoner på en slik fordeling er store forskjeller mellom resultatene fra RE og FE modellen og svært store konfidensintervaller i RE modellen.
- **Bruk av RE metaanalysen istedenfor FE:** Ved bruk av RE metaanalyse unngår man å beregne for små konfidensintervaller ved heterogenitet. Bruk av RE modellen er standard i Trafikksikkerhåndboken, med mindre det foreligger for få effektestimater. Bruk av RE modellen kan derimot ikke legitimere beregningen av en sammenlagt effekt i situasjoner hvor dette ikke er meningsfullt for eksempel fordi de enkelte effektestimaterne har grunnleggende forskjellige underliggende effekter (se punkt over og under).

- **Forklare heterogeniteten:** For å forklare heterogenitet, kan man gjøre **subgruppeanalyser** eller **metaregresjon** som er nærmere beskrevet i avsnitt 4.2.4. Med slike metoder kan man undersøke hvorvidt ulike variabler påvirker den sammenlagte effekten (for eksempel etter undersøkelses metode, skadegrad, ulykkestype, vegtype, tiltaksvariant). Den variabelen (eller de variablene) som man har delt opp etter, er moderatorvariabler (jf. avsnitt 4.2.8). Når subgruppene er mindre heterogene kan det være mest meningsfylt å kun rapportere sammenlagte effekter for hver subgruppe. For eksempel når man har resultater for ulykker med bilister og syklistere og tiltaket som er undersøkt, har forskjellige effekter blant bilister og syklistere, kan en sammenlagt effekt basert på alle ulykkene være meningsløs (fordi det ikke finnes noen trafikantgruppe hvor man vil finne denne effekten). I et slikt tilfelle vil det være bedre å rapportere ett resultat for bilister og ett for syklistere.
- **Effekter som funksjoner:** For noen tiltak gir det best mening å beskrive virkningene som en funksjon. Eksempelvis vil virkningene av politikontroll være en funksjon av omfanget av kontroll. Sammenhengen mellom fart og trafikksikkerhet kan også best beskrives med et sett av funksjoner. I slike tilfeller kan en sammenlagt effekt være misvisende fordi for eksempel «politikontroll vs. ikke politikontroll» kan være helt forskjellige ting (avhengig av omfanget av politikontroll) i ulike studier.

4.2.4 Tilleggsanalyser: Avviksfølsomhet, subgruppeanalyser og metaregresjon

Analysen av avviksfølsomhet kan vise hvorvidt en sammenlagt effekt i metaanalyse er påvirket av enkelte atypiske resultater. Subgruppeanalyser og metaregresjon er metoder for å forklare heterogeniteten ved å undersøke effekten av mulige moderatorvariabler. Moderatorvariabler kan for eksempel være tiltaksvarianter som har ulike effekter på antall ulykker eller skader.

Tilleggsanalyser omfatter sensitivitetsanalyser (avviksfølsomhet), subgruppeanalyser og metaregresjon.

Avviksfølsomhet

Avviksfølsomhet betegner hvor sterkt påvirket gjennomsnittet i en fordeling er av ett enkelt, atypisk resultat (uteligger, eller outlier på engelsk). Som regel vil slike avvikende resultater føre til at det oppdages heterogenitet i resultatene og dermed at resultatet/resultatene med de største statistiske vektene vil få noe reduserte statistiske vekt. Likevel kan resultatene som avviker mye fra de øvrige resultatene og som har stor statistisk vekt, påvirke den sammenlagte effekten.

Avviksfølsomhet kan også testes ved å beregne nye sammenlagte effekter der ett og ett resultat er utelatt. Hvis ingen av disse skiller seg betydelig fra den sammenlagte effekten som er basert på alle effektestimaterne, kan man si at den sammenlagte effekten ikke er påvirket av avvikende effekter.

Det finnes imidlertid ingen objektive kriterier for å skille mellom «normale» og «avvikende» resultater. Resultater som ser ut til å være avvikende, er derfor i

Trafikksikkerhåndboken som regel ikke ekskludert fra analysen, med mindre det finnes spesifikke grunner til det, som for eksempel at metoden er mye dårligere enn i de øvrige studiene eller at det undersøkte tiltaket er forskjellig fra de øvrige undersøkte tiltakene.

Subgruppeanalyser og metaregresjon

Formålet med subgruppeanalyser og metaregresjon er å finne moderatorvariabler, dvs. faktorer som påvirker størrelsen på den sammenlagte effekten. Subgruppeanalyser og metaregresjon er mest relevante når:

- Det er heterogenitet i den sammenlagte effekten
- Når fordelingen av effektestimater ikke tyder på at det finnes én felles effekt
- Når man har en konkret hypotese om ulike effekter i ulike grupper; i Trafikksikkerhåndboken gjelder dette især skadegrader og i mange tilfeller også ulykkestyper og/eller trafikantgrupper.

Når heterogeniteten er liten (eller ikke statistisk signifikant), vil en subgruppeanalyse eller metaregresjon som regel ikke avdekke relevante moderatorvariabler.

Typiske moderatorvariabler som undersøkes i metaanalyser som gjøres i forbindelse med Trafikksikkerhåndboken, er:

- Skadegrad
- Ulykkestype
- Trafikantgruppe
- Vegtype eller områdetype (ofte spredt- vs. tettbygd strøk)
- Tiltaksvariant
- Type studie, for eksempel studier med vs. uten kontroll for regresjonseffekter
- Publikasjonsår.

Ulykkesrisikoen uten tiltak kan også være en relevant moderatorvariabel i mange sammenhenger. For eksempel kan man tenke seg at kjøretøytiltak har størst effekt i typer kjøretøy eller blant førere som i utgangspunktet har veldig høy risiko for ulykker eller skader. Hvis man undersøker sammenhengen mellom risiko uten tiltak og tiltakseffekt, kan imidlertid resultatene være svært misvisende (Thompson et al., 1997).

Subgruppeanalyse: Ved subgruppeanalyser deler man effektestimater inn i grupper, for eksempel etter skadegrad, type skade, type tiltak, vegtype, persongruppe. Dersom man finner forskjeller i den sammenlagte effekten mellom grupper og når heterogeniteten i gruppene er redusert, tyder dette på at den respektive variabelen er en relevant moderator.

En risiko ved subgruppeanalyser er at man grupperer enkeltstudiene etter kriterier som gjør at de enkelte sammenlagte effektene ikke har (statistisk signifikant) heterogenitet, men uten at resultatene ville la seg replisere (Elvik, 2005B). En slik framgangsmåte ligner på såkalt **data dredging** (eller «fisking») etter forskjeller eller effekter). Det er derfor viktig å ha gode begrunnelser for valg av kriterier. For eksempel er skadegrad et kriterium som ofte vil være meningsfylt for å danne subgrupper fordi mange trafikksikkerhetstiltak har ulike effekter på ulike skadegrader. Danner man subgrupper etter flere kriterier, bør inndelingen være konsistent og i overensstemmelse med begrunnede antakelser om sammenhengen mellom kriteriene og effekt.

Meta-regresjon: Meta-regresjon er en multivariat analysemetode som tester simultant effekten av flere mulige moderatorvariabler, dvs. om effektestimater er systematisk forskjellige avhengig av verdien på en eller flere moderatorvariabler (Thompson & Sharp, 1999). Den avhengige variabelen i meta-regresjon er effektvariabelen, mens de potensielle moderatorvariablene er prediktorvariablene.

Fordelen med meta-regresjon framfor subgruppeanalyser er at man kan teste effekten av flere variabler simultant. Moderatorvariablene er alltid egenskaper ved de enkelte effektestimater (ikke egenskaper ved enhetene som inngår i de enkelte effektestimater). Når effekten av et tiltak for eksempel er undersøkt på ulike typer veg, kan effekten av type veg undersøkes med meta-regresjon kun hvis det foreligger effektestimater som gjelder ulike typer veg. Dersom effektene på ulike typer veg er slått sammen i alle studiene, kan effekten av type veg ikke undersøkes med meta-regresjon.

Som i subgruppeanalyser er det stor fare for å «*overtilpasse*» (overfit) regresjonsmodellene, dvs. å lage modeller (gjerne med mange prediktorvariabler) som er veldig «gode» i den forstand at de forklarer en stor andel av variansen i effektestimater og at prediktorvariablene i modellen er statistisk signifikante. Risikoen med for detaljerte modeller er at de er veldig godt tilpasset det aktuelle datasettet, men at de ikke vil la seg replisere. En generell anbefaling er derfor at man ikke bruker mer enn én prediktorvariabel per fem eller 10 effektestimater som inngår i regresjonsmodellen (van Houwelingen et al., 2002). Hvis antall prediktorvariabler er lik antall effektestimater minus én, vil man kunne forklare all varians mellom effektestimater (Thompson & Sharp, 1999), men dette betyr ikke at prediktorvariablene er meningsfylte eller at resultatene av meta-regresjonen vil la seg replisere.

4.2.5 Analyse av skjevheter

En vanlig og alvorlig form for skjevhet i metaanalyse er publikasjonsskjevhet. Denne oppstår når empiriske studier som finner ikke-signifikante eller «uønskede» resultater, publiseres i mindre grad enn studier med signifikante effekter i «ønsket» retning. Den mest vanlige måten å undersøke publikasjonsskjevhet på er en analyse av traktogrammer. Det er mulig å korrigere for publikasjonsskjevhet med den såkalte trim-and-fill metoden. Ulemper ved slike metoder er at de er lite treffsikre. De er heller ikke i stand til å skille mellom publikasjonsskjevhet og andre former for skjevhet.

Med skjevhet i fordelingen av effektestimater i metaanalyse menes at det er en sammenheng mellom enkeltresultatets størrelse og presisjon. Skjevhet i fordelingen av effektestimater kan vurderes visuelt ved hjelp av traktogrammer (se ovenfor, figur 11, avsnitt 4.2.3). Uten skjevhet er fordelingen symmetrisk, det er ingen sammenheng og det finnes omtrent like mange effekter med (stort) positivt avvik fra gjennomsnitt som det finnes effekter med (stort) negativt avvik.

En indikator for symmetri i en fordeling er avvik mellom vektet og uvektet gjennomsnitt, median (verdien som deler fordelingen i to like store deler) og modalverdi (høyeste punkt i fordelingen) (Elvik, 1998). I en fordeling som er helt symmetrisk vil disse verdiene være identiske. Det er også mulig å teste symmetri statistisk. Dette gjøres i forbindelse med testing av publikasjonsskjevhet som er forklart nedenfor. Hvis en fordeling er skjev, kan den sammenlagte effekten gi et misvisende bilde av hvor hovedtyngden av resultater ligger da de fleste enkeltresultatene vil ligge enten over eller under gjennomsnittet (Elvik, 1998). Fordelingen i figur 11 er ikke helt symmetrisk.

Å vurdere fordelingen av effektestimater i metaanalyse er relevant fordi usymmetriske fordelinger som regel skyldes en av de følgende faktorene (Anzures-Cabrera & Higgins, 2010; Moher et al., 2015):

- Publikasjonsskjevhet
- Selektiv rapportering i enkeltstudier
- Systematiske forskjeller mellom resultatene fra store og små studier

Disse tre faktorene er nærmere beskrevet i de følgende avsnittene.

Publikasjonsskjevhet

Publikasjonsskjevhet innebærer at undersøkelser som finner «uønskede» eller ikke-signifikante resultater ikke publiseres. At resultater er «uønsket» kan bl.a. skyldes at oppdragsgivere som finner ut at tiltak de selv har vært ansvarlige for, ikke reduserer ulykker. Ikke-signifikante resultater anses ofte som uinteressante av reviewere i fagfelleverderte tidsskrifter. Hvorvidt dette er berettiget (det er ikke det), er diskutert i avsnitt 3.4. Slike studier kan forbli upublisert, eller publiseres gjennom mindre tilgjengelige kanaler, for eksempel i rapporter i andre språk enn engelsk, slik at de lett kan bli oversatt i litteratursøk (Guyatt et al., 2011C). Når studier benytter multivariate modeller, er det relativt vanlig at kun variabler med signifikant effekt inkluderes i modellen. For ikke-signifikante variabler foreligger i slike tilfeller ingen informasjon om verken effektens størrelse eller varians, slik at det ikke er mulig å inkludere dem i en metaanalyse.

Publikasjonsskjevhet fører som regel til at den sammenlagte effekten er overestimert i forhold til hvis alle potensielle undersøkelser hadde vært tilgjengelige for metaanalysen.

Det finnes ulike metoder for å vurdere, teste og korrigere for publikasjonsskjevhet:

- Visuell inspeksjon av traktdiagrammer
- Sammenligning av resultater fra RE vs. FE metaanalyse
- Regresjonsbaserte metoder
- Trim-and-fill analyse.

Visuell inspeksjon av traktdiagrammer: Traktdiagrammer kan gi et inntrykk av om fordelingen av effektresultatene er symmetrisk eller skjev. En skjev fordeling indikerer at resultatene *kan* være påvirket av publikasjonsskjevhet (men skjevheten kan også skyldes andre faktorer). Publikasjonsskjevhet medfører som regel en skjev fordeling med en mangel på undersøkelser på den siden av fordelingen som viser virkninger i «ikke-ønsket» retning, især blant effektestimater med små statistiske vekter. Sistnevnte skyldes at virkninger i små undersøkelser må være større for å bli signifikante sammenlignet med virkninger i store undersøkelser. Traktdiagrammer er mest informative når de er basert på minst 10 studier; traktdiagrammer basert på færre studier bør unngås (Anzures-Cabrera & Higgins, 2010).

Bax et al (2008) viser at traktdiagrammer har begrenset verdi for å identifisere publikasjonsskjevhet. Forskere med mye erfaring i metaanalyse klarte i denne studien bare i veldig liten grad å vurdere graden av publikasjonsskjevhet i traktdiagrammer (ved visuell inspeksjon). Traktdiagrammene var basert på et stort antall simulerte metaanalyser med ulike grader av publikasjonsskjevhet, dvs. at den «sanne» graden av publikasjonsskjevhet var kjent på forhånd (men ikke av testpersonene). For andre typer diagrammer var det ikke større sammenheng mellom publikasjonsskjevhet og forskernes vurderinger.

FE vs. RE metaanalyse: Når resultatene fra en RE metaanalyse er mer fordelaktige enn resultatene fra en FE metaanalyse, kan dette tyde på at små studier har funnet større effekter i «ønsket» retning, fordi disse får relativt større vekt i RE metaanalyse enn i FE metaanalyse. Sterne et al. (2000) anbefaler i slike tilfeller å undersøke heterogeniteten.

Regresjonsbaserte metoder: Slike metoder undersøker sammenhengen mellom effektens størrelse og standardavviket (Egger et al., 1997). I log-odds metaanalyse ville man beregne sammenhengen mellom den naturlige logaritmen av effekten (log-odds) og standardavviket. I skjeve fordelinger vil man finne en sammenheng.

En generell ulempe med slike metoder er at de har lav styrke, dvs. at selv om et resultat ikke tyder på at det er skjevhet, kan det likevel være skjevhet. Dette gjelder især metaanalyser som er basert på få studier og hvor skjevheten ikke er stor. Metodene kan også gi mange falske positive resultater, dvs. at resultatene tyder på skjevhet selv om det ikke er noe skjevhet, især når effekten er stor og når de enkelte studiene er små (Sterne et al., 2000).

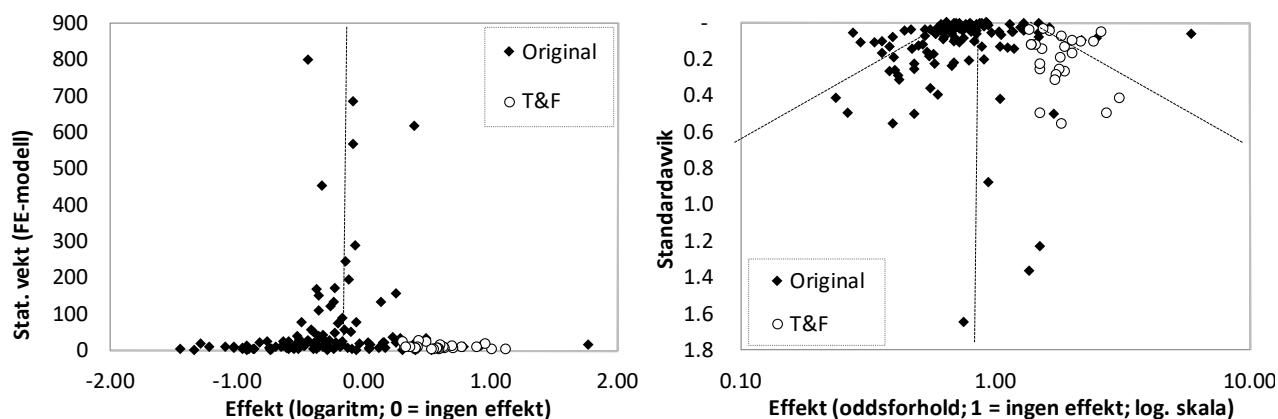
I log-odds metaanalyse (som benyttes i Trafikksikkerhetshåndboken) kommer i tillegg at det finnes en sammenheng mellom størrelsen på oddsforhold og deres standardavvik, selv når fordelingen ikke er skjev (Sterne et al., 2011). Dette kan også føre til falske positive resultater (Rücker et al., 2008).

Det er utviklet alternative regresjonsbaserte metoder som skal unngå denne ulempen (bl.a. Harbord et al., 2006; Peters et al., 2006), og disse anbefales å bruke av Sterne et al. (2011). Likevel kan begge metodene gi falske positive resultater når det er stor heterogenitet. Metoden til Peters et al. (2006) er basert på størrelsen på utvalget i de enkelte studiene og er derfor ikke mulig å bruke i mange meta-analyser i Trafikksikkerhetshåndboken hvor en del resultater er basert på multivariate studier. Metoden til Rücker et al. (2008) unngår dette problemet, men resultatene kan være vanskelige å tolke.

Trim-and-fill analyse: For å teste og korrigere for publikasjonsskjevhet kan man bruke den såkalte «trim-and-fill» (T&F) metoden (Anzures-Cabrera & Higgins, 2010; Christensen, 2003; Duval & Tweedie, 2000). Med trim-and-fill metoden beregnes simulerte nye effektestimater og vektorer, slik at fordelingen i traktogrammet blir symmetrisk, og det beregnes en ny sammenlagt effekt basert på både de originale og de simulerte effektestimaterne. Denne nye effekten antas å ligge nærmere den effekten som man ville ha funnet hvis det ikke hadde vært publikasjonsskjevhet.

Metoden er basert på symmetrien i traktogrammet og kan derfor ikke skille mellom ulike typer av skjevhet. Med andre ord, hvis skjevheten ikke skyldes publikasjonsskjevhet (men for eksempel en blanding av effekter på personskade- og dødsulykker), kan effekten som er «korrigert» for publikasjonsskjevhet, være svært misvisende.

Figur 12 viser resultatene av en trim-and-fill analyse av effektestimaterne i figur 11. De simulerte effektestimaterne fører til at fordelingen er blitt mer symmetrisk og at den nye sammenlagte effekten er mindre (nærmere null) enn den originale sammenlagte effekten. Den sammenlagte effekten som er beregnet med trim-and-fill, ligger også nærmere midten av fordelingen.



Figur 12: Resultatfordelingsdiagram for effektestimaterne fra figur 11 med simulerte nye effektestimater («T&F») som er beregnet med trim-and-fill metoden.

Anbefalinger for å undersøke publikasjonsskjevhet

Generelt anbefales det å undersøke hvorvidt resultatene fra metaanalysen kan være påvirket av publikasjonsskjevhet (eller andre former for skjevhet) (Egger et al., 1997; Harbord et al., 2006; Ioannidis & Trikalinos, 2007; Sterne et al., 2000, 2011; Sutton et al., 2000).

Ved tolkningen av resultatene er det viktig å ta hensyn til at skjevhet ikke alltid er et resultat av publikasjonsskjevhet, men at det kan finnes andre mulige forklaringer som også bør vurderes (Peters et al., 2008; Sterne et al., 2011). En mulig forklaring er beskrevet nedenfor i avsnittet «Systematiske forskjeller mellom resultatene fra store og små studier».

Ved bruk av formelle tester, er det i tillegg viktig å ta hensyn til at slike tester kan ha svak teststyrke og gi falske positive resultater. Sterne et al. (2011) anbefaler ved log-odds metaanalyse testene etter Harbord et al. (2006) og Peters et al. (2006), men disse kan ha falske positive resultater når det er stor heterogenitet. Testen etter Rücker et al. (2008) kan også brukes når det er stor heterogenitet, men resultatene kan være vanskelige å tolke.

For å teste publikasjonsskjevhet (og andre former for skjevhet) bør ifølge Ioannidis & Trikalinos (2007) de følgende forutsetningene være oppfylt:

- (1) Det er ikke stor heterogenitet i resultatene; mer konkret bør ikke I^2 (se avsnitt 4.2.3) være lik eller større enn 50%; alternativt skal Cochrans Q ikke være statistisk signifikant på 10%-nivået (tosidig test).
- (2) Antall studier eller effektestimater må være minst 10 (dette gjelder også ifølge Sterne et al., 2011)
- (3) Standardavviket bør være forskjellig mellom de enkelte effektestimaterne (gjelder også ifølge Sterne et al., 2011); mer konkret bør det største standardavviket være minst fire ganger så stor som det minste
- (4) Minst ett av de enkelte effektestimaterne bør være statistisk signifikant (dette fordi publikasjonsskjevhet antas å være tilstede når fortrinnsvis statistisk signifikante resultater publiseres).

Selektiv rapportering i enkeltstudier

Med selektiv rapportering i enkeltstudier menes manglende rapportering av enkelte resultater eller ufullstendig rapportering av informasjon om enkelte resultater (Chan et al., 2004; Moher et al., 2015). Slik selektiv rapportering kan prinsipielt medføre de samme problemene som publikasjonsskjevhet (Sterne et al., 2011).

I studiene som inngår i metaanalysene i Trafikksikkerhetskåndboken er det i hovedsak de følgende to formene for selektiv rapportering som kan forekomme:

- **Manglende rapportering av resultater:** Problemet oppstår (i forbindelse med Trafikksikkerhetskåndboken) som regel i multivariate studier hvor kun variabler med statistisk signifikante effekter inngår i de multivariate modellene. I metaanalysen er det ikke mulig å inkludere slike resultater; man kunne prinsipielt sette effekten lik null, men det vil fortsatt mangle informasjon om den statistiske vekten. I metaanalyser som gjøres i forbindelse med Trafikksikkerhetskåndboken opplyses om slike resultater. Når studier rapporterer for eksempel et statistisk signifikant resultat av én tiltaksvariant, trafikantgruppe eller lignende, men ikke et ikke-signifikant resultat for en annen tiltaksvariant, trafikantgruppe eller lignende, blir som regel ingen av resultatene med i metaanalysen.

Dette problemet kan kun avdekkes dersom studien gir informasjon om at for eksempel en faktor X er undersøkt, når det ikke er rapportert noe tilsvarende resultat. Det er derfor viktig å sjekke om det er rapportert resultater for alle variabler som er målt i studien (Kirkham et al., 2010).

- **Ufullstendig rapportering av informasjon om resultater:** Når studier rapporterer fullstendig informasjon kun for utvalgte resultater (for eksempel for statistisk signifikante resultater) men ikke for andre resultater, oppstår de samme problemene som ved manglende rapportering av resultater.

Systematiske forskjeller mellom resultatene fra store og små studier

Skjevhet i fordelingen av effektestimater kan oppstå som følge av at:

- Enkelte grupper av effektestimater har ulike sammenlagte effekter
- Det er systematiske forskjeller i effektstørrelsen mellom små og store studier.

Et relativt vanlig eksempel som kombinerer begge effektene, er fordelingen av effektestimater som er basert på ulykker med ulike skadegrader:

- Dødsulykker er langt sjeldnere enn mindre alvorlige ulykker, noe som fører til at effekter på dødsulykker ofte har langt større varians (mindre statistiske vektorer) enn effekter på mindre alvorlige ulykker.
- Virkningen på dødsulykker av trafikksikkerhetstiltak er ofte langt større enn på mindre alvorlige ulykker (især tiltak som påvirker fart).

Til sammen fører dette til at traktogrammer som viser effektestimater for både dødsulykker og mindre alvorlige ulykker, kan være usymmetriske med en «mangel» på effektestimater som viser små ulykkesreduksjoner og små vektorer. Dette skyldes at store effektestimater med små statistiske vektorer er overrepresentert på grunn av dødsulykkene. En slik resultatfordeling kan lett tolkes slik at resultatene er påvirket av publikasjonsskjevhet. I slike tilfeller vil en T&F analyse føre til misvisende resultater og det vil være mer korrekt å dele opp datamaterialet etter skadegrad før man eventuelt gjør en T&F analyse. Dette er gjort for de aller fleste tiltak i trafikksikkerhetsåndboken hvor det foreligger effektestimater for ulike skadegrader.

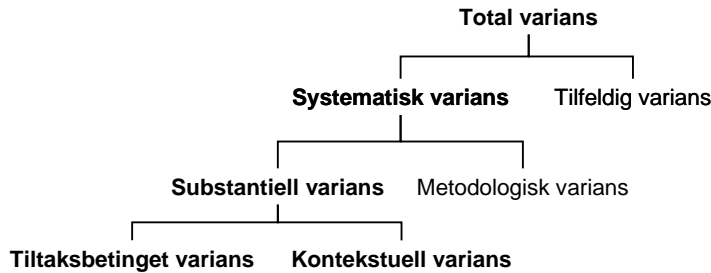
At mindre studier finner større effekter enn større studier, kan også forekomme når mindre studier har dårlig kvalitet enn større studier (Sterne et al., 2011). Dette fordi metodisk svakere studier ofte finner større effekter enn metodisk mer solide studier.

4.2.6 Vurdering av resultatenes pålitelighet

Dette avsnittet gir en generell oversikt over hvordan man kan vurdere hvor pålitelige resultatene fra metaanalyse er. De konkrete metodene for å gjøre slike vurderinger er beskrevet i avsnittene over.

Kilder til variasjon i forskningsresultater

For de fleste tiltak som beskrives i Trafikksikkerhetsåndboken foreligger det mer enn en undersøkelse om virkninger på ulykker eller skader. Det er vanlig at resultatene av ulike undersøkelser om virkninger av det samme tiltaket varierer. Figur 13 viser en modell av kilder til variasjon i forskningsresultater (Elvik, 1994A).



Figur 13: Dekomponering av varians i resultatene fra enkeltstudier ved metaanalyse.

Den delen av varianser som man forsøker å identifisere i empiriske studier er den tiltaksbetingede variansen i figur 13. Andre former for varians som ikke kan tolkes som effekter av tiltaket, er:

- **Kontekstuell varians:** Varians som kan forklares med at tiltaket har ulike virkninger under ulike forhold, for eksempel i ulike trafikkmiljøer
- **Metodologisk varians:** Systematisk variasjon som skyldes at undersøkelser som har brukt ulike metoder, kommer til forskjellige resultater
- **Tilfeldig varians:** Varians som ikke kan forklares med konkrete faktorer og som ikke har sammenheng med tiltakets virkning.

Formålet med metaanalyser er å identifisere den tiltaksbetingede variansen, dvs. effekten av et tiltak på ulykker/skader. Metodologisk og kontekstuell varians bidrar til systematiske forskjeller mellom resultater fra enkeltstudier, og det finnes flere metoder som er beskrevet i de avsnittene over, for å oppdage og korrigere for slike forskjeller (bl.a. tester for heterogenitet, analyse av skjevheter, subgruppeanalyser og metaregresjon).

Vurdering av påliteligheten til empiriske resultater

Det finnes flere tilnærminger for å definere resultatenes pålitelighet og for å spesifisere hvilke faktorer som bør undersøkes for å vurdere dem. Her beskrives to av dem, Kriteriene etter Elvik (2005A) og systemet GRADE.

Kriteriene etter Elvik (2005A): Ifølge Elvik (2005A) er det bl.a. de følgende faktorene som bør vurderes i en metaanalyse:

- Publikasjonsskjevhet (se avsnitt 4.2.5)
- Avviksfølsomhet (se avsnitt 4.2.4)
- Effekt av statistisk vektning av enkeltresultater (random effects vs. fixed effects meta-analyse vs. beregning av uvektet gjennomsnitt; se avsnitt 4.2.2)
- Vurdering av kvaliteten til enkeltstudiene.

Det siste punktet omfatter en vurdering av hvorvidt den metodiske kvaliteten til enkeltstudier påvirker resultatet av metaanalyse. Dette kan bl.a. vurderes ved å undersøke forskjeller mellom studier av ulik kvalitet eller sammenhengen mellom en kvalitetsscore og den estimerte effekten i enkeltstudier. I Trafikksikkerhåndboken er det typisk at ett tiltak er undersøkt av studier som har brukt ett av et begrenset antall forsøksopplegg. I metaanalysene er det (der det foreligger resultater for ulike typer studier) testet hvorvidt det er forskjeller mellom ulike typer studier. I mange metaanalyser er det for eksempel testet hvorvidt resultatene er forskjellige mellom:

- Før-etter studier med vs. uten kontrollgruppe
- Før-etter studier vs. multivariate studier
- Studier med vs. uten kontroll for regresjonseffekter.

Finner man betydelige forskjeller mellom resultatene fra ulike typer studier, blir som regel kun resultatene fra de metodisk mest solide studiene lagt til grunn.

GRADE: En vurdering av «styrken» til de empiriske resultatene er det siste skrittet i arbeidet med systematiske litteratursøk og metaanalyse i PRISMA (Moher et al., 2015). Moher et al. (2015) foreslår å bruke GRADE (Grading of Recommendations Assessment, Development and Evaluation; Guyatt et al., 2011A). GRADE er utviklet i forbindelse med kliniske studier, men hovedtrekkene er like relevante i trafikksikkerhetsforskningen. En vesentlig forskjell er at de aller fleste studier om trafikksikkerhet ikke er randomiserte eksperimenter (randomized controlled trials) slik at de aller fleste studiene ikke vil kunne oppnå status som «veldig sterk». GRADE brukes ikke systematisk i arbeidet med Trafikksikkerhetsåndboken, men de generelle prinsippene for å vurdere den metodiske kvaliteten til empiriske studier er likevel omtrent de samme.

GRADE gir retningslinjer for:

- Vurderingen av kvaliteten til empiriske studier og systematiske litteraturoversikter
- Anbefalinger basert på disse resultatene (Guyatt et al, 2011A; Balshem et al., 2011).

I utledningen av **anbefalinger** er det viktig å vurdere kvaliteten, men det er også viktig å ta hensyn til andre faktorer som størrelsen på effekten og mulige bivirkninger (Balshem et al., 2011).

Med **kvaliteten** på resultatene av systematiske litteraturstudier menes graden av tillit til at resultatene er korrekte. Kvaliteten vurderes for ulike faktorer og oppsummeres i fire kategorier etter graden av tillit til at den empiriske effekten ligger nær det empiriske resultatet (Balshem et al., 2011). Denne kan være høy, moderat, lav eller veldig lav. Kvaliteten vurderes både for hver enkelt studie og for det sammenlagte resultatet. Kriteriene for å vurdere kvaliteten er beskrevet i følgende ³.

- **Skjevheter:** Det anbefales at risikoen for skjevheter vurderes for hver enkel studie og at kun studier med ingen eller liten risiko inngår i syntesen (Guyatt et al., 2011B). Skjevhet omfatter i hovedsak:
 - Selektiv rapportering av enkeltresultater (se avsnitt 4.2.5)
 - Ulikheter mellom forsøks- og kontrollgruppe, bl.a. forskjeller på andre variabler enn tiltaket under evaluering og forskjeller i måling av eksponerings- eller resultatvariabler (se avsnitt 4.2.8)
 - Manglende kontroll for forstyrrende variabler (se avsnitt 4.2.8)
 - Publikasjonsskjevhet (se avsnitt 4.2.5).
- **Unøyaktighet (tilfeldige feil, random error):** Dette omfatter vurderingen av konfidensintervallet i forhold til effektens størrelse (se avsnitt 3.4).
- **Inkonsistens/heterogenitet:** Hvor like de enkelte effektestimatene er, i hvilken grad konfidensintervallene overlapper og statistiske kriterier (se avsnitt 4.2.3).
- **Indirekthet:** Resultater kan være indirekte relevante dersom:
 - Pasientene i studien er forskjellige fra dem tiltaket er ment for (i forbindelse med Trafikksikkerhetsåndboken kan «pasienter» erstattet med veger, kjøretøy, trafikanter og lignende)
 - Tiltaket i studien er forskjellig fra tiltaket som skal evalueres

³ Beskrivelsen av kriteriene inneholder kun faktorer som er relevante i forbindelse med Trafikksikkerhetsåndboken; faktorer som kun er relevant i forbindelse med medisinske studier og/eller eksperimenter (randomized controlled trials) er ikke tatt med da de ikke (eller kun i svært liten grad) er relevante for trafikksikkerhetsstudier.

- Resultatmålet ikke er det samme som det man egentlig er interessert i. Dette gjelder bl.a. såkalte surrogatmål. I trafikksikkerhetsstudier kan surrogatmål være endringer i fart eller antall konflikter når målet man egentlig er interessert i, er antall ulykker.

Hvor brukbare er ulike typer forsøksopplegg i trafikksikkerhetsforskningen?

Den følgende oversikten viser ulike typer forsøksopplegg som er typiske i trafikksikkerhetsforskningen og hvordan de generelt sett kan klassifiseres som mer eller mindre brukbare:

<p>Metodisk gode (optimale)</p> <ul style="list-style-type: none"> ▪ Eksperimenter, dvs. kontrollerte forsøk med tilfeldig fordeling av enheter til en tiltaksgruppe og en kontrollgruppe ▪ Multivariate analyser, som bygger på en eksplisitt modell og et begrunnet valg av funksjonsform og restleddsstruktur, og som har kontrollert for regresjonseffekt, generell ulykkesutvikling, endogenitet og andre faktorer som påvirker ulykkestall / -risiko <p>Brukbare</p> <ul style="list-style-type: none"> ▪ Før-etter undersøkelser som har kontrollert for regresjonseffekt, generell ulykkesutvikling og andre mulige forstyrrende variabler ▪ Tidsrekkeanalyser som har fjernet trender, sesongvariasjoner og tilfeldige variasjoner og som har en klart definert tiltaksvariabel ▪ Multivariate analyser, som bygger på en eksplisitt modell og et begrunnet valg av funksjonsform og restleddsstruktur og hvor det er lite trolig at resultatene er påvirket av endogenitet <p>Svake</p> <ul style="list-style-type: none"> ▪ Før-etter undersøkelser som kontrollerer for generell ulykkesutvikling, men ikke for regresjonseffekt ▪ Med-uten undersøkelser som kontrollerer for et fåtall feilkilder ved å dele inn materialet i undergrupper ▪ Tidsrekkeanalyser, som kun bygger på enkle modeller ▪ Multivariate analyser, som bygger på tilgjengelige data og forutsetter normalfordelte restledd, eller hvor resultatene kan være påvirket av endogenitet <p>Utilstrekkelige</p> <ul style="list-style-type: none"> ▪ Enkle før-etter undersøkelser uten kontroll for feilkilder ▪ Enkle med-og-uten undersøkelser uten kontroll for feilkilder ▪ Ulykkesanalyser uten kontrollgruppe som bygger på forutsetninger det ikke er mulig å undersøke holdbarheten av ▪ Teoretiske beregninger av tiltaks virkninger uten å undersøke holdbarheten av grunnlaget for beregningene

Denne listen over metoder er ikke fullstendig, men dekker de mest brukte metoder i undersøkelser om virkninger av trafikksikkerhetstiltak på ulykker eller skader.

Studienes alder: For en del tiltak foreligger resultater fra meget gamle studier (for eksempel fra før 1950). Det er i hvert enkelt tilfelle vurdert hvorvidt «gamle» studier skal inkluderes i metaanalyser eller litteraturoversikter og før hvilket år studier ev. ikke lenger skal inkluderes. For noen tiltak kan eldre studier fortsatt være relevante (for eksempel vegbelysning eller planskilte kryss), mens det for mange tiltak har det skjedd så store endringer av både tiltakene og relevante kontekstfaktorer at det ikke lenger gir mening å henvise til eldre studier (for eksempel kjøretøytekniske tiltak).

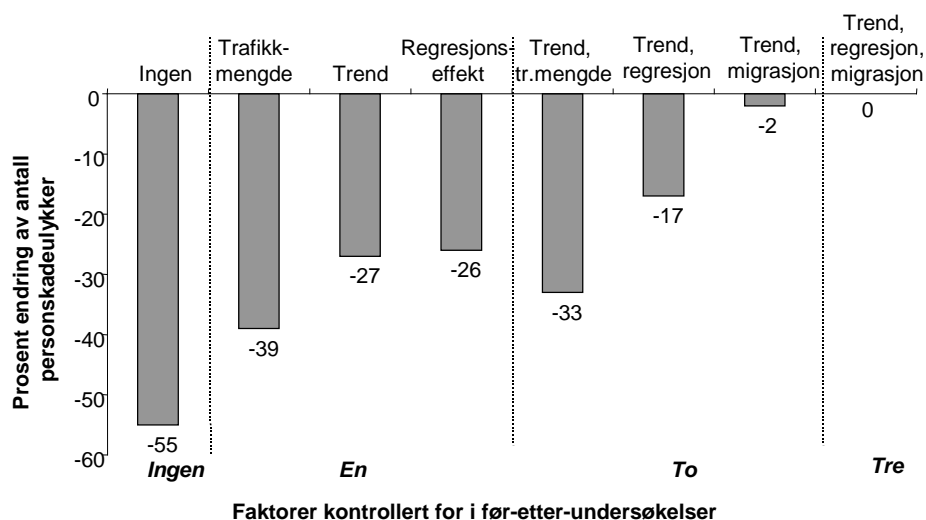
4.2.7 Eksempler på feilkildenes betydning for resultatene av undersøkelser

Dette avsnittet beskriver fire eksempler på hvordan typiske feilkilder kan påvirke resultatene av metaanalyse.

Undersøkelser om virkninger av trafikksikkerhetstiltak på ulykker og skader kan komme til ulike resultater, avhengig av hvor godt undersøkelsene har lyktes med å få kontroll over feilkildene som er nevnt foran. I dette avsnittet skal vi gi noen eksempler på dette, for en del av feilkildene som har blitt diskutert tidligere.

Eksempel 1: Systematiske målefeil i ulykkesdata. Resultatene av undersøkelser av bilbelters virkning på skader i ulykker kan bli sterkt påvirket av at beltebruken ofte er underrapportert blant uskadde eller lettere skadde (disse oppgir ofte at de brukte bilbeltet, selv om dette ikke har vært tilfellet, for å unngå problemer). Siden beltebruken blant alvorlig skadde og drepte i mindre grad er overrapportert, vil virkningen av beltebruken på risikoen for alvorlige skader være overestimert. For eksempel har Dean et al. (1995) estimert at beltebruken reduserer risikoen for å bli drept med 85% når man legger den rapporterte beltebruken til grunn, mens risikoen kun er redusert med 54% når man tar hensyn til overrapporteringen blant lettere skadde/uskadde.

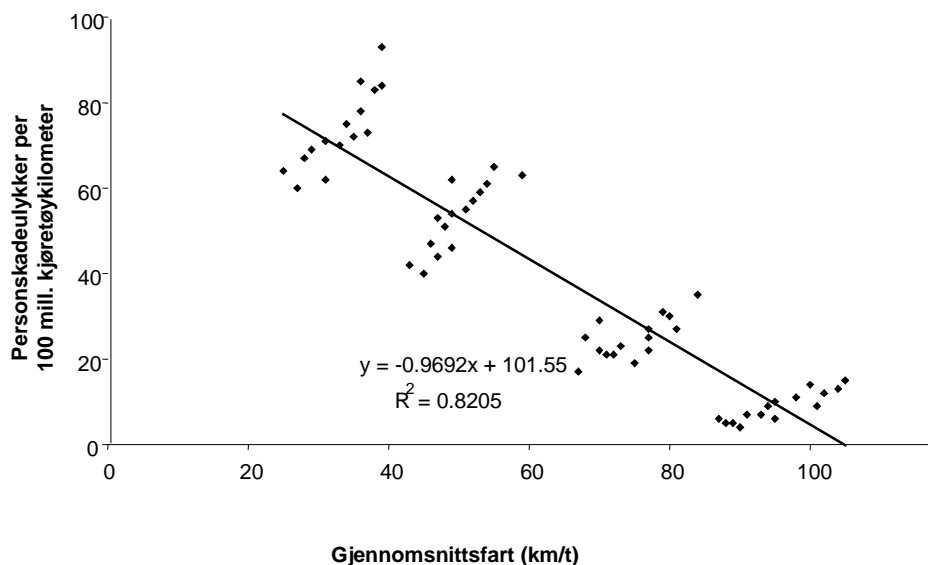
Eksempel 2: Kontroll for ulike feilkilder i undersøkelser om utbedring av ulykkesbelastede steder. Elvik (1997A) har viser at resultatene av undersøkelser om utbedring av ulykkesbelastede steder på vegnettet avhenger sterkt av hvilke feilkilder det er kontrollert for (figur 14). Figuren viser hvordan den observerte ulykkesnedgangen som tilskrives tiltaket, varierer avhengig av hvilke faktorer det er kontrollert for i undersøkelsene. Med kontroll for trend, ulykkesmigrasjon og regresjonseffekter er effekten lik null, selv om resultatene uten kontroll for disse faktorene viser en reduksjon på 55%.



Figur 14: Sammenhengen mellom hvilke faktorer det er kontrollert for og den beregnede virkning av utbedring av ulykkesbelastede steder i før-og-etterundersøkelser (Elvik, 1997A).

Eksempel 3: Feilspesifikasjon av formen på sammenhengen mellom endringer i fart og endringer i ulykkestall. Et tegn på at det er en årsakssammenheng mellom et tiltak og endringer i ulykkestall, er at man finner en såkalt dose-responsammenheng mellom tiltaket og endringen i ulykkestall. Sammenhengen beskrives i mange studier som lineære. Lineære sammenhenger kan noen ganger være egnet for å beskrive en sammenheng mellom aktuelle datapunkter, men er lite hensiktsmessig for å generalisere sammenhengen. Hovedgrunnen er at et tiltak aldri kan redusere ulykker med mer enn 100%. I tillegg har mange studier vist at det blir stadig vanskeligere å oppnå ytterligere ulykkesreduksjoner når antall ulykker allerede er lavt. Eksempelvis fant eldre studier av tiltak som skal øke bruken av bilbelte, ofte veldig store effekter. Beltebruken var da som regel i utgangspunkt lav, for eksempel under 30%. Når beltebruken imidlertid i utgangspunktet er høy (for eksempel godt over 90%), er det kun en spesiell gruppe personer som er resistente mot de fleste tiltak, som fortsatt ikke bruker bilbelte og tiltak for å øke beltebruken har langt mindre effekt, men mindre de gjør det umulig å kjøre uten belte (Høye, 2016B).

Eksempel 4: Tvetydig årsaksretning. Når man undersøker sammenhengen mellom for eksempel fartsgrense og ulykkesrisiko, kan resultatene bli absurde dersom man ikke tar hensyn til årsaksretningen. Figur 15 viser ulykkesrisikoen på veger med ulik gjennomsnittsfart. Når man ser alle vegene under ett, tyder resultatene på at jo høyere fartsgrensen er, desto lavere er risikonivået. Grunnen er dels at fartsgrensen på veier med høy risiko ofte settes ned og dels at veiene med de høyeste fartsgrensene har høy standard og bl.a. midtrekkverk, ingen plankryss og ingen myke trafikanter. Man kan følgelig ikke forvente at antall ulykker vil gå ned dersom man øker fartsgrensen uten å gjennomføre andre tiltak. Når man ser på veier med samme fartsgrense, medfører høyere fart høyere ulykkestall (hver av de fire gruppene med datapunkter i figur 15 representerer veier med samme fartsgrense).



Figur 15: Sammenheng mellom gjennomsnittsfart og ulykker (Elvik, 2008).

Sammenhengene er i mange tilfeller mye mer subtile enn i dette eksemplet og det er ikke alltid like opplagt hvilken retning årsakssammenhengen går i. Hvordan er kravene til gode undersøkelser brukt i Trafikksikkerhetsåndboken?

4.2.8 Validitet: Metodekrav og grunnlag for årsaksslutninger

Når resultater av empiriske studier er valide, betyr dette at studien har målt det den hadde til hensikt å måle. I forbindelse med Trafikksikkerhåndboken betyr det i hovedsak at resultatene viser effekten av det aktuelle tiltaket på ulykker eller skader (og ikke for eksempel effekter av andre faktorer enn tiltaket).

For å gi grunnlag for årsaksslutninger må en undersøkelse gi valide resultater. Med valide resultater menes resultater som vi har god grunn til å tro viser virkningen av det undersøkte tiltaket på en eller flere effektvariabler. Validitet betyr med andre ord at man har målt det man hadde tenkt å måle (i motsetning til reliabilitet som beskriver hvor nøyaktig man har målt det man faktisk har målt, uansett om det var det man hadde til hensikt å måle). Det kan skilles mellom fire former for validitet: Statistisk, teoretisk, intern og ekstern validitet. Disse er beskrevet i de følgende avsnittene.

Statistisk validitet

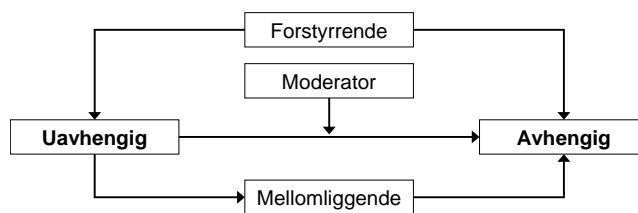
Med statistisk validitet menes nøyaktighet, feilfrihet og representativitet i resultatene. Et resultat er statistisk valid dersom det ikke skyldes tilfeldige eller systematiske feil i dataene og dersom det er representativt for en kjent populasjon av enheter. Følgende faktorer bidrar til den statistiske validiteten:

- **Undersøkelsens størrelse:** Enheter i trafikksikkerhetsstudier er som regel ulykker (eller skadde/drepte personer). Jo flere ulykker som inngår i en studie, desto mindre vil den statistiske usikkerheten i resultatene være.
- **Systematiske målefeil:** Systematiske målefeil bidrar til systematisk variasjon i resultatene, uten at denne kan tilskrives tiltaket, og kan føre til en systematisk over- eller undervurdering av virkningen. Et eksempel på systematiske målefeil er ufullstendig ulykkesrapportering.
- **Utvalgsmetode:** Ideelt sett bør tiltak iverksettes ved statistisk tilfeldig utvalgelse fra en kjent populasjon av for eksempel vegger, trafikanter eller lignende. Dette er som regel umulig, og resultater som er funnet ved et ikke-tilfeldig utvalg av for eksempel vegger kan ikke nødvendigvis generaliseres til andre typer for eksempel vegger.

Teoretisk validitet

Samsvar mellom det en undersøkelse tar sikte på å måle og det som faktisk måles. Et resultat er teoretisk valid dersom det bygger på en klar definisjonssammenheng mellom et teoretisk begrep og en målt størrelse og kan forklares ut fra en teori som støttes av resultatet. Dessverre bygger de fleste trafikksikkerhetsstudier ikke på en eksplisitt teoretisk bakgrunn. Et problem med slike studier er at man ofte kan finne «forklaringer» på de aller fleste resultater, slik at det ikke er mulig å be- eller avkrefte antakelser av hvordan et tiltak virker. Kriterier for teoretisk validitet er beskrevet i følgende avsnitt.

- **Relevante begreper og variabler:** Ulike typer variabler er skissert i figur 16. Den uavhengige variabelen er som regel et sikkerhetstiltak og den avhengige et effektmål (for eksempel antall ulykker). En *mellomliggende variabel* er en variabel som påvirkes av den uavhengige variabelen og som påvirker den avhengige variabelen. En *moderatorvariabel* påvirker sammenhengen mellom uavhengig og avhengig variabel, dvs. at sammenhengen er forskjellig mellom ulike nivåer eller grupper av moderatorvariabelen. Forstyrrende variabler er variabler som er relatert til den uavhengige og den avhengige variabelen, og sammenhengen mellom den forstyrrende og den avhengige variabelen kan forveksles med en virkning av den uavhengige variabelen. En moderator som ikke er kontrollert i undersøkelsen kan virke som er forstyrrende variabel.



Figur 16: Sammenhenger mellom ulike typer variabler.

Betydningen av de ulike typene variabler kan belyses med et eksempel. Hvorvidt en bil har elektronisk stabilitetskontroll (ESC; *uavhengig variabel*) kan påvirke ulykkesinnblandingen (*avhengig variabel*). Forklaringen er at ESC reduserer risikoen for at bilen får skrens og at skrens kan bidra til ulykker (*mellomliggende variabel*). Effekten av ESC på ulykkesinnblanding finner man især i utforkjørings- og møteulykker, men ikke bl.a. i påkjøring bakfra (dermed er ulykkestype *moderatorvariabel*). Biler med ESC har også av andre grunner færre ulykker enn biler uten ESC, bl.a. fordi de i gjennomsnitt er nyere og fordi nyere biler også generelt har bedre aktiv sikkerhet (*forstyrrende variabel*). Dersom man ikke kontrollerer for bilenes alder (eller andre egenskaper ved bilenes aktive sikkerhet) vil man følgelig overestimere virkningen av ESC.

- **Hypoteser om sammenhengen mellom variablene:** Slike hypoteser er formulert basert på en teoretisk bakgrunn. I tillegg blir de viktigste alternative hypoteser formulert, som kan forklare resultatene som ikke støtter hypotesene.
- **Hypoteser om virkningsmekanismen:** Virkningsmekanismer er faktorer som kan forklare virkningen av den uavhengige på den avhengige variabelen, dvs. mellomliggende variabler. Når man har en hypotese om hvilke virkninger av et tiltak som kan forklare virkningen på ulykker, og når man finner slike virkninger, øker dette den teoretiske validiteten. Virkningsmekanismer kan være konkrete risikofaktorer for ulykker (for eksempel fart eller promillekjøring), eksponering, eller risikofaktorer for skader.

Intern validitet

Intern validitet betyr at det er mulig å dra slutninger om årsakssammenhenger mellom en årsaksfaktor og virkningen av denne. Intern validitet fanger opp både statistisk, teoretisk og ekstern validitet. Med utgangspunkt i epidemiologisk teori er indikasjoner på at A har en årsakssammenheng med B følgende (Elvik, 2008):

- **Statistisk sammenheng mellom tiltak og virkning:** Det finnes en statistisk sammenheng mellom tiltak og virkning. Den interne validiteten er svekket hvis en statistisk sammenheng er inkonsistent og ikke kan forklares med (teoretisk begrunnede) moderatorvariabler.
- **Entydig retning av sammenhengen mellom tiltak og virkning:** Det finnes teoretiske og statistiske grunner for å påstå at det er tiltaket som forårsaker virkningen og ikke omvendt. Bl.a. bør tiltaket komme før virkningen i tid og man skal kunne forklare hvordan tiltaket påvirker virkningen.
- **Dose-responssammenheng:** Når et tiltak settes inn i ulike styrker forventer man at også styrken av virkningen varierer (jo mer tiltak, desto mer virkning).
- **Kontroll for forstyrrende variabler:** Forstyrrende variabler har sammenheng med effektvariabelen og kan føre til at endringer på effektvariabelen tilskrives tiltaket, selv om de skyldes den forstyrrende variabelen. Et typisk eksempel er den generelle ulykkesutviklingen. Antall ulykker har en tendens til å gå ned over tid; dersom man ikke kontrollerer for denne tendensen, vil man i en før-etterstudie kunne feilaktig konkludere at antall ulykker har gått ned på grunn av et tiltak, selv om antall ulykker uansett ville ha gått ned. Andre tiltak som er innført samtidig med tiltaket som skal evalueres, kan også fungere som forstyrrende variabler. Muligheter for å kontrollere for forstyrrende variabler er bl.a. forsøksopplegget (bl.a. bruk av kontrollgruppe) og statistisk kontroll (bl.a. ved bruk av multivariate modeller).

Noen generelle feilkilder i undersøkelser om virkninger av trafikksikkerhetstiltak som reduserer den interne validiteten er følgende:

- **Regresjonseffekt i ulykkestall:** Når enheter i en før-etterstudie som får et tiltak i før-situasjonen har veldig mange ulykker, er det som regel en tendens til at ulykkestallet går ned i etterperioden, selv uten effektivt tiltak (jf. avsnitt 3.2). Manglende kontroll for regresjonseffekt fører derfor i de fleste tilfeller til at tiltakets effekt overvurderes.
- **Endogenitet:** Når det ikke er tilfeldig hvor eller for hvilke personer tiltak brukes kan resultatene være påvirket av endogenitet. Endogenitet kan føre til at tiltakets virkning overestimeres (for eksempel hvis det kun er spesielt sikkerhetsbevisste personer som tar i bruk tiltaket) eller at den underestimeres eller at man finner en ulykkesøkning (for eksempel når tiltak settes inn på spesielt ulykkesbelastede steder som i en med-uten studie sammenlignes med mindre ulykkesbelastede steder).
- **Ulykkesmigrasjon:** Innføring av et tiltak et sted eller for en gruppe, kan føre til at ulykkene «forflyttes» til andre steder eller grupper. Manglende kontroll for ulykkesmigrasjon medfører nesten alltid en overvurdering av virkningene av et tiltak
- **Ekspertimenteffekter:** Dette er effekter som oppstår som følge av at det utføres et eksperiment eller at en størrelse blir målt. Eksempler er fartsmålinger som fører til redusert fart fordi måleposten ikke er godt nok skjult eller økt politikontroll som fører til at rapporteringsgraden øker.
- **Feil form av den uavhengige variabelen eller av regresjonsmodellen:** Når man for eksempel beregner en lineær sammenheng mellom to variabler, mens sammenhengen er U-formet, vil man kun finne en svak statistisk sammenheng, selv om de to variablene er perfekt korrelert. Hvor sterk sammenheng man finner mellom to variabler er derfor avhengig av hvilke transformasjoner av variablene (for eksempel lineær eller logaritmisk) og hvilken modellform man bruker.

- **Kollinearitet og utelatte variabler:** Resultater av multivariate modeller kan gi et skjevt bilde av sammenhengen mellom to variabler både når det er for mange prediktorvariabler i modellen som er korrelert med hverandre, og når relevante tredjevariabler er utelatt. Problemet er at man aldri kan være sikker på at alle relevante tredjevariabler er inkludert i modellen.

Siden man ikke kan være sikker på at man kjenner alle kontrollvariabler, er kontroll som ikke bygger på randomisering alltid mangelfull. Kravet om kontroll for tredjevariabler må her formuleres som et krav om kontroll for alle *til enhver tid kjente* kontrollvariabler.

Ekstern validitet

Resultater av empiriske studier er eksternt valide dersom de er generaliserbare. Ekstern validitet kan best bedømmes ved å sammenlikne resultater av ulike undersøkelser om samme tiltak under ulike forhold. Resultater har høy ekstern validitet dersom de er stabile i tid og rom og på tvers av ulike undersøkelsesmetoder (Elvik, 2012). Eventuelle forskjeller i resultater mellom undersøkelser bør kunne forklares ut fra kjente egenskaper ved metodene, landene eller forholdene på den tiden undersøkelsen ble utført. Ekstern validitet kan, til en viss grad, kompensere for manglende teoretisk bakgrunn. Den eksterne validiteten til resultater fra undersøkelser om virkninger av trafikksikkerhetstiltak er redusert når det finnes effekter som er spesifikke for den kontekst en undersøkelse er utført i.

4.2.9 Presentasjon av resultater fra metaanalyse i Trafikksikkerhåndboken

Resultatene fra metaanalyse presenteres i Trafikksikkerhåndboken i et standard-oppsett hvor resultatene presenteres med beste anslag på effekten med 95%-konfidensintervall. Resultatene presenteres (der det er relevant) for ulike varianter av tiltaket, ulike skadegrader og ulike ulykkestyper. Resultater som er presentert i tabellform kan ikke nødvendigvis tolkes som «fasitsvar»; eventuelle forbehold er beskrevet i teksten.

I Trafikksikkerhåndboken er resultatene av undersøkelser om et tiltak alltid gruppert etter følgende kjennetegn:

- Varianter av tiltaket
- Ulykkers eller skaders alvorlighetsgrad
- Ulykkestype som påvirkes.

For hver tiltaksvariant, skadegrad og ulykkestype presenteres effektens størrelse og 95% konfidensintervallet.

Effektens størrelse: Når det er funnet en stor virkning av et tiltak kan dette skyldes at tiltaket er effektivt, men det kan også finnes andre forklaringer. Som tidligere nevnt kan metodiske svakheter og publikasjonsskjevheter ha bidratt til at store virkninger ble funnet. Hvorvidt dette kan være tilfelle, er i Trafikksikkerhåndboken så langt som mulig testet empirisk eller vurdert kvalitativt.

Statistisk signifikans og konfidensintervaller: En estimert effekt betegnes som statistisk signifikant når konfidensintervallet ikke inneholder null. Konfidensintervaller og statistisk signifikans sier imidlertid ingenting om hvor pålitelig resultatet er (jf. avsnitt 4.2.6) eller hvorvidt det er praktisk relevant.

Det kan derfor være misvisende å benytte kun virkningens størrelse og statistisk signifikans for å vurdere hvorvidt et resultat er av betydning; i tillegg er det viktig å ta i betraktning hvor stor effekten er, hvor gode de empiriske studiene er, og hvorvidt resultatet kan antas å være generaliserbar. Dette er diskutert i avsnitt 3.4.

5 Nyttekostnadsanalyse og ulykkeskostnader

I dette kapitlet beskrives hovedtrekkene i nyttekostnadsanalyser av trafikksikkerhetstiltak og de ulykkeskostnader som benyttes i slike analyser. Hensikten med beskrivelsen er å gi leseren en forståelse av de nyttekostnadsanalyser som presenteres i tiltakskapitlene. Disse nyttekostnadsanalysene er, så langt det er mulig, i samsvar med de retningslinjer for slike analyser Statens vegvesen benytter. Det henvises derfor til disse retningslinjene og kun hovedtrekkene i dem beskrives her (Statens vegvesen, Håndbok V712, 2018).

5.1 Samfunnsøkonomiske analyser og nyttekostnadsanalyser

I hvert tiltakskapittel i Trafikksikkerhåndboken (del 2) er det et avsnitt om nyttekostnadsvurderinger av tiltaket. Formålet med disse avsnittene er å gi opplysninger som kan si noe om hvor kostnadseffektivt eller samfunnsøkonomisk lønnsomt et tiltak kan være. Slike opplysninger har interesse dersom man ønsker å prioritere trafikksikkerhetstiltak på grunnlag av kostnadseffektivitet eller samfunnsøkonomisk lønnsomhet. Slike prioriteringskriterier representerer likevel bare en av flere mulige tilnæringsmåter for å avgjøre hvilke trafikksikkerhetstiltak som bør gjennomføres.

Samfunnsøkonomiske analyser

Samfunnsøkonomiske analyser omfatter blant annet kostnadseffektivitetsanalyser og nyttekostnadsanalyser (Norges Offentlige Utredning 2012:16).

Kostnadseffektivitetsanalyse: En kostnadseffektivitetsanalyse går ut på å finne den billigste måten å nå et gitt mål på, eksempelvis hvilke tiltak som gir den største reduksjonen av ulykker, skadde eller skadeposter per krone det koster å gjennomføre tiltaket. Det er ikke nødvendig å omregne det målet man skal nå, til en pengeverdi. Kostnadseffektiviteten kan i trafikksikkerhetssammenheng defineres som for eksempel antall unngåtte ulykker eller skader per investert krone. Hvis et tiltak har flere virkninger enn å forebygge ulykker og hvis det er mulig å uttrykke de andre virkningene (enn færre eller mindre alvorlige skader) som pengeverdi, kan man i kostnadseffektivitetsanalyser bruke en samfunnsøkonomisk nettokostnad (Mattsson, 1991). Dette betyr at man trekker pengeverdien av tilleggsnytt fra kostnadene.

Nyttekostnadsanalyse: I en nyttekostnadsanalyse regnes alle fordeler og ulemper ved et tiltak om til kroneverdier. På denne måten gjøres alle typer virkninger sammenlignbare og det blir mulig å avgjøre om nytten er større eller mindre enn kostnadene. Et tiltak hvor nytten (fordelene) er større enn kostnadene (ulempene) er samfunnsøkonomisk lønnsomt. Den samfunnsøkonomiske lønnsomhet kan uttrykkes på ulike måter:

- **Netto nåverdi:** Differanse mellom nytte og kostnad
- **Nytte-kostnadsforhold:** Nyten delt på kostnadene; er denne større enn én er tiltaket samfunnsøkonomisk lønnsom
- **Netto-nytte per budsjettkrone:** Nettonytte (summen av alle nyttekomponentene minus kostnadene) delt på budsjettkostnadene (Statens vegvesen, Håndbok V712, 2018). Når denne er større enn null er tiltaket samfunnsøkonomisk lønnsom.

Nyttekostnadsanalyser

Nyttekostnadsanalyser er blitt brukt i evaluering av trafikksikkerhetstiltak i mange år (Trilling, 1978). Bruken av nyttekostnadsanalyser er likevel kontroversiell (Hauer, 2011). Nyttekostnadsanalyser brukes som regel for problemstillinger eller prosjekter med følgende kjennetegn:

- Prosjektet medfører offentlige utgifter, ofte investeringer
- Det finnes flere mål som må tas hensyn til, og disse kan stå helt eller delvis i motsetning til hverandre
- Ett eller flere mål gjelder ikke-markedsgoder, som f.eks. miljø, kriminalitet eller trafikksikkerhet
- Ressursene er begrensede og det er ønskelig å bruke dem på en mest mulig effektiv måte.

Disse kjennetegnene passer godt for trafikksikkerhetstiltak som iverksettes av offentlige myndigheter. Slike tiltak medfører offentlige utgifter, inngår i en politikk som også har andre mål enn å bedre trafikksikkerheten, produserer et ikke-markedsgode (bedre trafikksikkerhet) og må finansieres av begrensede budsjetter. Generelle innføringer i nyttekostnadsanalyser er gitt blant annet av Mishan og Quah (2007), Boardman et al. (2018), Norges Offentlige Utredninger (2012:16) og Statens vegvesen (Håndbok V712, 2018).

Faktorer som inngår i dagens nyttekostnadsanalyser

Statens vegvesen har utarbeidet et opplegg for nyttekostnadsanalyser av vegprosjekter (Statens vegvesen, håndbok V712, Konsekvensanalyser, 2018). Tabell 1 viser hvilke prissatte faktorer som inngår i disse analysene.

Tabell 1: Oversikt over konsekvenser som beskrives i Statens vegvesens håndbok om konsekvensanalyser (Statens vegvesen, håndbok V712, 2018).

Aktører	Tema	Deltema
Trafikanter og transportbrukere	Trafikant- og transportbrukernytte	<ul style="list-style-type: none"> ▪ Distanseavhengige kjørekostnader ▪ Andre reiseutlegg ▪ Tidsbruk ▪ Ulempkostnader i ferjesamband og ved vegstengning ▪ Helsevirkninger av økt gang- og sykkeltrafikk ▪ Utrygghet for gående og syklende
Operatører	Operatørnytte	<ul style="list-style-type: none"> ▪ Operatørselskapenes kostnader (kollektivselskap, bompengeselskap, ferjeselskap, parkeringselskap) ▪ Brukerinntekter og overføringer
Det offentlige	Budsjettvirkning	<ul style="list-style-type: none"> ▪ Investering, drift og vedlikehold, tilskudd til kollektivtrafikk, ▪ Skatteinntekter
Miljø og samfunn for øvrig	Trafikkulykker	<ul style="list-style-type: none"> ▪ Personskadeulykker og materiellskadeulykker
	Restverdi	<ul style="list-style-type: none"> ▪ Framtidig nytte av tiltaket etter beregningsperioden (restverdi)
	Skattekostnad	<ul style="list-style-type: none"> ▪ Effektivitetstap knyttet til skattefinansiering, 20 % av offentlige utgifter
	Støy og luftforurensning	<ul style="list-style-type: none"> ▪ Støyplage innendørs. ▪ Lokal og regional luftforurensning ▪ Klimaeffekter

5.2 Ulykkeskostnader

Det ble i 2009 og 2010 gjennomført en omfattende verdsettingsstudie der de aller fleste goder uten markedspris inngikk og nye økonomiske verdsettinger av disse ble beregnet (Samstad et al., 2010). Denne verdsettingsstudien omfattet også ulykkeskostnadene (Veisten et al., 2010) som også benyttes av Statens vegvesen (2018). Ulykkeskostnadene er summen av fem hovedposter:

Medisinske kostnader: Alle kostnader til medisinsk behandling av personskader i trafikkulykker, inkludert hjemmesykepleie og private medisinske kostnader.

Materielle kostnader: Alle kostnader til reparasjon eller erstatning av skadde kjøretøy eller andre gjenstander som blir skadet eller ødelagt ved trafikkulykker. I hovedsak dekkes disse kostnadene av trafikantene gjennom motorvognforsikring.

Administrative kostnader: Økt ressursbruk til administrasjon som følge av trafikkulykker. Dette omfatter forsikringsadministrasjon, trygdeadministrasjon, politiets kostnader til ulykkesrapportering og kostnader til rettssaker som har sitt utspring i trafikkulykker. Kostnader til forsikringsadministrasjon forutsettes dekket på samme måte som de materielle kostnadene. Øvrige administrative kostnader forutsettes dekket av offentlig sektor. Utbetalte erstatninger og trygder inngår ikke direkte i ulykkeskostnadene da disse er en måte å dekke ulykkeskostnadene på. Den samfunnsøkonomiske kostnaden trygder og erstatninger dekker, er inntektsbortfall som skyldes at man ikke lenger kan delta i yrkeslivet.

Produksjonsbortfall: Verdien av den produksjonsmengde som går tapt på grunn av trafikkskader. Kostnadene omfatter både betalt og ubetalt produksjon. Kostnadene ved ubetalt husholdsproduksjon er beregnet ved å ta utgangspunkt i hva det koster å kjøpe slike tjenester. Ved dødsfall er kostnader ved tapt framtidig produksjon medregnet fram til forventet pensjonsalder.

Velferdstap ved trafikkskader: Enhver nedsettelse av velferden som ulykkene medfører. Velferd er den grad av velvære og trivsel den enkelte opplever, materielt og mentalt. Å ha høy velferd betyr at man får dekket grunnleggende materielle og fysiske behov, samtidig som man fungerer normalt i dagliglivet og opplever en tilstand av fysisk, mentalt og sosialt velvære.

De tre førstnevnte postene kalles tradisjonelt for direkte kostnader fordi de i prinsippet kan gjenfinnes i form av konkrete utbetalinger eller tapte inntekter for de trafikkskadde og andre berørte parter i samfunnet. De direkte kostnadene kan stort sett beregnes ved å bygge på eksisterende markedspriser. Produksjonsbortfall kalles indirekte kostnader, fordi det i mange tilfeller ikke dreier seg om direkte utlegg men om tapte produksjonsmuligheter, ikke en konkret observerbar nedgang i produksjon.

En beregning eller estimering av direkte eller indirekte kostnader er ikke mulig når det gjelder **velferdstap** ved trafikkskader. Den økonomiske verdsettingen av velferdstap er ikke et spørsmål om hva ulykkene koster når de først har skjedd, men om hvor mye ressurser samfunnet er villig til å bruke på å redusere antall ulykker og deres konsekvenser. Velferdstapet verdsettes følgelig ut fra hvor mye folk er villige til å betale for å redusere risikoen for trafikkskader. I verdsettingsstudien ble en rekke metoder benyttet for å finne ut hvor mye folk er villige til å betale for bedre trafikksikkerhet. Hensikten med å benytte flere metoder, var å få vite hvor mye valget av metode påvirker resultatene. Tidligere studier viser at resultatene av verdsettingsstudier påvirkes betydelig av metoden som brukes. De verdier som anbefales i nyttekostnadsanalyser er forholdsvis konservative, det vil si at verdsettingsstudien inneholder en rekke anslag på verdien av trafikksikkerhet som er høyere enn dem som anbefales brukt.

Totale ulykkeskostnader som benyttes i samfunnsøkonomiske analyser er (i 2016-kr.):

- 30,2 mill. kr. per dødsfall
- 27,1 mill. kr. per meget alvorlig skadd
- 9,6 mill. kr. per alvorlig skadd
- 0,73 mill. kr. per lett skadd
- 0,038 mill. kr. per materiell skade (Statens vegvesen, 2018).

Kostnaden ved en drept person ble ved siste revisjon av Statens vegvesens håndbok V712 redusert fra 35,3 millioner kroner (2013-verdi) til 30,2 millioner kroner (2016-verdi). Reduksjonen er ikke begrunnet. Det er henvist til NOU 2012:16, men verdsettingen av en drept i den utredningen er i 2012-verdi. Kostnadstallene for drepte/skadde gjelder drepte/skadde i politirapporterte personskadeulykke. Det er tatt hensyn til underrapportering av ulykker og skader i offisiell ulykkesstatistikk. Kostnadene varierer mellom ulike ulykkestyper.

Kostnaden ved en hardt skadd person (meget alvorlig pluss alvorlig skade) er 11,2 millioner kroner (2016). Gjennomsnittlig kostnad ved en personskadeulykke er 3,0 millioner kroner (2016).

Usikkerhet i ulykkeskostnadene. Ulykkeskostnadene som ble beregnet i verdsettingsstudien har en usikkerhet på minst 20% i begge retninger. Det meste av denne usikkerheten knytter seg til hvilken verdsettingsmetode som er best. Den rent statistiske usikkerheten i verdiene er relativt liten.

Hva fanger ulykkeskostnadene ikke opp? Aspekter ved trafikksikkerhet som ikke uten videre inngår i ulykkeskostnadene, i den form disse foreligger og brukes i dag, er bl.a. utjevning av forskjeller i risiko mellom trafikantgrupper og katastrofepotensialet ved ulykker. Skulle man ivareta et mål om å redusere forskjeller i risiko mellom trafikantgrupper, måtte kostnaden per ulykke være proporsjonal med risikoen per personkilometer, slik at trafikantgrupper med høy risiko ville få en høy kostnad per ulykke. Potensialet for katastrofer i vegtrafikken er trolig mindre i vegtrafikken enn i andre transportsektorer. Det er hovedsakelig knyttet til ulykker på motorveg i tett tåke, ulykker i tunneler, store bussulykker og ulykker under transport av farlig gods. Hvis man skulle legge spesiell vekt på å forebygge katastrofer, måtte kostnaden per skadet eller drept person ved katastrofer være høyere enn ved andre vegtrafikkulykker.

Det er viktig å huske at den dominerende posten i ulykkeskostnadene, verdsetting av velferdstap, gjenspeiler det trafikantene er villige til å betale for en redusert risiko for dødsfall eller skader i trafikken. Dette kalles ofte *ex ante* (på forhånd) verdsetting. Når risikoen materialiserer seg i form av ulykker og skader, kan den erstatning ofrene måtte få for å oppnå samme velferdsnivå som før de ble skadet, godt tenkes å overstige betalingsvilligheten for en redusert risiko for skaden. Det kan være en vesentlig forskjell på å verdsette et hypotetisk utfall (en skade som ikke har skjedd) og et faktisk utfall (en skade som har skjedd).

Verdsetting av andre goder uten markedspris

I nyttekostnadsanalyser inngår ikke bare ulykkeskostnader, men også en økonomisk verdsetting av andre goder uten markedspris. Verdsettinger som i dag anbefales benyttet i nyttekostnadsanalyser er oppsummert i Statens vegvesens håndbok V712, konsekvensanalyser (2018).

Tiltaks teknisk-økonomiske levetid

I nyttekostnadsanalyser regnes nytten med i hele et tiltaks teknisk-økonomiske levetid. Framtidig nytte omregnes til nåverdi ved hjelp av en kalkulasjonsrente. Nåverdien av framtidig nytte eller framtidige kostnader er den verdi vi tillegger nytten eller kostnadene i dag. Ulike trafikksikkerhetstiltak har ulike teknisk-økonomisk levetid. Tabell 2 viser den teknisk-økonomiske levetid som er forutsatt for ulike hovedgrupper av tiltak.

For en del tiltak vil den teknisk-økonomiske levetiden variere avhengig av trafikkmengden. Dette gjelder særlig en del vegvedlikeholdstiltak. Spennvidden i levetid er fra ca. 1 år til ca. 10 år. For slike tiltak er den teknisk-økonomiske levetiden beregnet i hvert tilfelle ved hjelp av modeller utviklet av Statens vegvesen.

Tabell 2: Teknisk-økonomisk levetid for ulike grupper av trafikksikkerhetstiltak.

Gruppe av tiltak	Teknisk-økonomisk levetid
Veginvesteringstiltak	40 år
Kjøretøytiltak	18 år
Trafikkskilt, mindre utbedringer av veger	10 år
ATK-utstyr	10 år
Vegoppmerking	1-10 år (avh. av trafikkmengde)
Reasfaltering, nye vegdekker	1-10 år (avh. av trafikkmengde)
Vinterdriftstiltak	1 år (1 vinter)
Kjøretøykontroll	1 år
Føreropplæringstiltak	1-3 år
Informasjonskampanjer	1 år
Politikontroll	1 år

Andre tiltak har bare virkning samtidig med at tiltaket er i bruk. Dette gjelder f.eks. informasjonskampanjer og, i det minste som en brukbar tilnærming, politikontroller. For slike tiltak er varigheten av tiltaket og av virkningen av praktiske grunner satt lik ett år. Nytte og kostnad for tiltaket kommer da i samme år. Kalkulasjonsrenten ved beregningen av nåverdien av tiltak er satt lik 4% per år (Finansdepartementet, 2005). Det betyr at verdien i dag (ofte kalt nåverdien) av en fremtidig nytte eller kostnad reduseres med 4% per år.

Beregning av tiltakenes nytte

I mange av de nyttekostnadsanalyser som presenteres i Trafikksikkerhåndboken, er det brukt en beregningsmodell av følgende type for å beregne nytten av et tiltak:

$$\text{Nytte} = \text{Eksposering} \cdot \text{Risiko} \cdot \text{Tiltakets virkning} \cdot \text{Ulykkeskostnad}$$

Produktet av eksposering (trafikkmengde, kjørelengde) og risiko angir antallet ulykker eller skader et tiltak virker på. Produktet av eksposering, risiko og tiltakets virkning angir antallet unngåtte ulykker eller skader. Dette antallet ganget med ulykkeskostnaden gir innsparte ulykkeskostnader.

Beregningene krever med andre ord at vi gjør visse forutsetninger om eksposering og risiko for ulike vegtyper, kjøretøytyper og trafikantgrupper. Som en hovedregel er det forutsatt at eksposering og risiko er lik gjennomsnittet for vedkommende gruppe for vegtyper, kjøretøygrupper og trafikantgrupper. Risikoen (antall ulykker/drepte/skadde per mill. kjøretøykilometer) for ulike vegtyper kan beregnes med hjelp av ulykkesmodellene som er utviklet av Høye (2016). Mange tiltak på vegnettet påvirker dessuten fartsnivået. For disse tiltakene er det derfor også gjort antakelser om gjennomsnittlig fartsnivå.

6 Referanser

- Amrhein, V., Greenland, S., & McShane, B. (2019). Retire statistical significance. *Nature*, 567, 305-307.
- Amundsen, F. H. & Ranes, G. (1997). Trafikkulykker i vegtunneler. En analyse av trafikkulykker fra 1992-96 i vegtunneler på riksvegnettet. Rapportutkast. Oslo, Vegdirektoratet, Transport og trafikksikkerhetsavdelingen, Transportanalysekontoret.
- Amundsen, F.H. & Engebretsen, A. (2008). Traffic accidents in road tunnels. An analysis of traffic accidents in tunnels on national roads for the period 2001-2006. TS Report 7-2008. Oslo: Public Roads Administration.
- Anzures-Cabrera, J., & Higgins, J. P. T. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods*, 1(1), 66-80.
- ASA (2016). Statement on statistical significance and p-values. American Statistical Association.
- Balsheim, H., Helfand, M., Schünemann, H. J., Oxman, A. D., Kunz, R., Brozek, J., . . . Guyatt, G. H. (2011). GRADE guidelines: 3. Rating the quality of evidence. *Journal of Clinical Epidemiology*, 64(4), 401-406.
- Bax, L., Ikeda, N., Fukui, N., Yaiu, Y., Tsuruta, H., & Moons, K. G. M. (2008). More Than Numbers: The Power of Graphs in Meta-Analysis. *American Journal of Epidemiology*, 169(2), 149-255.
- Bjørnskau, T. & S. Fosser (1996). Bilisters atferdstilpasning til innføring av vegbelysning. Resultater fra en før- og etterundersøkelse på E-18 i Aust-Agder. TØI Report 332. Oslo: Institute of Transport Economics.
- Bjørnskau, T. (1994B). Hypoteser om atferdstilpasning (risikokompensasjon). Arbeidsdokument TST/0512/94. Oslo, Transportøkonomisk institutt.
- Bjørnskau, T. (2000). Risiko i vegtrafikken 1997-1998. TØI-rapport 483/2000. Oslo: Transportøkonomisk institutt.
- Bjørnskau, T. (2003). Risiko i vegtrafikken 2001-2002. TØI-rapport 694/2003. Oslo: Transportøkonomisk institutt.
- Bjørnskau, T. (2008). Risiko i vegtrafikken 2005-2007. TØI-rapport 986/2008. Oslo: Transportøkonomisk institutt.
- Bjørnskau, T. (2011). Risiko i vegtrafikken 2009-2010. TØI-rapport 1164/2011. Oslo: Transportøkonomisk institutt.
- Bjørnskau, T. (2015). Risiko i vegtrafikken 2013-2014. TØI-rapport 1448/2015. Oslo: Transportøkonomisk institutt.
- Boardman, A. E., Greenberg, D. H., Vining, A. R., & Leimer, D. L. (2018). Cost-benefit analysis. Concepts and practice. Fifth edition. Cambridge, Cambridge University Press.

- Borenstein, M., Higgins, J.P.T., Hedges, L.V., & Torhstein, H.T. (2017). Basics of meta-analysis: I2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8, 5-18.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T. & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, John Wiley and Sons,.
- Borger A. (1995). Underrapportering av trafikkkulykker. Metaanalyse av rapporteringsgrad. Arbeidsdokument TST/0690/95. Oslo: Institute of Transport Economics.
- Chan, A. W., Hrobjartsson, A., Haahr, M. T., Gotzsche, P. C., & Altman, D. G. (2004). Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*, 291(20), 2457-2465.
- Christensen, P. (2003). *Topics in meta-analysis*. Report 692/2003. Oslo: Institute of Transport Economics.
- Cochran, W. G. (1954). The combination of estimates from different experiments. *Biometrics*, 10, 101-129.
- Cook, T. D. & D. T. Campbell (1979). *Quasi-Experimentation. Design and Analysis Issues for Field Settings*. RandMcNally, Chicago.
- Cooper, C., Lovell, R., Husk, K., Booth, A., & Garside, R. (2018). Supplementary search methods were more effective and offered better value than bibliographic database searching: A case study from public health and environmental enhancement. *Research Synthesis Methods*, 9(2), 195-223.
- Dean, J. M., J. C. Reading & P. J. Nechodom (1995). Overreporting and Measured Effectiveness of Seat Belts in Motor Vehicle Crashes in Utah. *Transportation Research Record*, 1485, 186-191.
- Duval, S. & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455-463.
- Egger, M., Smith, G. D., & Sterne, J. A. C. (2001). Uses and abuses of meta-analysis. *Clinical Medicine*, 1(6), 478-484.
- Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, 315, 629-634.
- Elvik, R., T. Vaa & E. Østvik (1989). *Trafikksikkerhetshåndbok*. Revidert utgave. Oslo: Institute of Transport Economics.
- Elvik, R. (1994A). Metaanalyse av effektmålinger av trafikksikkerhetstiltak. TØI rapport 232. Oslo, Transportøkonomisk institutt.
- Elvik, R. (1995A). An analysis of official economic valuations of traffic accident fatalities in 20 motorized countries. *Accident Analysis and Prevention*, 27, 237-247.
- Elvik, R. & R. Muskaug (1994). Konsekvensanalyser og trafikksikkerhet. TØI Report 281. Oslo: Institute of Transport Economics.
- Elvik, R. (1997A). Evaluations of road accident blackspot treatment: A Case of the Iron Law of evaluation studies? *Accident Analysis and Prevention*, 29, 191-199.
- Elvik, R., Mysen, A.B. & Vaa, T. (1997). *Trafikksikkerhetshåndboken (Handbook of road safety measures)*. Oslo: Institute of Transport Economics.

- Elvik, R. (1998). Evaluating the statistical conclusion validity of weighted mean results in meta-analysis by analyzing funnel graph diagrams. *Accident Analysis and Prevention*, 30, 225-266.
- Elvik, R. (1999). Can injury prevention efforts go too far? Reflections on some possible implications of Vision Zero for road accident fatalities. *Accident Analysis and Prevention*, 31, 265-286.
- Elvik, R. & A. B. Mysen. (1999). Incomplete accident reporting. Meta-analysis of studies made in 13 countries. *Transportation Research Record*, 1665, 133-140.
- Elvik, R. (2004). To what extent can theory account for the findings of road safety evaluation studies? *Accident Analysis & Prevention*, 36(5), 841-849.
- Elvik, R. & Vaa, T. (2004). *Handbook of road safety measures*. Amsterdam: Elsevier.
- Elvik, R. & Bjørnskau, T. (2005). How accurately does the public perceive differences in transport risks? An exploratory analysis of scales representing perceived risk. *Accident Analysis and Prevention*, 37, 1005-1011.
- Elvik, R. (2005A). Can we trust the results of meta-analysis? *Transportation Research Record*, 1980, 221-229.
- Elvik, R. (2005B). Introductory guide to systematic reviews and meta-analysis. *Transportation Research Record*, 1908, 230-235.
- Elvik, R. (2005C). A catalogue of risks of accidental death in various activities. Arbeidsdokument SM/1661/2005. Oslo: Transportøkonomisk institutt.
- Elvik, R. (2009). Potensmodellen for sammenhengen mellom fart og trafikksikkerhet. En oppdatering. TØI-rapport 1034/2009. Oslo: Transportøkonomisk institutt.
- Elvik, R. (2008). Criteria for causal inference: lessons from epidemiology applied to safety evaluation. Paper prepared for Workshop on Future Directions in Highway Crash Data Modeling, Washington D. C. November 20-21, 2008.
- Elvik, R., Høy, A., Vaa, T., & Sørensen, M. (2009). *The handbook of road safety measures*. Bingley, UK: Emerald.
- Elvik, R., Kolbenstvedt, M., Elvebakk, B., Hervik, A., & Bræin, L. (2009B). Costs and benefits to Sweden of Swedish road safety research. *Accident Analysis & Prevention*, 41(3), 387-392.
- Elvik, R. & Høy, A., & Christensen, P. (2009C). Elementary Units of Exposure. *Transportation Research Record*, 2103, 25-31.
- Elvik, R. (2012). The range of replications technique for assessing the external validity of road safety evaluation studies. *Accident Analysis and Prevention*, 45, 272-280.
- Elvik, R. (2013A). A re-parameterisation of the Power Model of the relationship between the speed of traffic and the number of accidents and accident victims. *Accident Analysis & Prevention*, 50, 854-860.
- Elvik, R. (2013B). Can a safety-in-numbers effect and a hazard-in-numbers effect co-exist in the same data? *Accident Analysis and Prevention*, 60, 57-63.
- Elvik, R. (2017). Fartsgrensepolitikk. Virkninger på trafikksikkerhet av ulike fartsgrenser. TØI-rapport 1589. Oslo, Transportøkonomisk institutt.

- Elvik, R. (2018). Analyse av syklistskader i Oslo. Rapporteringsgrad, helsekonsekvenser og sammenligning med svenske data. Arbeidsdokument SM/51134. Oslo, Transportøkonomisk institutt.
- Elvik, R., Høye, A., Vaa, T., & Sørensen, M. W. J. (2013). El Manual de Medidas de Seguridad Vial. Emerald.
- Elvik, R., Høye, A., Vaa, T., & Sørensen, M. (2015). O Manual de Medidas de Seguridad Viaria. Madrid: Fundación mapfre.
- Elvik, R. (2017). The Value of Life - The rise and fall of a scientific research programme. Doctoral theses at NTNU, 2017:340. NTNU Norwegian University of Science and Technology.
- Elvik, R. Sørensen, M.W.J., & Nævestad, T.O. (2013). Factors influencing safety in a sample of marked pedestrian crossings selected for safety inspections in the city of Oslo. *Accident Analysis and Prevention*, 59, 64-70.
- Elvik, R., Vadeby, A., Hels, T., & van Schagen, I. (2019). Updated estimates of the relationship between speed and road safety at the aggregate and individual levels. *Accident Analysis & Prevention*, 123, 114-122.
- Elvik, R. & Goel, R. (2019). Safety-in-numbers: An updated meta-analysis of estimates. *Accident Analysis & Prevention*, 129, 136-147.
- Erke, A. & Elvik, R. (2006). Effektkatalog for trafikksikkerhetstiltak (Road safety measures: A catalogue of estimated effects). TØI-Report 851/2006. Oslo: Institute of Transport Economics.
- Farstad, E. (2018). Transportytelser i Norge 1946–2016. TØI-rapport 1613/2018. Oslo: Transportøkonomisk institutt.
- Finansdepartementet (2005). Veileder i samfunnsøkonomiske analyser.
- Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions*. Second Edition. John Wiley and Sons, New York, NY.
- Forsyth, E., G. Maycock & B. Sexton (1995). Cohort study of learner and novice drivers: Part 3, accidents, offences and driving experience in the first three years of driving. Project Report 111. Transport Research Laboratory, Crowthorne, Berkshire.
- Fosser, S., F. Sagberg & I.-A. Sætermo (1996). Atferdstilpasning til kollisjonspuget og blokkeringsfrie bremses. TØI Report 335. Oslo: Institute of Transport Economics.
- Fridstrøm, L. , Ifver, J., Ingebrigtsen, S., Kulmala & Thomsen, L.K. (1993). Explaining the variation in road accident counts. Report Nord 1993:35. Nordic Council of Ministers, Copenhagen.
- Fyhri, A., Sundfør, H.B., Bjørnskau, T., & Laureshyn, A. (2017). Safety in numbers for cyclists—conclusions from a multidisciplinary study of seasonal change in interplay and conflicts. *Accident Analysis and Prevention*, 105, 124-133.
- Gaudry, M. & S. Lassarre, S. (Eds) (2000): *Structural road accident models. The international DRAG family*. Pergamon Press, Oxford.
- Graham, P. L., & Moran, J. L. (2012). Robust meta-analytic conclusions mandate the provision of prediction intervals in meta-analysis summaries. *Journal of Clinical Epidemiology*, 65, 503-510.
- Grime, G. (1987). *Handbook of road safety research*. London, Butterworths.

- Guyatt, G.H., Oxman, A.D., Schunemann, H.J., Tugwell, P., Knottnerus, A. (2011A). GRADE guidelines: A new series of articles in the Journal of Clinical Epidemiology. *Journal of Clinical Epidemiology*, 64, 380-382.
- Guyatt, G. H., Oxman, A. D., Vist, G., Kunz, R., Brozek, J., Alonso-Coello, P., . . . Schunemann, H. J. (2011B). GRADE guidelines: 4. Rating the quality of evidence--study limitations (risk of bias). *J Clin Epidemiol*, 64(4), 407-415.
- Guyatt, G. H., Oxman, A. D., Montori, V., Vist, G., Kunz, R., Brozek, J., . . . Schunemann, H. J. (2011C). GRADE guidelines: 5. Rating the quality of evidence--publication bias. *J Clin Epidemiol*, 64(12), 1277-1282.
- Guyatt, G. H., Oxman, A. D., Kunz, R., Woodcock, J., Brozek, J., Helfand, M., . . . Schünemann, H. J. (2011G). GRADE guidelines: 7. Rating the quality of evidence— inconsistency. *Journal of Clinical Epidemiology*, 64(12), 1294-1302.
- Harbord, R. M., Egger, M., & Sterne, J. A. C. (2006). A modified test for small-study effects in meta-analyses of controlled trials with binary endpoints. *Statistics in Medicine*, 25(20), 3443-3457.
- Hauer, E. (1995). On exposure and accident rate. *Traffic Engineering and Control*, 36, 134-138.
- Hauer, E. (2011). Computing what the public wants. Some issues in road safety cost-benefit analysis. *Accident Analysis and Prevention*, 43, 151-164.
- Hauer, E. & A. S. Hakkert. (1988). Extent and Some Implications of Incomplete Accident Reporting. *Transportation Research Record*, 1185, 1-10.
- Hedlund, J. (2000). Risky business: Safety regulations, risk compensation, and individual behavior. *Injury Prevention*, 6, 82-90.
- Helmer, D., Savoie, I., Green, C., & Kazanjian, A. (2001). Evidence-based practice: extending the search to find material for the systematic review. *Bulletin of the Medical Library Association*, 89(4), 346-352.
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, 327, 557-562.
- Higgins, J. P. T., Thompson, S. G., & Spiegelhalter, D. J. (2009). A re-evaluation of random effects meta-analysis. *Journal of the Royal Statistical Society*, 172(1), 137-159.
- Høyе, A., Elvik, R., Sørensen, M., & Vaa, T. (2012). *Trafikksikkerhåndboken*. Oslo: Transportøkonomisk institutt.
- Høyе, A. (2014). Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge. TØI-rapport 1323/2014. Oslo: Transportøkonomisk institutt.
- Høyе, A. & Hesjevoll, I.S. (2016). Synlige sykklister - Bruk av sykkellys i Norge og effekt på ulykker. TØI-Rapport 1478/2016. Oslo: Transportøkonomisk institutt.
- Høyе, A. (2016A). Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge (2010-2015). TØI-rapport 1522/2016. Oslo: Transportøkonomisk institutt.
- Høyе, A. (2016B). How would increasing seat belt use affect the number of killed or seriously injured light vehicle occupants? *Accident Analysis and Prevention*, 88, 175-186.
- Høyе, A. & Nævestad, T.O. (2019). Veggtunneler – branner, ulykker og havarier. TØI-Rapport 1705/2019. Oslo: Transportøkonomisk institutt.

- Høye, A. (2019). Vehicle registration year, age, and weight – Untangling the effects on crash risk. *Accident Analysis & Prevention*, 123, 1-11.
- Høye, A., Bjørnskau, T. & Elvik, R. (2014). Hva forklarer nedgangen i antall drepte og hardt skadde i trafikken fra 2000 til 2012?. TØI-rapport 1299/2014. Oslo: Transportøkonomisk institutt.
- Hutchinson, T. P. (1984). Medical statistics on road accident injury in several countries. In Yagar, S. (Ed): *Transport Risk Assessment*, 43-76. University of Waterloo Press, Waterloo, Ontario.
- Hvoslef, H. (1980). Risikoanalyse av trafikksystemet i Haugesund 1970-76. En analyse av trafikkulykker og trafikkrisiko. Arbeidsdokument 30.9.1980 (prosjekt 4237, blandingstrafikk-gater). Oslo, Transportøkonomisk institutt.
- Hvoslef, H. (1995). Ulykkesfrekvenser som foreslås benyttet i vegplanlegging for bruer og tunneler. Notat datert 5. januar 1995, Vegdirektoratet, Trafikksikkerhetskontoret.
- Hvoslef, H. (1996). Trafiksikkerheten - kan den bli bättre? Vad händer när trafiken ökar? I Nordiska trafiksikkerhetsdagar 1995, 73-83. Rapport TemaNord 1996:511. København, Nordisk Ministerråd.
- Ioannidis, J. P., & Trikalinos, T. A. (2007). The appropriateness of asymmetry tests for publication bias in meta-analyses: a large survey. *CMAJ*, 176(8), 1091-1096.
- Johansson, R. & Naeslund, A-L. (1986). Upplevd och verklig olycksrisk - möjligheter till påverkan. TFB-rapport 1986:18. Stockholm, Transportforskningsberedningen.
- Kirkham, J. J., Dwan, K. M., Altman, D. G., Gamble, C., Dodd, S., Smyth, R., & Williamson, P. R. (2010). The impact of outcome reporting bias in randomised controlled trials on a cohort of systematic reviews. *Bmj*, 340, c365.
- Kommunikationsdepartementet. (1996). Nollvisionen. En rapport från två trafiksikkerhetsdagar nittonhundranittiosex. Stockholm.
- Krøyer, H. R. G. (2015). Is 30km/h a 'safe' speed? Injury severity of pedestrians struck by a vehicle and the relation to travel speed and age. *IATSS Research*, 39(1), 42-50.
- Langan, D., Higgins, J. P. T., Gregory, W., & Sutton, A. J. (2012). Graphical augmentations to the funnel plot assess the impact of additional evidence on a meta-analysis. *Journal of Clinical Epidemiology*, 65(5), 511-519.
- Lund, J. (2019). Helsevesenbasert skaderegistrering som verktøy for å forebygge trafikkulykker. Rapport. Oslo: Trygg Trafikk.
- Lund, J. & Bjerkedal, T. (2001). Permanent impairments, disabilities, and disability pensions related to accidents in Norway. *Accident Analysis and Prevention*, 33, 19-30.
- Mattsson, B. (1991). Samhällsekonomisk beräkningsmetod. Bilaga 1 till Samhällsekonomisk prioritering av trafiksikkerhetsåtgärder. TFB&VTI forskning/research 7:1, 1991. Stockholm og Linköping, Transportforskningsberedningen og Statens Väg- och Trafikinstitut.
- Mishan, E. J. & Quah, E. (2007). *Cost benefit analysis*. Fifth edition. Abingdon, Routledge.
- Moher, D., Liberati, A., Tetzlaff, J., & Altman, D. (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. *PLOS Medicine*, 6(7), 1-6.

- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1), 1-9.
- Norges Offentlige Utredninger (NOU). Samfunnsøkonomiske analyser. NOU 2012:16. Oslo, Departementenes servicecenter.
- Nuzzo, R. (2014). Statistical errors. *Nature*, 506, 150-152.
- OECD (1994). Methods for evaluating road safety measures. Paris, OECD, Road Research Group.
- Papaioannou, D., Sutton, A., Carroll, C., Booth, A., & Wong, R. (2010). Literature searching for social science systematic reviews: consideration of a range of search techniques. *Health Info Libr J*, 27(2), 114-122.
- Patil, S., Geedipally, S. R., & Lord, D. (2012). Analysis of crash severities using nested logit model—Accounting for the underreporting of crashes. *Accident Analysis & Prevention*, 45, 646-653.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2006). Comparison of Two Methods to Detect Publication Bias in Meta-analysis. *JAMA*, 295(6), 676-680.
- Peters, J. L., Sutton, A. J., Jones, D. R., Abrams, K. R., & Rushton, L. (2008). Contour-enhanced meta-analysis funnel plots help distinguish publication bias from other causes of asymmetry. *Journal of Clinical Epidemiology*, 61(10), 991-996.
- Riley, R. D., Higgins, J. P. T., & Deeks, J. J. (2011). Interpretation of random effects meta-analyses (Vol. 342).
- Rücker, G., Schwarzer, G., & Carpenter, J. (2008). Arcsine test for publication bias in meta-analyses with binary outcomes. *Statistics in Medicine*, 27(5), 746-763.
- Sælensminde, K. (2004). Metodikk for nytte-kostnadsanalyse av tiltak for gående og syklende. Arbeidsdokument SM/1570/2004. Oslo: Transportøkonomisk institutt.
- Sagberg, F., S. Fosser, & I.-A. Sætermo. (1997). An investigation of behavioural adaptation to airbags and antilock brakes among taxi drivers. *Accident Analysis and Prevention*, 29, 293-302.
- Sakshaug, K. & Johannessen, S. (2005). Revisjon av håndbok 115 “Analyse av ulykkessteder”: Verdier for normal ulykkesfrekvens og skadekostnad ved normal og god standard. SINTEF notat datert 2005-05-03. Trondheim, SINTEF teknologi og samfunn, transportsikkerhet og –informatikk.
- Samferdselsdepartementet (2016-2017). Meld. St. 33 (2016-2017). Melding til Stortinget Nasjonal transportplan 2018-2029.
- Samstad, H., Ramjerdi, F., Veisten, K., Navrud, S., Magnussen, K., Flügel, S., Killi, M., Harkjerr-Halse, Elvik, R., & San Martin, O. (2010). Den norske verdsettingsstudien – Sammendragsrapport. TØI-Rapport 1053/2010. Oslo: Transportøkonomisk institutt.
- Shadish, W. R. & C. K. Haddock. (1994). Combining estimates of effect size. In Cooper, H.; Hedges, L. V. (Eds): *The Handbook of Research Synthesis*, Chapter 18, 261-281. Russell Sage Foundation, New York, NY.
- Sivak, M., & Tsimhoni, O. (2008). Improving traffic safety: Conceptual considerations for successful action. *Journal of Safety Research*, 39(5), 453-457.

- Statens vegvesen (2018). Håndbok V712. Konsekvensanalyser. Oslo, Statens vegvesen, Vegdirektoratet.
- Sterne, J. A. C., Gavaghan, D., & Egger, M. (2000). Publication and related bias in meta-analysis: Power of statistical tests and prevalence in the literature. *Journal of Clinical Epidemiology*, 53, 1119-1129.
- Sterne, J. A. C., & Egger, M. (2001). Funnel plots for detecting bias in meta-analysis: Guidelines on choice of axis. *Journal of Clinical Epidemiology*, 54(10), 1046-1055.
- Sterne, J., Sutton, A.J., Ioannidis, J.P.A. et al. (2011). Recommendations for examining and interpreting funnel plot asymmetry in meta-analysis of randomized controlled trials. *BMJ*, 342.
- Sutton, A. J., Song, F., Gilbody, S. M., & Abrams, K. R. (2000). Modelling publication bias in meta-analysis: a review. *Statistical Methods in Medical Research*, 9(5), 421-445.
- Thompson, S. G., & Pocock, S. J. (1991). Can meta-analyses be trusted? *The Lancet*, 338(8775), 1127-1130.
- Thompson, S. G., Smith, T. C., & Sharp, S. J. (1997). Investigating underlying risk as a source of heterogeneity in meta-analysis. *Stat Med*, 16(23), 2741-2758.
- Thompson, S. G., & Sharp, S. J. (1999). Explaining heterogeneity in meta-analysis: a comparison of methods. *Statistics in Medicine*, 18(20), 2693-2708.
- Tingvall, C. (1997). The Zero Vision - A Road Transport System Free from Serious Health Losses. *Transportation, Traffic Safety and Health*, 37-57.
- Tingvall, C. & Haworth, N. (1999). Vision Zero - An ethical approach to safety. Paper presented to the 6th ITE International Conference Road Safety & Traffic Enforcement: Beyond 2000, Melbourne, 6-7 September 1999
- Anzures-Cabrera, J., & Higgins, J. P. T. (2010). Graphical displays for meta-analysis: An overview with suggestions for practice. *Research Synthesis Methods*, 1(1), 66-80.
- Vaa, T. (1991). Vurdering av sammenheng mellom opplevd utrygghet og ulykkesrisiko. Bilføreres opplevde utrygghet: Vurdering av måleproblemer. Arbeidsdokument TST/0264//91. Oslo, Transportøkonomisk institutt.
- Vaa, T., Glad, A., Sagberg, F., Bjørnskau, T. & Berge, G. (2002). Faktorer som påvirker kjørefart – Litteraturstudier og hypoteser. TØI rapport 601/2002. Oslo: Transportøkonomisk institutt.
- Vaa, T. (2016). Utrygghet og risiko i transport: En diskusjon. TØI-rapport 1525/2016. Oslo: Transportøkonomisk institutt.
- Veisten, K., Flügel, S., & Elvik, R. (2010). Den norske verdsettingsstudien. Ulykker - Verdien av statistiske liv og beregning av ulykkenes samfunnskostnader. TØI-Rapport 1053c/2010. Oslo: Transportøkonomisk institutt.
- Vägverket. (1996). Nollvisionen. En idé om trafiksäkerhet. Statens Vägverk, Borlänge.
- van Houwelingen, H. C., Arends, L. R., & Stijnen, T. (2002). Advanced methods in meta-analysis: Multivariate approach and meta-regression. *Statistics in Medicine*, 21(4), 589-624.
- Wasserstein, R. L., & Lazar, N. A. (2016). The ASA's Statement on p-Values: Context, Process, and Purpose. *The American Statistician*, 70(2), 129-133.

- Wilde, G. J. S. (1982). The theory of risk homeostasis: implications for safety and health. *Risk Analysis*, 2, 209-225.
- Wilde, G. J. S. (1986). Beyond the concept of risk homeostasis: suggestions for research and applications towards the prevention of accidents and life style related disease. *Accident Analysis and Prevention*, 18, 377-401.
- Wilde, G. J. S. (1988). Risk homeostasis theory and traffic accidents: propositions, deductions and discussion of dissension in recent reactions. *Ergonomics*, 31, 441-468.
- Wilde, G. J. S. (1994). *Target Risk. Dealing with the danger of death, disease and damage in everyday decisions*. PDE Publications, Toronto.
- Wold, H. (1995). *Trafikkulykker i planskilte kryss. Hovedoppgave i samferdselsteknikk, høsten 1995*. Trondheim, Norges Tekniske Høgskole, Institutt for samferdselsteknikk.
- Ziliak, S.T. & McCloskey, D.N. (2008). *The Cult of Statistical Significance*. University of Michigan Press.

Transportøkonomisk institutt (TØI) Stiftelsen Norsk senter for samferdselsforskning

TØI er et anvendt forskningsinstitutt, som mottar basisbevilgning fra Norges forskningsråd og gjennomfører forsknings- og utredningsoppdrag for næringsliv og offentlige etater. TØI ble opprettet i 1964 og er organisert som uavhengig stiftelse.

TØI utvikler og formidler kunnskap om samferdsel med vitenskapelig kvalitet og praktisk anvendelse. Instituttet har et tverrfaglig miljø med rundt 90 høyt spesialiserte forskere.

Instituttet utgir tidsskriftet Samferdsel på internett og driver også forskningsformidling gjennom TØI-rapporter, artikler i vitenskapelige tidsskrifter, samt innlegg og intervjuer i media. TØI-rapportene er gratis tilgjengelige på instituttets hjemmeside www.toi.no.

TØI er partner i CIENS Forskningscenter for miljø og samfunn, lokalisert i Forskningsparken nær Universitetet i Oslo (se www.ciens.no). Instituttet deltar aktivt i internasjonalt forsknings-samarbeid, med særlig vekt på EUs rammeprogrammer.

TØI dekker alle transportmidler og temaområder innen samferdsel, inkludert trafiksikkerhet, kollektivtransport, klima og miljø, reiseliv, reisevaner og reiseetterspørsel, arealplanlegging, offentlige beslutningsprosesser, næringslivets transport og generell transportøkonomi.

Transportøkonomisk institutt krever opphavsrett til egne arbeider og legger vekt på å opptre uavhengig av oppdragsgiverne i alle faglige analyser og vurderinger.

Besøks- og postadresse:

Transportøkonomisk institutt
Gaustadalléen 21
NO-0349 Oslo

22 57 38 00
toi@toi.no
www.toi.no