

Denne artikkel er publisert i det elektroniske tidsskrift
Udvalgte Artikler fra Trafikdage på Aalborg Universitet
(Selected Proceedings from the Annual Transport
Conference at Aalborg University)

ISSN 1903-1092

www.trafikdage.dk/artikelarkiv

Modtaget 12.10.2011

Accepteret den 07.02.2012



Hva bør gjøres når en evaluering ikke kan anvende beste metode?

Rune Elvik (re@toi.no)

Transportøkonomisk institutt og Aalborg Universitet

Abstrakt

Utgangspunktet for dette paperet er at det ikke alltid er mulig å gjennomføre en før-og-etterundersøkelse av et trafikksikkerhetstiltak med den beste metode. Spørsmålet er hva man i så fall skal gjøre. Man kan enten velge å gjøre en evaluering av tiltaket med en "nest-beste" metode, eller man kan avstå fra å gjøre en undersøkelse i det hele tatt. For å kunne velge mellom å gjøre en nest-beste studie, eller ingen studie i det hele tatt, bør det tas utgangspunkt i de krav som stilles til gode evalueringer av trafikksikkerhetstiltak og man bør kritisk vurdere hvor mange av disse krav en evalueringsstudie kan forventes å oppfylle. Dersom en studie kan forventes å oppfylle de fleste av kravene til en god undersøkelse, bør man gjennomføre studien. En nest-beste studie kan ofte styrkes ved at man innhenter en del lett tilgjengelige supplerende opplysninger om forhold som erfaringsmessig påvirker trafikksikkerheten, slik som folketall og økonomiske konjunkturer. I dette paperet illustreres denne fremgangsmåten ved å studere virkninger av kampanjer mot promillekjøring blant unge førere i Nordjylland i perioden 1988-1993, 1995 og etter 1996. Ved å innhente opplysninger om folketall i amtene og om utviklingen av realinntekter, er det mulig å evaluere kampanjenes virkninger ved hjelp av en multivariat negativ binomial regresjonsmodell. Selv om denne studien ikke oppfyller alle krav til en god studie, oppfyller den disse kravene tilstrekkelig godt til at man kan gi sterkere grunner for å tro på resultatene av undersøkelsen enn for ikke å tro på disse resultatene.

1 Bakgrunn og problemstilling

De siste 15-20 år er metodene for ikke-eksperimentelle før-og-etterundersøkelser av trafikksikkerhetstiltak utviklet betydelig (Hauer 1997, Hauer et al 2002, Persaud og Lyon 2007, Highway Safety Manual 2010). Mange betrakter i dag empirisk Bayes metode som den beste metoden for å gjøre før-og-etterundersøkelser av trafikksikkerhetstiltak. Denne metoden er utviklet over lang tid og finnes i dag i flere versjoner. I sin mest avanserte form stiller metoden store krav til datagrunnlag og statistiske analyser. Den krever at det utvikles en ulykkesmodell som forklarer så mye av den systematiske variasjonen i ulykkestall som mulig. Videre krever metoden at man kjenner trafikkmengden og ulykkestallene hvert år i den perioden man studerer. Gjennom en god empirisk Bayes undersøkelse oppnår forskeren kontroll over en del viktige

feilkilder i før-og-etterundersøkelser, slik som regresjon mot gjennomsnittet, langsiktige trender i antall ulykker og endringer i trafikkmengde.

De nødvendige data til å utføre en empirisk Bayes før-og-etterundersøkelse foreligger ikke alltid. Spørsmålet er hva man da bør gjøre. Bør man, når den beste metode ikke kan anvendes, avstå fra å gjøre en evaluering av et tiltak? Eller bør man forsøke å gjøre en undersøkelse med "nest beste" metode, der man så langt tilgjengelige data gjør det mulig forsøker å tilnærme seg den beste metode?

Svaret på dette spørsmålet avhenger av hvor god en nest-beste undersøkelse kan bli. Hvis man kan gjøre en undersøkelse der man oppnår en viss grad av kontroll over de samme feilkilder som i en empirisk Bayes undersøkelse, kan en nest-beste undersøkelse forsvares. Hvis man derimot ikke kan oppnå kontroll over disse feilkilder, er det umulig å vite om de endringer man finner i ulykkestall i hovedsak skyldes tiltaket eller de feilkilder man ikke har kontroll over.

I dette paperet vil det bli gitt et eksempel på en nest-beste evalueringsstudie og forklart hvordan man kan bedømme kvaliteten på en slik undersøkelse på grunnlag av kriterier for kvalitet. Før studien presenteres, vil de viktigste krav til en god evaluering av virkninger av et trafikksikkerhetstiltak kort bli drøftet.

2 Krav til gode evalueringer av trafikksikkerhetstiltak

Det foreligger en relativt bred enighet om hva som er de viktigste krav til en god ikke-eksperimentell evaluering av et trafikksikkerhetstiltak (Hauer 2005, Highway Safety Manual 2010, Elvik 2007, 2008, 2011). Tabell 1 oppsummerer disse kravene i form av ni kriterier for å fastslå at det er en årsakssammenheng mellom innføring av et trafikksikkerhetstiltak og endringer i trafikksikkerhet etter at tiltaket er innført.

Ikke alle de ni kriteriene som beskrives i tabell 1 er like viktige. De viktigste kriteriene er 1, 4, 5, 7 og 8 (i de tilfeller kriterium 8 kan anvendes). Det første kriteriet sies ganske enkelt at en virkning skal kunne påvises. Mange vil stille et krav om at denne virkningen skal være statistisk signifikant, men det kan lett gis eksempler på virkninger av stor praktisk interesse som ikke var statistisk signifikante (Hauer 2004). Kravet er derfor formulert slik at statistisk signifikante endringer i ulykkestall ikke er nødvendig, men må regnes som en styrke hvis de kan påvises.

I de tilfeller der de påviste endringer i ulykkestall ikke er statistisk signifikante, får kriteriene 2 og 3 økt betydning. Kriterium 2 sier at de endringer man statistisk kan knytte til innføring av et trafikksikkerhetstiltak bør være større enn andre endringer i ulykkestall som kan påvises i datamaterialet, for eksempel de gjennomsnittlige endringer i ulykkestall fra år til år i perioden før tiltaket ble innført. Kriterium 3 sier at en årsakssammenheng er kjennetegnet av en høy grad av lovmessighet og stabilitet. Dette kriteriet innebærer at dersom man kan dele inn datamaterialet i grupper, som alle er likt påvirket av det samme tiltaket, så bør man finne de samme endringer i ulykkestall i alle grupper – innenfor rammene av tilfeldig variasjon i ulykkestall. I mange undersøkelser gir imidlertid tilfeldig variasjon i ulykkestall et stort bidrag til de endringer man kan observere. Lovmessighetskriteriet kan følgelig ikke anvendes for strengt i trafikksikkerhetsstudier.

Kriterium 4 sier at årsaksretningen må være entydig. Det betyr at av to variabler, A og B, må man kunne avgjøre om A er årsak til B eller omvendt. Årsakspåvirkninger kan i noen tilfeller gå i begge retninger. I evalueringer av trafikksikkerhetstiltak er man imidlertid kun interessert i årsakspåvirkning fra tiltaket til endringer i ulykkestall.

Selv om hovedinteressen er knyttet til hvilken virkning et tiltak har på ulykkene, kan man i prinsippet ikke utelukke at ulykkeshistorien påvirker sannsynligheten for at et tiltak innføres. Dette kan i det minste tenkes i to tilfeller: (1) Når det over lengre tid har vært en ugunstig utvikling av antall ulykker; (2) Når man et bestemt år observerer et unormalt høyt antall ulykker. I begge disse tilfeller er endringer i antall ulykker før et tiltak innføres en mulig feilkilde (konkurrerende forklaring) i en før-og-etterundersøkelse av et trafikksikkerhetstiltak. De viktigste mulige feilkilder i slike undersøkelser (kriterium 5) er:

Tabell 1: Kriterier som må oppfylles for å påvise en årsakssammenheng mellom et trafikksikkerhetstiltak og endringer i trafikksikkerhet

Kriterium	Operasjonell definisjon av kriteriet
1: Det skal være en statistisk sammenheng mellom innføring av tiltaket og endringer i ulykkestall (en virkning skal finnes).	Endring i antall ulykker fra før til etter innføring av et tiltak; endringene bør fortrinnsvis være statistisk signifikante, men kan vise reelle virkninger selv om de ikke er det.
2: Sammenhengen mellom et tiltak og endringer i ulykkestall bør være sterkere enn andre sammenhenger som finnes i datamaterialet.	Endringer i antall ulykker som har statistisk sammenheng med et tiltak bør være større enn endringer i antall ulykker som har sammenheng med andre forhold.
3: Sammenhengen mellom et tiltak og endringer i ulykkestall bør være konsistent enten i ulike deler av datamaterialet eller over tid.	Dersom man kan observere endringer i ulykkestall i ulike grupper av datamaterialet bør disse være tilnærmet like store (lik årsak gir lik virkning).
4: Årsaksretningen skal være entydig; det vil si at det kan avgjøres hvilken av to variabler som er årsak til endringer i den andre.	Hvis man først innfører og deretter avskaffer et tiltak, bør endringene i ulykkestall gå i motsatt retning i de to tilfeller.
5: Sammenhengen mellom tiltaket og endringer i ulykkestall skal bestå etter kontroll for de viktigste alternative forklaringer på endringene.	Det bør kontrolleres statistisk for de viktigste feilkilder, det vil si alternative forklaringer på endringer i ulykkestall, og de endringer i ulykkestall som kan knyttes til tiltaket bør kunne påvises etter man har eliminert de viktigste feilkilder.
6: Det bør kunne påvises en årsaksmekanisme som forklarer sammenhengen mellom tiltaket og endringer i ulykkestall.	Det bør kunne identifiseres en eller flere risikofaktorer tiltaket påvirker; endringer i disse risikofaktorer bør kunne påvises og knyttes til innføring av tiltaket.
7: Sammenhengen mellom tiltaket og endringer i ulykkestall bør ha støtte fra, eller i det minste ikke motsi, veletablert vitenskapelig kunnskap.	De påviste endringer i ulykkestall bør fremstå som plausible i lys av fysiske lovmessigheter, biomekanisk kunnskap eller kunnskap om menneskers persepsjon og informasjonbearbeiding.
8: Dersom tiltaket forekommer i ulike doser, bør det finnes en dose-responsammenheng mellom dosen av tiltaket og størrelsen på endringer i ulykkestall.	Jo større mengde av tiltaket man innfører, eller jo høyere standard av tiltaket man anvender, desto større bør endringene i ulykkestall være; kriteriet kan kun anvendes i de tilfeller variasjoner i tiltakets dose eller standard kan måles tilfredsstillende.
9: Dersom tiltaket bare innføres i en bestemt, klart definert gruppe, bør endringer i ulykkestall primært forekomme i denne gruppen og ikke i grupper tiltaket ikke er rettet mot.	Endringer i ulykkestall bør kunne påvises i tiltakets målgruppe, men ikke utenfor denne; kriteriet kan kun anvendes når et tiltak har en klart definert målgruppe og kun er innført i denne.

1. Regresjon mot gjennomsnittet, som i de fleste tilfeller innebærer at tilfeldig høye ulykkestall reduseres i en etterfølgende periode.
2. Langsiktige trender i ulykkestall, som betyr at antall ulykker viser en tendens til å øke eller synke over tid.
3. Endringer i trafikkmengde som ikke skyldes tiltaket.

4. Andre hendelser som inntreffer samtidig med innføring av tiltaket og som kan tenkes å ha de samme virkninger på antall ulykker som tiltaket.
5. Samtidig innføring av mer enn ett tiltak som påvirker de samme ulykkene i en situasjon der man ønsker å måle virkningene av hvert av de tiltak som er innført.
6. En tendens til at innføring av tiltak på ett sted fører til at ulykkene forflytter seg til andre steder der tiltak ikke er innført (ulykkesmigrasjon).

De tre førstnevnte feilkildene på denne listen er de viktigste. Enhver før-og-etterundersøkelse av et trafikksikkerhetstiltak bør kontrollere for disse feilkildene. Når det gjelder de tre sistnevnte feilkildene, bør man vurdere sannsynligheten for at de er til stede.

Ethvert trafikksikkerhetstiltak påvirker trafikksikkerheten ved å endre en eller flere risikofaktorer som har sammenheng med antall ulykker eller ulykkesens alvorlighetsgrad. Kriterium 6 sier at man bør kunne identifisere de risikofaktorer et tiltak påvirker og måle endringer i disse. Dette gjøres i de færreste tilfeller. I de tilfeller der man måler endringer i en eller flere risikofaktorer, er det ikke alltid slik at resultatene av undersøkelsen dermed blir lettere å tolke (Elvik 2003). Man bør derfor legge noe mer vekt på kriterium 7 enn på kriterium 6. Kriterium 7 sier at resultatene av en undersøkelse bør ha støtte i veletablert vitenskapelig teori, eller i det minste ikke motsi slik teori. Teorigrunnlaget for trafikksikkerhetsforskningen er ikke særlig godt utviklet; kriteriet gir derfor begrenset støtte ved tolkning av resultater av undersøkelser. Man kan likevel i en del tilfeller betrakte noen resultater som teoretisk mer plausible enn andre. Det er, for eksempel, mer plausibelt at lavere fart er forbundet med færre og mindre alvorlige ulykker enn det motsatte. Det er mer rimelig at dem som bruker bilbelte har større sannsynlighet for å overleve en ulykke enn at ikke-brukere av bilbelte har det. Videre er det mer sannsynlig at vegbelysning reduserer antall ulykker i mørke enn at tiltaket øker dem.

Kriterium 8, dose-responsammenheng, er svært nyttig i de tilfeller man kan benytte det. I mange tilfeller er imidlertid dosen eller kvaliteten på et trafikksikkerhetstiltak ukjent og da kan kriteriet ikke benyttes. Kriterium 9, virkning bare i tiltakets målgruppe, er også nyttig når det kan benyttes.

Selv om enhver tilordning av vektorer til kriteriene er vilkårlig, kan en vektning av kriteriene likevel være hensiktsmessig for å formalisere bruken av dem. Kriteriene er tilordnet følgende vektorer:

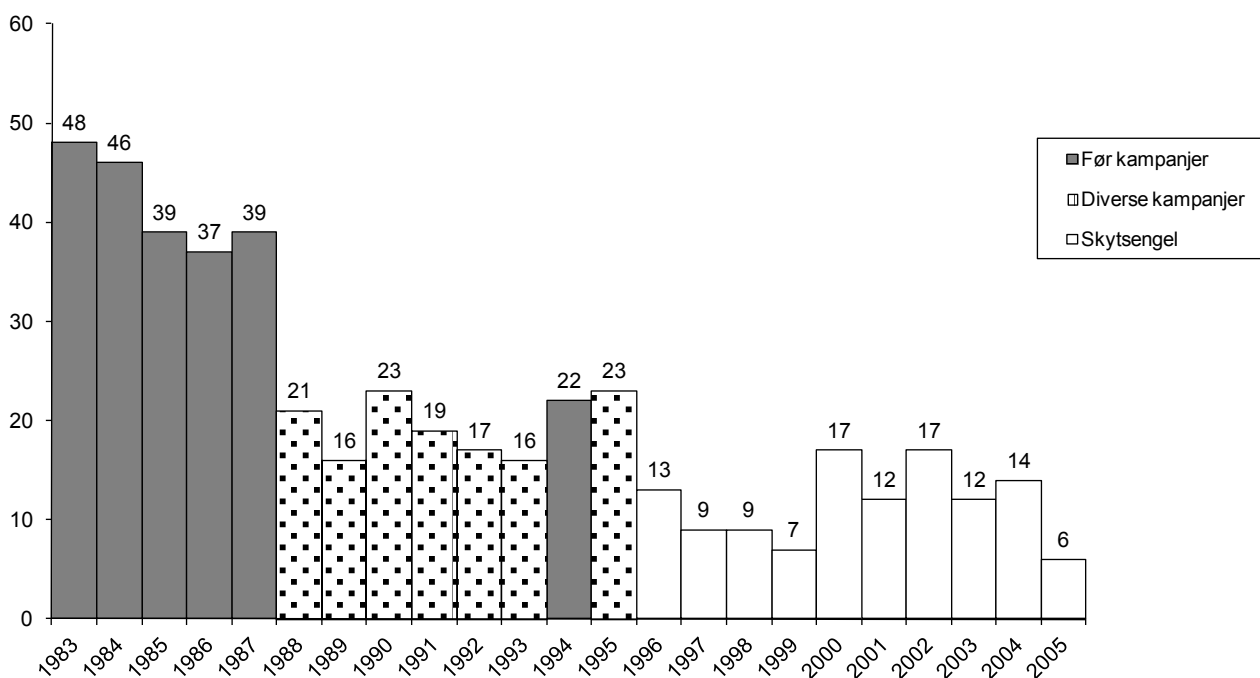
Kriterium 1:	10 poeng
Kriterium 2:	5 poeng
Kriterium 3:	5 poeng
Kriterium 4:	10 poeng
Kriterium 5:	60 poeng
Kriterium 6:	5 poeng
Kriterium 7:	5 poeng
Kriterium 8:	10 poeng
Kriterium 9:	10 poeng

De første syv kriteriene kan benyttes i alle undersøkelser og kan gi opp til 100 poeng hvis alle er oppfylt. Kriteriene 8 og 9 kan gi inntil 20 tilleggspoeng. Poengskalaen er ment å være kontinuerlig. Således kan en undersøkelse oppnå, for eksempel 45 av 60 mulige poeng for kriterium 5 eller 7 av 10 poeng for kriterium 4. Bruken av skalaen vil bli illustrert med utgangspunkt i en konkret undersøkelse.

3 Evaluering av kampanjer mot promillekjøring blant unge førere i Nordjylland

I det følgende vil en undersøkelse om virkninger av kampanjer mot promillekjøring blant unge førere i Nordjylland bli benyttet som eksempel på en nest-beste evaluering. I utgangspunktet bygger denne evalueringen kun på de data som er oppgitt i figur 1 (Studsholt 1990, Bech 2006).

Antall ulykker med påvirkede førere 17-20 år i Nordjylland i mai, juni og juli



Figur 1: Ulykker blant unge førere med promille i Nordjylland 1983-2005

Ulykkestallene som er oppgitt i figur 1 kan deles inn i fem perioder. I perioden 1983-1987 (1) var det ingen kampanjer. Fra og med 1988 til og med 1993 (2) ble det gjennomført diverse kampanjer. I året 1994 (3) var det ingen kampanje. I 1995 (4) var det på ny en kampanje mot promillekjøring. I årene fra og med 1996 til 2005 (5) er Skytsengel-kampanjen mot promillekjøring blant unge førere gjennomført. Spørsmålet er hvordan man kan få gjennomført en relativt godt kontrollert evaluering av kampanjene i perioden 1988-1993 samt 1995 og av Skytsengel-kampanjen fra 1996 til 2005 på grunnlag av disse dataene?

Det første spørsmål en slik evaluering må besvare er: Hva ville ha skjedd dersom de ulike kampanjene ikke var blitt gjennomført? For å svare på dette spørsmålet, må man beregne hvordan ulykkesutviklingen ville ha vært dersom kampanjene ikke var blitt gjennomført. Dette kan gjøres på flere måter. Man kan for eksempel føye en trendlinje til dataene for perioden 1983-1987 og forlengende denne trendlinjen til de etterfølgende år. Det er to problemer knyttet til dette. For det første har man kun fem datapunkter, noe som gjør enhver fremskrivning av trendlinjen meget usikker. For det andre kan mer enn en trendlinje føyes til datapunktene og de ulike trendlinjene gir meget ulike fremskrivninger av forventet ulykkesutvikling.

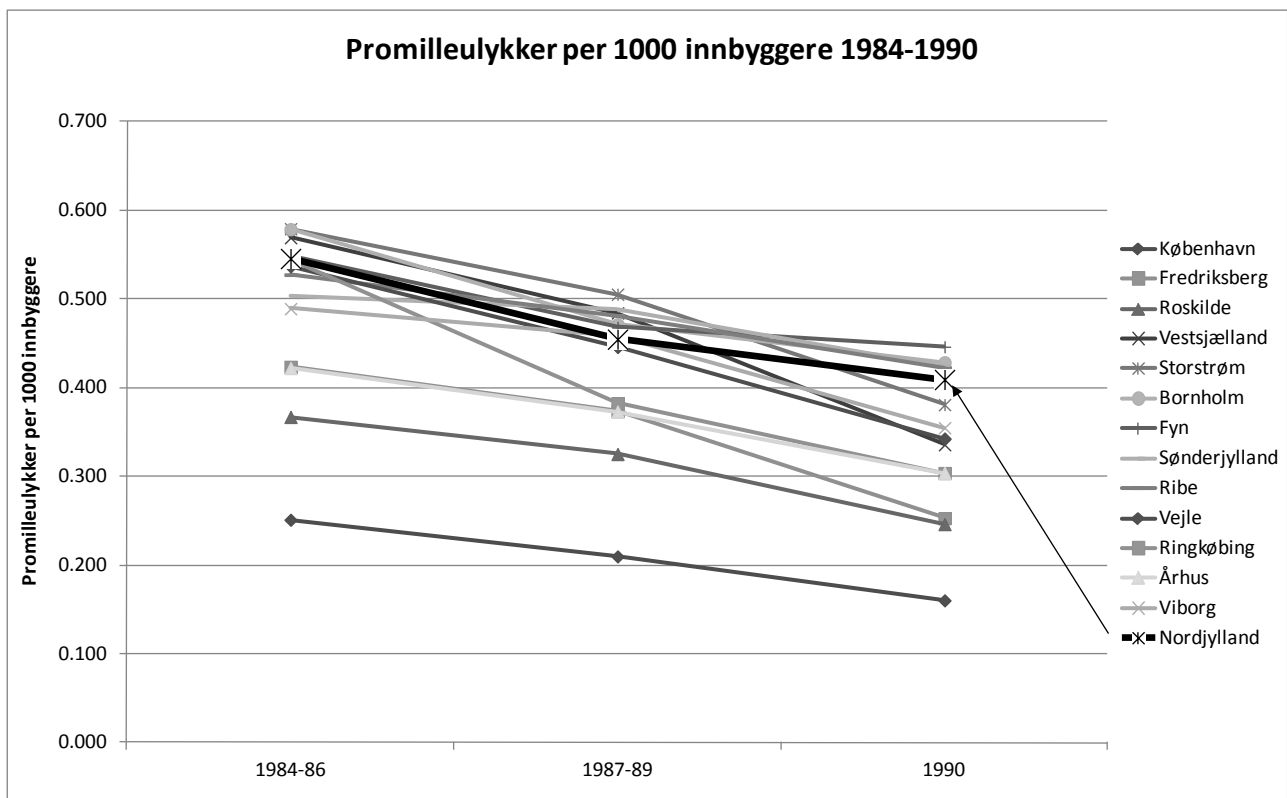
En annen tenkelig metode er å benytte ulykker med promilleførere i resten av Danmark som kontrollgruppe. Det er imidlertid ikke opplagt at ulykker med promilleførere i resten av Danmark gir en god prediksjon av utviklingen fra år til år i antall ulykker med promilleførere i Nordjylland. For årene fra og med 1984 til og med 1987 er de predikerte ulykkestallene for Nordjylland, når man benytter promilleulykker i resten av Danmark som kontrollgruppe, 48 (1984), 48 (1985), 37 (1986) og 33 (1987). Disse tallene avviker en del fra de registrerte tallene og tyder på at resten av Danmark ikke fungerer godt som en kontrollgruppe for Nordjylland.

Ved å innhente lett tilgjengelige supplerende opplysninger, kan man gjøre en bedre evaluering av kampanjene enn ved kun å bygge på de opplysninger som er gitt i figur 1. Følgende lett tilgjengelige opplysninger er innhentet:

1. Antall promilleulykker i hele Danmark hvert år fra 1983 til 2005. Dette er hentet fra den årlige statistikken over veitrafikkulykker.

2. Antall promilleulykker fordelt på amt i årene 1984-86, 1987-89 og 1990. Disse opplysningene er også hentet fra statistikken over veitrafikkulykker.
3. Antall innbyggere i de ulike amt. For Nordjylland er opplysninger om antall innbyggere i alderen 17-20 år innhentet hvert år fra 1983 til 2005.
4. Bruttonasjonalprodukt per innbygger regnet i faste priser (hentet fra Verdensbankens database over statistikk om økonomiske forhold i ulike land).

Ved hjelp av disse opplysninger vil det først bli undersøkt om antallet promilleulykker per 1.000 innbyggere var unormalt høyt i Nordjylland i perioden før de ulike kampanjene startet. Dette vil fortelle om det er grunn til å tro at regresjon mot gjennomsnittet har påvirket ulykkestallene i Nordjylland. Figur 2 viser antall promilleulykker per 1.000 innbyggere i 14 amt i periodene 1984-86, 1987-89 og 1990. Den første av disse periodene og deler av den andre kan betraktes som før-periode for kampanjene i Nordjylland.



Figur 2: Antall promilleulykker per 1.000 innbyggere i 14 amt 1984-86, 1987-89 og 1990

Nordjylland lå femte høyest blant 14 amt i perioden 1984-86 og syvende høyest i perioden 1987-89. Fra og med 1988 begynte kampanjene mot promillekjøring i Nordjylland. Denne statistikken viser at det ikke er grunnlag for å si at antallet promilleulykker i Nordjylland i perioden før kampanjene var unormalt høyt. Det er følgelig ikke grunn til å tro at de årlige ulykkestallene i perioden 1983-1987 er nevneverdig påvirket av regresjonseffekter. Beregninger tyder på at en eventuell regresjonseffekt høyst er i størrelsesorden ca 1 prosent. I det følgende betraktes derfor ulykkestallene i før-perioden som forventningsrette.

4 Negativ binomial regresjonsanalyse av kampanjenes virkninger på antall ulykker

Det er utført en evaluering av virkningene på antall ulykker av kampanjene mot promillekjøring i Nordjylland ved hjelp av en negativ binomial regresjonsanalyse. Datagrunnlaget for denne analysen fremgår av tabell 2.

Tabell 2: Datagrunnlag for negativ binomial regresjonsanalyse av virkninger av kampanjer mot promillekjøring i Nordjylland

Promille-ulykker med unge førere i Nordjylland	Promille-ulykker i resten av Danmark	År (1983 = 1; 2005 = 23)	Dummy for kampanjer 1988-1993 og 1995	Dummy for Skytsengel-kampanje 1996-2005	Befolkning i alderen 17-20 år i Nordjylland	BNP per innbygger (US dollar 2000-priser)
48	2267	1	0	0	31977	20815
46	2246	2	0	0	32921	21691
39	2364	3	0	0	32630	22555
37	2241	4	0	0	31546	23639
39	2015	5	0	0	30307	23679
21	1908	6	1	0	28446	23632
16	1959	7	1	0	27888	23753
23	1590	8	1	0	28055	24102
19	1472	9	1	0	27972	24349
17	1517	10	1	0	27893	24754
16	1407	11	1	0	27955	24641
22	1327	12	0	0	27376	25922
23	1259	13	1	0	26788	26599
13	1235	14	0	1	26331	27177
9	1244	15	0	1	25750	27928
9	1117	16	0	1	25065	28441
7	1195	17	0	1	24094	29070
17	1255	18	0	1	23363	29993
12	1128	19	0	1	22412	30104
17	1213	20	0	1	21954	30136
12	1177	21	0	1	22179	30180
14	1071	22	0	1	22330	30793
6	864	23	0	1	22727	31450

Det ble føyd en negativ binomial regresjonsmodell med promilleulykker med unge førere i Nordjylland som avhengig variabel og år, dummy for kampanjer 1988-1993 og 1995, dummy for Skytsengel-kampanjen 1996-2005, folketall i alderen 17-20 år i Nordjylland og bruttonasjonalprodukt per innbygger (i faste 2000-priser) som forklaringsvariabler. Sistnevnte variabel er ment å fange opp virkninger av økonomiske konjunkturer som kan endres fra år til år og påvirke antall ulykker blant unge førere. Deretter ble en tilsvarende modell føyd med antall promilleulykker i resten av Danmark som avhengig variabel. Dersom kampanjene i Nordjylland har medvirket til å redusere antall ulykker forventes det at:

1. Koeffisientene for dummyvariablene som representerer kampanjene er negative i analysen der ulykker i Nordjylland er avhengig variabel.
2. Koeffisientene for disse variablene er ikke statistisk signifikante i analysen der ulykker i resten av Danmark er avhengig variabel.

Tabell 3 viser de estimerte koeffisienter og den statistiske usikkerhet i disse.

Tabell 3: Estimerte koeffisienter i negativ binomial regresjonsanalyse

Variabler	Modell for ulykker i Nordjylland		Modell for ulykker i resten av Danmark	
	Koeffisient	Standardfeil	Koeffisient	Standardfeil
Konstantledd	5,118	3,344	7,838	0,527
År	-0,068	0,055	-0,084	0,008
Dummy for kampanjer	-0,566	0,172	-0,083	0,026
Skytsengeldummy	-0,738	0,269	-0,061	0,039
Folketall 17-20 år	-0,00005045	0,00007108	-0,00003713	0,00001016
BNP per innbygger	0,00001736	0,00010000	0,00005736	0,00001624

Modellen for ulykker i Nordjylland har en meget god føyning til data. Koeffisientene for kampanjevariablene tyder på at kampanjene har bidratt til å redusere antall ulykker. Man kan beregne hvor stor virkningen er ved å beregne predikert ulykkestall når variablene for kampanjene utelates fra modellen. Dette gir følgende resultater:

Predikert ulykkestall for årene 1988-1993 og 1995 uten kampanjedummy: 237,46
 Predikert ulykkestall for årene 1988-1993 og 1995 med kampanjedummy: 134,83

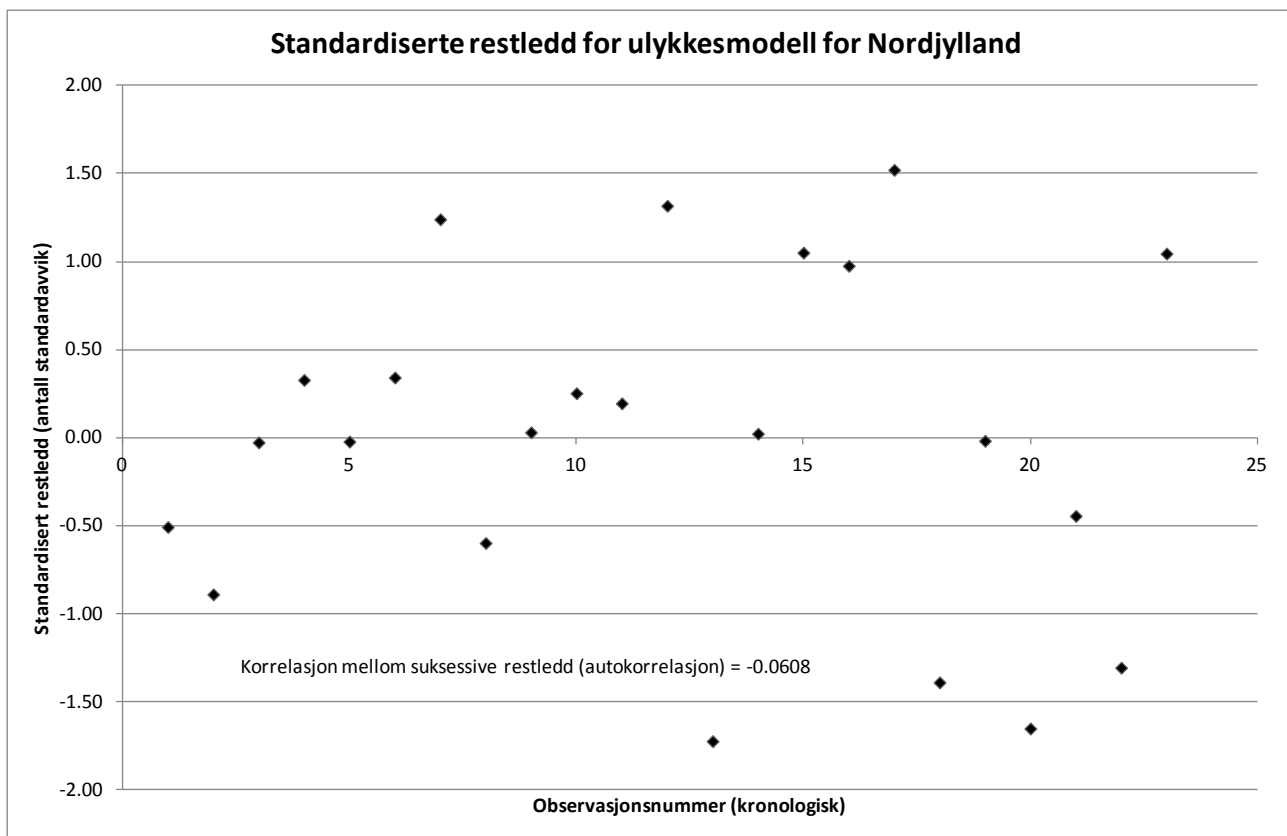
Differansen utgjør 102,63 ulykker. Standardfeilen til denne differansen, beregnet ut fra en antakelse om de predikerte ulykkestallene er Poisson-fordelte er 19,29. Den beregnede nedgang i ulykkestall er følgelig klart statistisk signifikant. Den tilsvarer en nedgang på 43 prosent. For Skytsengel-kampanjen fra 1996 til 2005 gir beregningene følgende resultater:

Predikert ulykkestall 1996-2005 uten kampanjedummy: 242,15
 Predikert ulykkestall 1996-2005 med kampanjedummy: 115,76

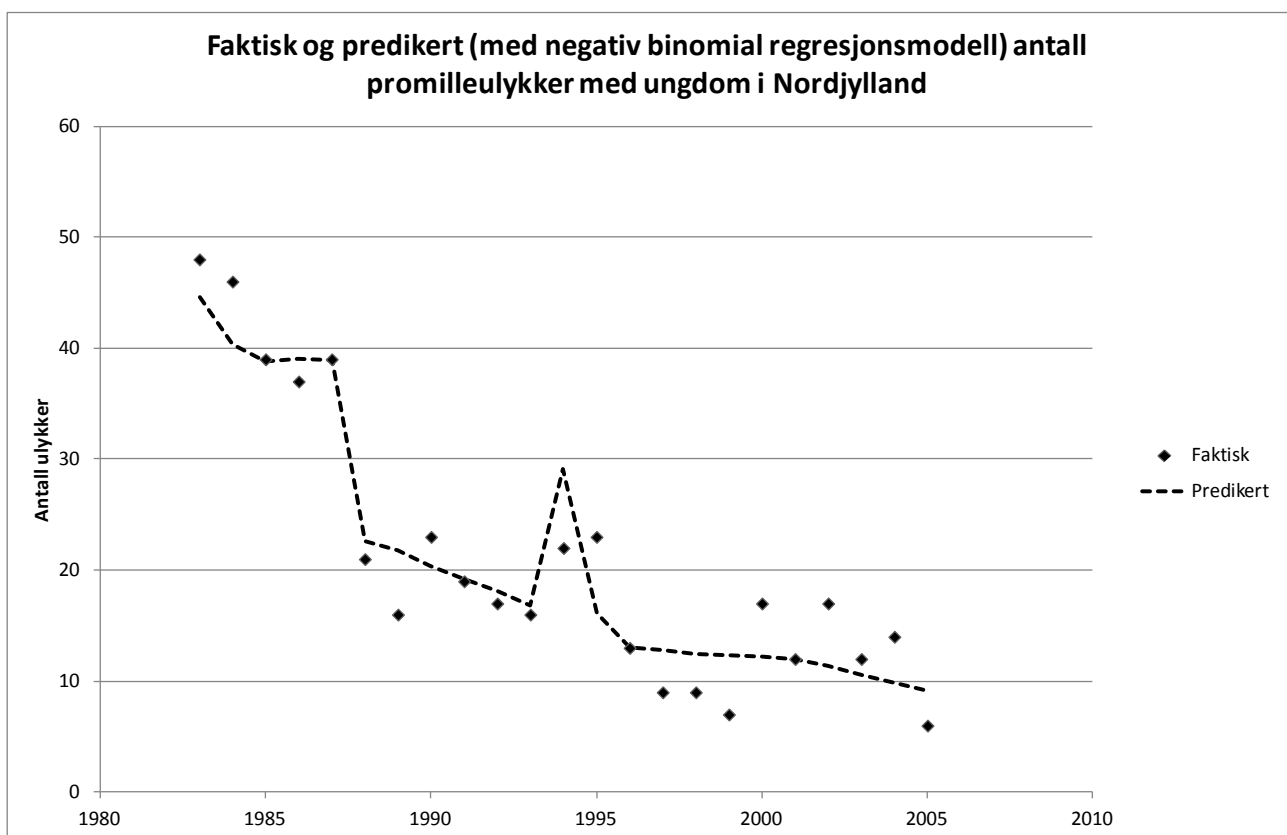
Differansen er på 126,38 ulykker. Standardfeilen til denne differansen er på 18,92 ulykker. Nedgangen er med andre ord statistisk signifikant og tilsvarer en nedgang i forventet ulykkestall på 52 prosent.

Figur 3 viser standardiserte restledd for modellen der ulykker i Nordjylland er avhengig variabel. De standardiserte restleddene viser avvikene mellom predikerte og registrerte ulykkestall regnet i form av antall standardavvik. Alle restledd ligger innenfor et 95 prosent konfidensintervall (pluss eller minus 1,96 standardavvik). Restleddene er symmetrisk fordelt rundt de predikerte verdier, med 11 negative og 12 positive restledd. Figur 4 viser registrerte og predikerte ulykkestall for de årene undersøkelsen omfatter.

Koeffisientene i modellen der ulykker i resten av Danmark er benyttet som avhengig variabel er forbausende like koeffisientene i modellen med ulykker i Nordjylland som avhengig variabel. Koeffisientene som representerer kampanjene har lavere tallverdi enn i modellen for Nordjylland. I modellen for resten av Danmark er koeffisienten for Skytsengel-kampanjen ikke statistisk signifikant. Ved å inkludere koeffisienten for dummyvariabelen for kampanjer i perioden 1988-1993 og 1995 predikeres ulykkestallene bedre enn ved å utelate denne koeffisienten i seks av syv år. For de ti årene Skytsengel-kampanjen inngår i datamaterialet er prediksjonene best med koeffisienten for denne kampanjen i syv av ti år, best uten denne koeffisienten i tre av ti år.



Figur 3: Standardiserte restledd for ulykkesmodell med ulykker i Nordjylland som avhengig variabel



Figur 4: Registrert og predikert antall promilleulykker med unge førere i Nordjylland 1983-2005

Dette tyder på at kriteriet om at man kun skal finne en virkning i tiltakets målgruppe, ikke utenfor denne, ikke fullt ut er oppfylt i dette tilfellet. På den annen side kan det tenkes at tilsvarende kampanjer har vært utført også i andre deler av Danmark enn Nordjylland. Da vil resten av Danmark ikke vise den ulykkesutvikling man kan forvente når de aktuelle kampanjene ikke finner sted.

5 Hvor godt oppfyller evalueringen kriteriene for en god undersøkelse?

I tabell 4 er det angitt hvor godt evalueringen kan sies å oppfylle de krav til en god undersøkelse som er beskrevet i avsnitt 2.

Tabell 4: Oppfyllelse av kravene til en god evaluering

Kriterium	Oppfyllelse i denne undersøkelsen	Poeng
1: Statistisk påvisbare endringer i ulykkestall	Oppfylt. Signifikante reduksjoner påvist	10
2: Sterkere effekt av tiltak enn andre forhold	Oppfylt. Koeffisientene for kampanjene har lavere p-verdi enn andre koeffisienter i modellen for Nordjylland	5
3: Konsistent effekt	Oppfylt. Lavere ulykkestall enn siste år før i seks av syv år med kampanje 1988-1993 og 1995 og alle ti fra 1996 til 2005	5
4: Entydig årsaksretning	Oppfylt. Nedgang fra 1987 til 1988, økningen fra 1993 til 1994, men ikke nedgang i registrert ulykkestall fra 1994 til 1995. Modellprediksjonene er helt konsistente	8
5: Kontroll for andre forklaringer	Delvis oppfylt. Modellen kontrollerer for (1) langsiktige trender og (2) endringer i eksponering (gjennom folketall). Det er dessuten lite sannsynlig at regresjonseffekt er en feilkilde	45
6: Kjent årsaksmechanisme	Ikke oppfylt. Endringer i omfanget av promillekjøring er ikke målt	0
7: Samsvar med veletablert teori	Ikke oppfylt. Det foreligger ingen veletablert teori om tiltakets virkninger	0
8: Dose-responsammenheng	Ikke oppfylt. Det foreligger ikke brukbare opplysninger om kampanjenes omfang	0
9: Virkning bare i tiltakets målgruppe	Delvis oppfylt, men tilsynelatende kan effekter påvises også utenfor Nordjylland	3
Samlet poengsum (av 110 mulige)		76

Det første kriterium, at man kan påvise en statistisk sammenheng mellom innføring av et tiltak og endringer i ulykkestall, er klart oppfylt. Det samme gjelder kriteriet om at virkningene av tiltaket bør være sterkere enn virkninger av andre forhold som påvirker ulykkestallene. Dette kriteriet er oppfylt fordi signifikanssannsynligheten til koeffisientene som representerer kampanjene er lavere enn de øvrige koeffisienter i modellen for Nordjylland (det vil si at koeffisientene er mer signifikante enn andre koeffisienter). Kravet om en konsistent effekt (kriterium 3) må også sies å være oppfylt, siden det i alle år det har pågått en kampanje, unntatt ett, er registrert lavere ulykkestall enn i de år det ikke har pågått noen

kampanje. Det ene unntaket fra denne regelen kan skyldes tilfeldig variasjon. Kravet om en entydig årsaksretning (kriterium 4) kan bedømmes med utgangspunkt i fortegnet på endringer i ulykkestall når et tiltak innføres og oppheves. Ulykkestallet sank fra 1987 til 1988, men økte fra 1993 til 1994, da det ikke var noen kampanje. Fra 1994 til 1995 gikk derimot ikke antall ulykker ned, etter at kampanjevirkningen ble gjenopptatt. Modellberegningene viser konsekvent nedgang i ulykkestall når en kampanje begynner og økning når den slutter. Undersøkelsen er tildelt 8 av 10 mulige poeng med hensyn til dette kriteriet.

Med hensyn til kriterium 5, kontroll for andre forklaringer, er seks mulige feilkilder i før-og-etterundersøkelser tidligere nevnt. Gjennom den multivariate analysen kan undersøkelsen sies å ha kontrollert for tre av disse: langsiktige trender i ulykkestall, endringer i trafikkmengde og andre hendelser. Trender er representert ved koeffisienten for årstall. Folketall er brukt som indikator på trafikkmengde, siden det ikke foreligger reisevanedata som viser det faktiske omfang av ferdsel i Nordjylland i den aldersgruppen kampanjene var rettet mot. Endringer i realinntekt er ment å indikere andre hendelser som kan påvirke antall ulykker i tillegg til de øvrige variabler som inngår i modellen. Videre er det dokumentert at regresjonseffekter som følge av unormalt høye ulykkestall i før-perioden sannsynligvis ikke er en feilkilde i undersøkelsen.

Det foreligger ikke opplysninger om eventuelle andre tiltak som er innført samtidig med kampanjene. Det er derfor umulig å vite om dette er en feilkilde i undersøkelsen.

Det må anses som lite sannsynlig at kampanjene mot promillekjøring i Nordjylland har medført ulykkesmigrasjon, det vil si ført at promilleulykkene har økt andre steder i Danmark. Tendensen i tallene tyder snarere på en motsatt virkning, det vil si at kampanjene i Nordjylland kan ha stimulert til lignende tiltak andre steder.

Alt i alt tildeles studien 45 av 60 oppnåelige poeng når det gjelder kontroll for andre forklaringer på resultatene.

Kriteriene 6, 7 og 8, som gjelder avdekking av en årsaksmekanisme, samsvar med veletablert teori og forekomst av en dose-responsammenheng er enten ikke anvendelige, eller kan ikke sies å være oppfylt for denne undersøkelsen. Derimot kan man bedømme undersøkelsen ut fra kriterium 9, at en virkning bare skal påvises i målgruppen for tiltaket, ikke utenfor denne. Dette kriteriet er i liten grad oppfylt og undersøkelsen tildeles 3 av 10 oppnåelige poeng for dette kriteriet.

Undersøkelsen kan dermed vurderes på grunnlag av 8 av de 9 kriteriene (alle unntatt dose-responsammenheng) og kan maksimalt oppnå 110 poeng. Undersøkelsen er vurdert til å oppnå 76 poeng. Dersom skalaen konverteres til verdier mellom 0 og 1 tilsvarer dette en score på 0,691. Dette er over skalaens medianverdi (0,500) og kan tolkes dit hen at det kan gis sterkere grunner for å tro at de påviste sammenhenger er uttrykk for årsakssammenhenger enn de grunner som kan gis for ikke å tro på dette.

6 Drøfting og konklusjoner

Det byr alltid på metodiske utfordringer å gjøre gode før-og-etterundersøkelser av trafiksikkerhetstiltak. Empirisk Bayes metode er av mange betraktet som den beste metoden for å gjøre slike undersøkelser. Denne metoden kan imidlertid ikke alltid anvendes. Hva bør da gjøres?

I dette paperet er det gitt et eksempel på hvordan man, ved å innhente lett tilgjengelige supplerende data, kan gjøre en evaluering av brukbar, om ikke fullgod, kvalitet også når empirisk Bayes metode ikke kan benyttes. Eksemplet omhandler kampanjene mot promillekjøring blant ungdom i Nordjylland. Effektene av disse kampanjene ble evaluert ved å føye en negativ binomial regresjonsmodell til dataene. Mange vil hevde at dette er fullstendig meningsløst og at man trenger et langt større utvalg for å kunne benytte en slik analyseteknikk (Lord 2006). I dette tilfellet ser imidlertid den negative binomiale regresjonsmodellen ut til å ha fungert tilfredsstillende. Alle koeffisientene har de ventede fortegn og modellen forklarer praktisk talt all systematisk variasjon i dataene. Fordelen med å benytte en slik modell, er at man får kontrollert statistisk for alle variabler som inngår i modellen. Modellen gir likevel bare forventningsrette resultater

dersom man kan gå ut fra at de registrerte ulykkestallene ikke var unormalt høye eller lave. I dette tilfellet ble dette vurdert ved å sammenligne antall promilleulykker per 1.000 innbyggere i Nordjylland amt med tilsvarende tall i andre danske amt. Nordjylland skilte seg da ikke ut med et spesielt høyt antall promilleulykker sammenlignet med andre amt.

De viktigste konklusjoner som kan trekkes av denne undersøkelsen kan oppsummeres i følgende punkter:

1. Når man skal evaluere virkninger av trafikksikkerhetstiltak, er det viktig å oppnå så god kontroll over mulige feilkilder i undersøkelsen som tilgjengelige data gjør mulig.
2. Hvis foreliggende data for eksempel ikke gjør det mulig å avgjøre om ulykkestallene var unormalt høye før tiltaket ble gjennomført, bør man innhente lett tilgjengelige supplerende data fra andre kilder for å gjøre det mulig å sammenligne ulykkestallene der hvor tiltaket er innført med ulykkestall fra andre tilsvarende steder.
3. Når det ikke foreligger eksponeringsdata i form av antall kjørte kilometer, bør andre lett tilgjengelige indikatorer benyttes, eksempelvis folketall eller antall førerkortinnehavere.
4. Som indikator på andre hendelser som kan forekomme samtidig med innføring av et tiltak, og som kan påvirke ulykkestallene, er et mål på økonomisk utvikling godt egnet, da erfaring viser at antall ulykker samvarierer med økonomiske konjunkturer. Realinntektsutvikling per innbygger er et eksempel på en lett tilgjengelig og brukbar indikator.
5. For samtidig å kunne beregne virkningene på antall ulykker av alle de variabler man har opplysninger om, må det benyttes en multivariat analyseteknikk, for eksempel negativ binomial regresjon. En negativ binomial regresjonsanalyse av data for Nordjylland for perioden 1983-2005 ga meningsfulle resultater.

Referanser

Bech, V. Evaluering af Skytsengelkampagnen i Nordjyllands Amt. Upublisert manuskript. Aalborg, Nordjyllands Amt, 2006.

Elvik, R. Assessing the validity of road safety evaluation studies by analysing causal chains. *Accident Analysis and Prevention*, 35, 741-748, 2003.

Elvik, R. Operational criteria of causality for observational road safety evaluation studies. *Transportation Research Record*, 2019, 74-81, 2007.

Elvik, R. Making sense of road safety evaluation studies. Developing a quality scoring system. Report 984. Oslo, Institute of Transport Economics, 2008.

Elvik, R. Assessing causality in multivariate accident models. *Accident Analysis and Prevention*, 43, 253-264, 2011.

Hauer, E. *Observational before-after studies in road safety*. Oxford, Pergamon Press (Elsevier Science), 1997.

Hauer, E. The harm done by tests of significance. *Accident Analysis and Prevention*, 36, 495-500, 2004.

Hauer, E. Cause and effect in observational cross-section studies on road safety. Report. Highway Safety Information System. US Department of Transportation, Federal Highway Administration, Turner-Fairbank Highway Research Center, McLean, Virginia, 2005.

Hauer, E., Harwood, D. W., Council, F. M., Griffith, M. S. Estimating safety by the Empirical Bayes method. A tutorial. *Transportation Research Record*, 1784, 126-131, 2002.

Highway Safety Manual. First edition. American Association of State Highway and Transportation Officials (AASHTO), Washington DC, 2010.

Lord, D. Modeling motor vehicle crashes using Poisson-gamma models: examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. *Accident Analysis and Prevention*, 38, 751-766, 2006.

Persaud, B., Lyon, C. Empirical Bayes before-after safety studies: Lessons learned from two decades of experience and future directions. *Accident Analysis and Prevention*, 39, 546-555, 2007.

Studsholt, P. Campaigns against drunken driving by young drivers. Proceedings of road safety and traffic in Europe (Conference Gothenburg, Sweden, September 26-28). VTI-report 365A, 26-33. Linköping, Swedish Road and Traffic Research Institute, 1990.