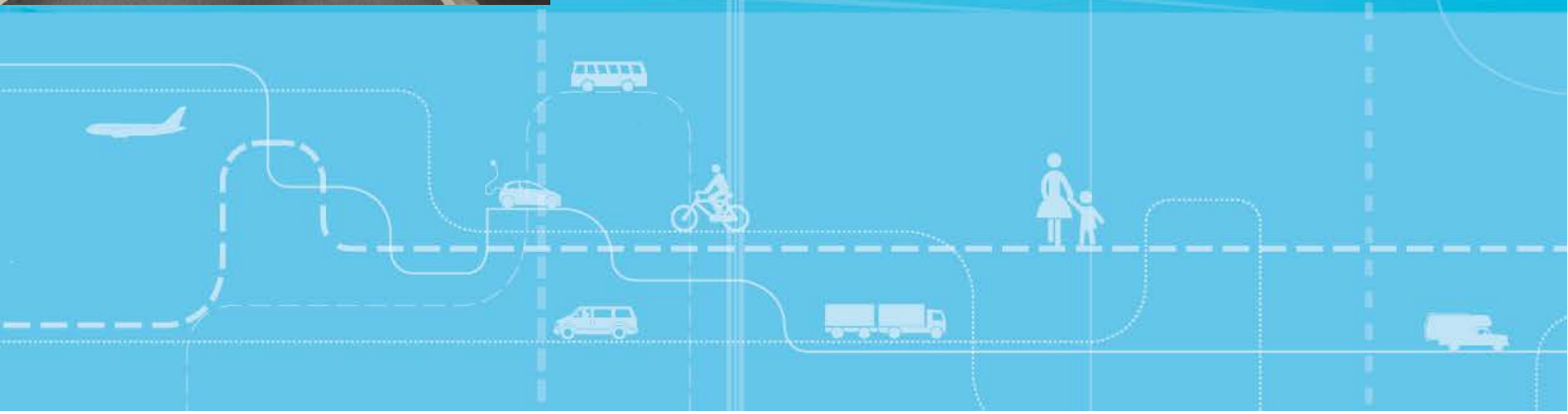


Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge



Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge

Alena Høye

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.

ISSN 0808-1190

ISBN 978-82-480-1528-4 Papirversjon

ISBN 978-82-480-1527-7 Elektronisk versjon

Oslo, mai 2013

Tittel: Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge

Forfattere: Alena Høye

Dato: 05.2014

TØI rapport: 1323/2014

Sider 45

ISBN Papir: 978-82-480-1528-4

ISBN Elektronisk: 978-82-480-1527-7

ISSN 0808-1190

Finansieringskilde: Statens vegvesen Vegdirektoratet

Prosjekt: 3751 - Nye modeller for beregning av normale skadegrader og ulykkesrisiko

Kvalitetsansvarlig: Rune Elvik

Emneord: Modell

Trafikksikkerhet

Ulykker

Sammendrag:

Ulykkesmodeller er utviklet for riks- og fylkesvegnettet i Norge for å beregne normale antall personskadeulykker, lett skadde, hardt skadde, drepte og samlet antall drepte eller hardt skadde som en funksjon av trafikkmengde, strekningslengde, vegtype, fartsgrense, antall kjørefelt, antall kryss og andre vegegenskaper. Den valgte modellformen (negative binomialmodeller med variabel overspredningsparameter) gjør det mulig å benytte resultatene bl.a. i evalueringer av trafikksikkerhetstiltak med den empiriske Bayes metoden som i dag anses som den beste for å redusere risikoen for å systematisk overestimere virkninger av trafikksikkerhetstiltak i før-etter studier. Resultatene kan også brukes i ulike verktøy som brukes av Statens vegvesen for å gjøre analyser av ulykkessituasjonen i vegnettet og for å beregne forventede virkninger av trafikksikkerhetstiltak.

Title: Development of crash prediction models for national and county roads in Norway

Author(s): Alena Høye

Date: 05.2014

TØI report: 1323/2014

Pages 45

ISBN Paper: 978-82-480-1528-4

ISBN Electronic: 978-82-480-1527-7

ISSN 0808-1190

Financed by: The Norwegian Public Roads Administration

Project: 3751 - Nye modeller for beregning av normale skadegrader og ulykkesrisiko

Quality manager: Rune Elvik

Key words: Accident

Model

Road safety

Summary:

Crash prediction models for national and county roads in Norway were developed in order to calculate predicted numbers of injury crashes, slightly injured, seriously injured, fatalities, and the total number of killed or seriously injured, as a function of traffic volume, segment length, road category, speed limit, number of lanes, number of intersections and other road characteristics. Models were calculated as generalized negative binomial models (negative binomial models with a variable overdispersion parameter). Results can be used in evaluations of road safety measures with the empirical Bayes method which is regarded as state of the art with respect to control for regression to the mean. The models can also be applied in conjunction with several tools of the Norwegian Public Roads Administration for road safety analyses and estimation of expected effects of road safety measures.

Language of report: Norwegian

Forord

På oppdrag av Vegdirektoratet er det utviklet ulykkesmodeller for å beregne normale antall ulykker, skadde og drepte på riks- og fylkesveger i Norge. Normale ulykkes- og skadetall som beregnes med modellene kan benyttes eksempelvis i programmer som TS-Effekt og S-kost. Den valgte modellformen gjør det også mulig å benytte resultater fra modellberegninger i evalueringer av trafikksikkerhetstiltak med den empiriske Bayes metoden.

TØIs prosjektleder har vært Alena Høye som også har skrevet rapporten. Oppdragsgivers kontaktperson har vært Arild Ragnøy. Arild Ragnøy, Svenn Fjeld Olsen og Finn Harald Amundsen, alle fra Vegdirektoratet, har bidratt med nyttige kommentarer til tidligere versjoner av rapporten. Datagrunnlaget for modellberegningene er i hovedsak skaffet til veie av Håvard Lerstad fra Triona, basert på data i den nasjonale vegdatabanken (NVDB). Statens vegvesen har i tillegg skaffet informasjon om forsterket midtoppmerking.

Rune Elvik har stått for kvalitets sikring av rapporten. Trude Rømming har tilrettelagt rapporten for utgivelse i papirform og elektronisk.

Oslo, mai 2014
Transportøkonomisk institutt

Gunnar Lindberg
direktør

Michael W. J. Sørensen
avdelingsleder

Innhold

Sammendrag

Summary

1	Bakgrunn	1
2	Modellform og prediktorvariabler	3
	2.1 Modellform.....	3
	2.2 Segmentlengde og antall år.....	4
	2.3 Trafikkmengde.....	6
	2.4 Fartsgrense.....	8
	2.5 Antall kjørefelt.....	10
	2.6 Kryss, rundkjøringer og ramper.....	11
	2.7 Kurver.....	13
	2.8 Stigninger.....	15
	2.9 Vegklasse.....	17
	2.10 Midtdeler og midtrekkverk.....	18
	2.11 Forsterket midtoppmerking.....	19
	2.12 Fylke.....	20
	2.13 Konstantterm.....	21
	2.14 Sammenheng mellom ulike vegegenskaper.....	21
	2.15 Overspredningsparameter.....	24
3	Modellene	25
	3.1 Koeffisientene.....	25
	3.2 Hvor gode er modellene?.....	28
4	Trendfaktor	32
	4.1 Trendjustering av predikerte (normale) ulykkestall.....	32
	4.2 Trendjustering av overspredningsparameteren.....	34
	4.3 Trendjustering av trafikkmengden.....	37
5	Bruerveiledning for Ulykkesmodeller.xlsx	39
6	Referanser	42

Sammendrag:

Utvikling av ulykkesmodeller for ulykker på riks- og fylkesvegnettet i Norge

TØI rapport 1323/2014
Forfatter: Alena Høye
Oslo 2014 45 sider

Ulykkesmodeller er utviklet for riks- og fylkesvegnettet i Norge for å beregne normale antall personskadeulykker, lett skadde, hardt skadde, drepte og samlet antall drepte eller hardt skadde som en funksjon av trafikkmengde, strekningslengde, vegtype, fartsgrense, antall kjørefelt, antall kryss og andre vegegenskaper. Den valgte modellformen (negative binomialmodeller med variabel overspredningsparameter) gjør det mulig å benytte resultatene bl.a. i evalueringer av trafikksikkerhetstiltak med den empiriske Bayes metoden som i dag anses som den beste for å redusere risikoen for å systematisk overestimere virkninger av trafikksikkerhetstiltak i før-etter studier. Resultatene kan også brukes i ulike verktøy som brukes av Statens vegvesen for å gjøre analyser av ulykkesituasjonen i vegnettet og for å beregne forventede virkninger av trafikksikkerhetstiltak.

Ulykkesmodellene som er utviklet på oppdrag av Vegdirektoratet er beregnet med ulykkesdata fra den nasjonale vegdatabanken (NVDB) fra 2006-2011. Rapporten bygger på tidligere rapporter om utvikling av ulykkesmodeller for det norske vegnettet (Ragnøy, Christensen & Elvik, 2002; Ragnøy & Elvik, 2003). Modellene som er utviklet i den aktuelle rapporten bygger på mer aktuelle ulykkesdata og en videreutvikling av modellene og metodene som er brukt av Ragnøy et al. (2002). Modellutviklingen er også beskrevet i et arbeidsdokument (Høye, 2013).

Modellene kan brukes for å beregne normale antall personskadeulykker (PSU), lett skadde (LS), hardt skadde (HS), drepte (D) og det samlede antall drepte eller hardt skadde (D/HS) på riks- og fylkesvegnettet. Normale ulykkestall beregnes som en funksjon av bl.a. trafikkmengde og ulike vegegenskaper. Et normalt ulykkestall er antall ulykker som skjer i gjennomsnitt på en veg av en gitt lengde med en gitt trafikkmengde og gitte vegegenskaper. Modellformen som er benyttet (negative binomialmodeller med variabel overspredningsparameter) gjør det også mulig å beregne forventede ulykkestall som en funksjon av normale ulykkestall, registrerte ulykkestall og overspredningsparameteren. Forventede ulykkestall er antall ulykker som man på lang sikt forventer at vil skje på en vegstrekning, ut fra generelle vegegenskaper og ulykkene som har skjedd på strekningen.

Overspredningsparameteren brukes for å vekte sammen normale og registrerte ulykkestall. Overspredningsparameteren varierer som en funksjon av trafikkmengde og strekningslengde, slik at registrerte ulykkestall får en større vekt på vegstrekninger med store normale ulykkestall.

Når man beregner det forventede antall ulykker på en strekning med unormalt mange ulykker antar man at resultatet viser hvor mange ulykker man kan forvente på strekningen på lang sikt. Man tar dermed hensyn til at antall ulykker kan ha vært tilfeldig høyt og at antall ulykker på langt sikt mest sannsynlig vil ligge nærmere det normale ulykkestallet. Ved å benytte forventede ulykkestall, istedenfor registrerte ulykkestall, kan man i en før-etter studie unngå å overestimere virkningen av trafikksikkerhetstiltak. Tiltak settes ofte inn på vegstrekninger hvor det har vært spesielt mange ulykker og derfor vil antall ulykker i mange tilfeller gå ned (komme nærmere gjennomsnittet) uansett om man setter inn et tiltak eller ikke.

Modell og prediktorvariabler

Modellformen som er brukt for alle skadegrader er en negativ binomialmodell (NB modeller) med variabel overspredningsparameter. Modellformen er følgende:

$$E(n) = e^{\sum_i \text{Prediktor}_i * \text{Koeff}_i}$$

$E(n)$ er det predikerte (normale) ulykkestall (dvs. antall PSU, LS, HS, D eller D/HS), prediktorene er trafikkmengden og ulike vegegenskaper, i er subskript for prediktorene. I tillegg inneholder modellen prediktorer og koeffisienter for overspredningsparameteren.

Fordelen med denne modellformen er at den tar hensyn til overspredningen i ulykkesdata og at overspredningen varierer som en funksjon av trafikkmengde, segmentlengde og antall år. Dermed er det mulig å benytte resultatene bl.a. i evalueringer av trafikksikkerhetstiltak med den empiriske Bayes (EB) metoden. Denne anses i dag som den beste for å redusere risikoen for å systematisk overestimere virkninger av trafikksikkerhetstiltak i før-etter studier. I en EB studie sammenligner man det registrerte ulykkestallet etter at et tiltak er satt inn på en veg med det forventede ulykkestallet. Det forventede ulykkestallet beregnes ved å vekte sammen det registrerte ulykkestall og det normale ulykkestall. Det normale ulykkestall er antall ulykker som er predikert av modellen, dvs. det gjennomsnittlige antall ulykker på en vegstrekning med de gitte vegegenskapene. Vekten er en funksjon av overspredningsparameteren og det normale antall ulykker. Det forventede antall ulykker ligger alltid mellom det normale og det registrerte antall.

Prediktorvariablene i modellene for alle skadegrader er følgende:

- **Segmentlengde og antall år:** Segmentlengde og antall år inngår i alle modellene som eksponeringsvariabler, slik at det normale antall ulykker alltid øker proporsjonalt med segmentlengde og antall år (koeffisientene for $\text{Ln}(\text{segmentlengde} * \text{antall år})$ er satt lik 1). Antall år er prediktor fordi det ikke foreligger data fra alle år for alle segmentene; segmenter med vesentlige endringer (som for eksempel endringer av fartsgrensen) inngår kun med ulykkesdata fra etter endringen i datafilen som ligger til grunn for modellberegningene.
- **Trafikkmengde: $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$.** Trafikkmengden inngår i modellene som $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$. Dermed øker normale ulykkestall med økende trafikkmengde, men økningen blir stadig svakere ved høyere trafikkmengde.

- **Fartsgrense: Dummyvariabler.** For hver fartsgrense er det definert en dummyvariabel som er én ved den respektive fartsgrensen og null ellers. Dummyvariabler er valgt framfor en tallvariabel for å ta hensyn til generelle forskjeller mellom veger med ulike fartsgrenser som gjør at antall ulykker kan gå opp og ned mellom ulike fartsgrenser.
- **Antall kjørefelt: Dummyvariabler.** For antall kjørefelt er det definert dummyvariabler på samme måte og av de samme grunnene som beskrevet for fartsgrense. Segmenter med kun ett kjørefelt er utelatt fra datasettet fordi dette er en liten og svært heterogen gruppe av helt ulike typer veger og dermed ikke noen meningsfylt kategori.
- **Plankryss, rundkjøringer og ramper (planskilte kryss): $\text{Ln}(\text{antall} + 1)$.** For X-kryss, T-kryss, rundkjøringer, av-ramper, på-ramper og uspesifiserte ramper er det beregnet variabler som er den naturlige logaritmen av antallet kryss/rundkjøringer/ramper pluss én (pluss én for å unngå å ta logaritmen av null). Dette har vist seg å gi mer logiske resultater enn andre typer variabler (f.eks. uttransformert eller kvadrert). X-kryss er firearmede plankryss, T-kryss er trearmede plankryss (det skilles ikke mellom ulike typer kryssregulering), rundkjøringer omfatter alle typer rundkjøringer, av- og påramper er ramper til og fra hovedveger med planskilte kryss, og uspesifiserte ramper er ramper hvor det i datafilen ikke er spesifisert om det handler om av- eller på-ramper (uspesifiserte ramper vil ikke være relevante for brukere av modellen, men måtte likevel være med i modellene).
- **Kurver: $\text{Ln}(\text{antall kurver} + 1)$, delt opp etter fartsgrense:** Hver del av vegstrekningen på 50 m lengde med kurveradius under 300 m regnes som én kurve. Kurvevariabelen i modellene er den naturlige logaritmen av antallet slike kurver pluss 1. Det er også gjort forsøk med en lignende kurvevariabel (deler av vegstrekningen på 30 m lengde med en kurveradius på under 200 m), men resultatene virket litt mindre logiske. Kurvevariablene sier lite om veggeometrien fordi det ikke er tatt hensyn til retningsendringer; en strekning med syv "kurver" etter denne definisjonen kan ha alt fra én langstrakt kurve til syv korte krappe kurver med til sammen åtte retningsendringer. For å ta hensyn til at kurver kan ha ulik sammenheng med antall ulykker / skader er kurvevariabelen delt opp etter fartsgrense. Forsøk med å skille mellom enslige kurver og kurver på kurverike strekninger (enslige kurver har som regel høyere risiko enn kurver som har flere andre kurver i nærheten) ga ikke meningsfulle resultater, trolig fordi kurvevariablene ikke tar hensyn til retningsendringer.

- **Stigninger: $\ln(\text{antall stigninger} + 1)$, delvis delt opp etter fartsgrense:** Hver del av vegstrekningen på 200 m lengde med en stigning på minst 4% regnes som 1 stigning. Stigningsvariabelen i modellene er den naturlige logaritmen av antallet slike stigninger pluss 1. Det er også gjort forsøk med en lignende stigningsvariabel (deler av vegstrekningen på 200 m lengde med en stigning på minst 2%), men resultatene virket litt mindre logiske. Problemene med stigningsvariabelene er de samme som med kurvevariabelen, det er ikke tatt hensyn til retningsendringer (høy-/lavbrekk). For å ta hensyn til at stigninger kan ha ulik sammenheng med antall ulykker / skader er stigningsvariabelen delt opp etter fartsgrense i modellene for PSU og LS. I modellene for HS og D er en slik inndeling ikke gjort fordi det er for få HS / D på segmenter med stigninger og spesifikke fartsgrenser og det er derfor kun brukt én stigningsvariabel i disse modellene.
- **Vegklasser: Fem dummyvariabler.** Det er definert fem dummyvariabler for ulike vegklasser: 1 = motorveg; 2 = tofeltsveg med planskilte kryss (tidligere betegnet som motortrafikkveg); 3 = TEN-T-veg (ikke motorveg / tofeltsveg med planskilte kryss); 4 = øvrig europa-/riksveg (ikke motorveg / tofeltsveg med planskilte kryss / TEN-T-veg); 5 = fylkesveg. Disse vegklassene er valgt fordi det finnes generelle forskjeller i vegstandarden (bl.a. kjørefelt- og skulderbredde) mellom klassene; f.eks. har alle motorveger midtrekkverk og TEN-T-veger har generelt en noe høyere standard enn øvrige europa- og riksveger. Det er også gjort forsøk med en inndeling av de enkelte vegklassene i veger med og uten midtdeler / midtrekkverk. Dette ga til sammen 11 vegtyper. Modeller med de 11 vegtypene har delvis vist seg å være noe bedre enn modellene med de 5 vegklassene, forskjellene er imidlertid små og ikke helt konsistente for ulike skadegrader og indikatorer for goodness of fit. De 5 vegklassene er valgt som prediktorer (istedenfor de 11 vegtypene) fordi resultatene er mer oversiktlige og lettere å tolke.
- **Midtdeler / -rekkverk: Fire dummyvariabler for veger med / uten midtdeler / midtrekkverk.** Alle segmentene er delt inn i segmenter med både midtdeler og midtrekkverk, segmenter med midtdeler og uten midtrekkverk, segmenter med midtrekkverk og uten midtdeler, og segmenter uten verken midtdeler eller midtrekkverk. For hver av de fire gruppene er det definert én dummyvariabel. Det er ikke definert to separate dummyvariabler (én for med/uten midtdeler og én for med/uten midtrekkverk) fordi det kan være interaksjonseffekter mellom midtdeler og midtrekkverk, dvs. at virkningen av midtrekkverk kan være forskjellig på veger med og uten midtdeler og at virkningen av midtdeler kan være forskjellig på veger med og uten midtrekkverk.
- **Forsterket midtoppmerking: To dummyvariabler for bred og smal forsterket midtoppmerking.** Forsterket midtoppmerking er kombinasjonen av vanlig oppmerking med rumleriller. Hvis hele midtoppmerkingen (inkludert rumleriller) er 1 m bred eller bredere, kalles det i NVDB-datafilen "midtfelt", mens smalere forsterket midtoppmerking går under betegnelsen "midtrill".
- **Fylke: Dummyvariabler.** For hvert fylke er det definert én dummyvariabel. Disse skal fange opp generelle forskjeller mellom fylkene (f.eks. forskjeller i topografi, vær og befolkningstetthet).
- **Konstantterm:** Alle modellene inneholder en konstantterm.

For hver av modellene er det i tillegg estimert koeffisienter for å beregne **overspredningsparameteren** som en funksjon av segmentlengde, antall år og trafikkmengde. Overspredningen synker med økende segmentlengde, antall år og trafikkmengde.

Hvor gode er modellene?

Det er beregnet flere ulike goodness-of-fit (GOF) indikatorer. Når man ser på andelen av variansen i de registrerte antall som er forklart av modellene, er den endelige modellen for PSU den beste, mens modellene for HS og D er forholdsvis svake. Når man sammenligner ulike modeller (med ulike prediktorer) viser GOF indikatorene kun svært små forskjeller mellom ulike modellvarianter for hver skadegrad. De eneste modellene som skiller seg ut er modellene med kun $\ln(\text{ÅDT})$ som prediktor som er klart dårligere enn modellene med øvrige prediktorer. Blant disse er modellen med $\ln(\text{ÅDT})$ som eneste trafikkmengde-prediktor noe dårlige enn modellene med flere trafikkmengde-prediktorer for alle skadegrader. Testberegninger som er gjort for å finne de beste variabeldefinisjonene for enkelte prediktorer har også vist at modellenes GOF nesten ikke påvirkes av hvordan man definerer enkelte prediktorer eller om enkelte prediktorer er med eller ikke. Dette betyr i praksis at modellene ikke vil endre seg i stor grad hvis man endrer eller utelater enkelte prediktorer (for predikerte ulykkes-/skadetall på enkelte segmenter eller vegstrekninger kan det likevel ha betydning hvilke variabler som er med i modellen).

Det er to problemer med modellresultatene som det ikke er funnet noen løsning på:

- Det første problemet er at modellen for LS predikerer alt for mange LS (2,45% flere enn det faktiske antallet i hele datafilen; avviket er størst ved ÅDT over ca. 3000). Mulige forklaringer er: A) at det finnes relevante prediktorvariabler som ikke er inkludert i modellen (for eksempel forholdet mellom kapasitet og trafikkmengde); B) at antall LS i større grad er underrapportert på veger med høy trafikkmengde enn på andre veger og C) at segmenter med høy ÅDT i gjennomsnitt er representert med data fra færre år slik at usikkerheten (og dermed avvikene) her er større enn på segmenter med lavere ÅDT .
- Det andre problemet er at variansen i de predikerte antallene HS og D er mindre enn gjennomsnittet, noe som pleier å være omvendt. Forklaringen er trolig at det finnes svært mange segmenter med null HS og D.

For øvrig ser modellene generelt logiske ut i den forstand at de aller fleste koeffisientene har omtrent forventet størrelse og fortegn. Det er kun noen enkelte koeffisienter i modellene for HS og D som er justert manuelt i regnearket fordi koeffisientene var absurd store, noe som trolig skyldes små antall HS og D i segmenter med de respektive variablene.

Avvikene mellom de faktiske og predikerte antallene er forholdsvis små og kun i liten grad systematisk forskjellig fra null når man ser på antall drepte / skadde per kilometer per år på lengre sammenhengende strekninger. Dette gjelder både LS og D/HS. Kun ved ÅDT over omtrent 10 000 ser man at de gjennomsnittlige predikerte antallene LS er noe høyere enn de faktiske, men heller ikke her er avvikene dramatisk store.

Regneark Ulykkesmodeller

I vedlagte regneark Ulykkesmodeller.xlsx kan man beregne:

- Predikerte (normale) antall PSU, LS, HS og D som en funksjon av de prediktorvariablene som er beskrevet ovenfor
- En overspredningsparameter for hver skadegrad
- En vekt for hver skadegrad som kan brukes til å beregne forventede ulykkes/-skadetall
- Hvis man i tillegg oppgir registrerte ulykkes- og skadetall, beregnes også forventede antall PSU, LS, HS og D.

Det er også mulig å omregne resultater til år mellom 1997 og 2020.

Summary:

Development of crash prediction models for national and county roads in Norway

TØI Report 1323/2014

Author: Alena Høy

Oslo 2014, 45 pages Norwegian language

Crash models for national and county roads in Norway were developed in order to calculate predicted numbers of injury crashes, slightly injured, seriously injured, fatalities and the total number of killed or seriously injured, as a function of traffic volume, segment length, road category, speed limit, number of lanes, number of intersections and other road characteristics. Models were calculated as generalized negative binomial models (negative binomial models with a variable overdispersion parameter). Results can be used in evaluations of road safety measures with the empirical Bayes method which is regarded as state of the art with respect to control for regression to the mean. The models can also be applied in conjunction with several tools of the Norwegian Public Roads Administration for road safety analyses and estimation of expected effects of road safety measures.

Crash models were developed on behalf of the Norwegian Public Roads Administration based on crash data in the national roads data base from the years 2006-2011. The models can be used to calculate predicted numbers of injury crashes, slightly injured, seriously injured, fatalities and the total number of killed or seriously injured (KSI) on national and county roads as a function of traffic volume, segment length, road category, speed limit, number of lanes, number of intersections and other road characteristics. Models were calculated as generalized negative binomial models (negative binomial models with a variable overdispersion parameter) which allows the calculation of expected crash numbers as a function of the model predictions, crash counts and an overdispersion parameter. Expected crash numbers refer to the number of crashes that can be expected on a road segment in the long run, based on general road characteristics and the specific crash history of the road segment in question. The statistical weights for model predictions and crash counts are calculated as functions of the overdispersion parameter. The overdispersion parameter varies as a function of traffic volume and segment length. The statistical weight for the crash counts increases with increasing segment length and traffic volume.

The expected numbers of crashes will always be between the model prediction and the observed numbers of crashes. In before-after studies of the safety effects of measures at high-crash locations, regression to the mean is less likely to affect the results when the analysis is based on expected crash numbers in the before period instead of observed crash counts. Regression to the mean occurs when the observed number of crashes in the before period was exceptionally high. One would then expect the number of crashes to decrease in the after period even without any (effective) safety measures.

Model and predictors

The models were developed as generalized negative binomial (NB) models in which the overdispersion parameter is estimated as a function of traffic volume, segment length and number of years. The model form is as follows:

$$E(n) = e^{\sum_i \text{Predictor}_i * \text{Coeff}_i}$$

$E(n)$ is the predicted number of crashes (i.e. number of injury crashes / injuries / fatalities), predictors are traffic volume and a number of road characteristics, and i is the subscript for the predictors.

A generalized NB model was chosen because it takes into account that crash counts usually are overdispersed, and that overdispersion is not a constant but depending on traffic volumes, segment length and the number of years. The model results can be used in before-after evaluations of road safety measures with the empirical Bayes (EB) method. The EB method controls for regression to the mean by comparing the observed number of crashes in the after period with the expected number of crashes. The expected number of crashes is a function of the actual crash count and the model prediction of the number of crashes for the same road section and period of time. Crash counts and model predictions are weighted with a function of the overdispersion parameter. The expected number of crashes is always between the crash count and the model prediction.

The crash models for all dependent variables (injury crashes and numbers of injuries / fatalities) are based on the following predictors:

- **Segment length and number of years: $\ln(\text{segment length})$ and $\ln(\text{number of years})$.** The coefficients for the natural logarithm of segment length and number of years are set equal to one, thus normal crash numbers increase proportionally with segment length and number of years. The number of years is a predictor because segments with substantial changes (e.g. speed limit reductions) are represented with crash data only from after the change was made in the data the models are based on.
- **Traffic volume: $\ln(\text{AADT})$ and $\ln(\text{AADT})^2$.** Volume predictors are the natural logarithm of AADT (annual average daily traffic) and the squared natural logarithm of AADT. Predicted crash numbers increase with increasing traffic volume, but at a decreasing pace.

- **Speed limit: Dummy variables.** For each speed limit a dummy variable is defined in order to take into account that crash numbers not necessarily are a monotonous (or other) function of speed limit because of general differences between roads with different speed limits (e.g. many roads with a 70 km/h speed limit had previously an 80 km/h speed limit that was reduced because of exceptionally high crash numbers).
- **Number of lanes: Dummy variables.** For each number of lanes a dummy variable is defined for the same reasons as for speed limit. Segments with one lane are a highly heterogeneous group of different kinds of roads and only a small proportion of all data. They are therefore omitted from the data.
- **At-grade intersections, roundabouts and ramps (grade separated intersections): $\ln(\text{Number of } \dots + 1)$.** For four-armed intersections, three-armed intersections, roundabouts, off-ramps and on-ramps the natural logarithm of the number of intersections / roundabouts / ramps plus one is calculated (plus one in order to avoid taking the logarithm of zero on segments without any of these).
- **Curves: $\ln(\text{Number of curves} + 1)$ according to speed limit.** The curve variable that was available was the number of parts of each road segment that is 50 m long and has a curve radius below 300 m. The natural logarithm of the number of curves according to this definition plus one is calculated and one such variable is defined for each speed limit ($\ln(\text{number of curves} + 1)$ at the respective speed limit and zero at all other speed limits) in order to take into account different effects of curves on crash numbers at different speed limits. The available curve variable has only a weak relationship with the actual number of curves or other definitions of road curvature as it does not take into account turning points in road curvature.
- **Grades: $\ln(\text{Number of grades} + 1)$ according to speed limit.** The grade variable that was available was the number of parts of each road segment that are 200 m long and that have a vertical grade of at least 4%. The natural logarithm of the number of grades according to this definition plus one is calculated and one such variable is defined for each speed limit ($\ln(\text{number of grades} + 1)$ at the respective speed limit and zero at all other speed limits) in order to take into account different effects of grades on crash numbers at different speed limits. The available grade variable has only a weak relationship with the actual number of grades and contains no information about crest or sag curves.
- **Road category: Five dummy variables.** A dummy variable is defined for each of the following road categories: Motorway; two-lane road with grade separated intersections; TEN-T road (other than motorway or two-lane road with grade separated intersections); European or state highway (other than those previously mentioned); county road.
- **Median and median guardrail: Four dummy variables.** A dummy variable is defined for each of the following: Median with guardrail; median without guardrail; guardrail that separates opposing directions of traffic without median; neither guardrail nor median. Dummy variables for the presence vs. absence of either median or guardrail would not allow to detect interaction effects between guardrail and median presence, which is why four dummy variables for each possible combination were defined.

- **Center line rumble strips: Two dummy variables.** Narrow (below one meter) and wide (one meter or wider) center line rumble strips are represented in the model by dummy variables.
- **County: Dummy variables.** For each county a dummy variable is defined. These variables are meant to represent general differences between counties such as topography, weather and population density.
- **Constant:** A constant term is included in all models.

Additionally, all models contain coefficients for calculating the **overdispersion parameter** as a function of the natural logarithms of segment length, number of years and traffic volume. The overdispersion parameter decreases with increasing segment length, number of years and traffic volume.

How good are the models?

Several goodness-of-fit (GOF) indicators were calculated for each model: R^2 , mean square prediction error, and Elvik-index. Additionally Cure plots were plotted. The number of injury crashes is the one with the smallest mean square prediction error.

Comparing models with different sets of predictor variables or with different functions of road characteristics (e.g. speed limit dummies vs. speed limit as a numerical variable) there are only small differences between the GOF indicators, although estimated normal crash numbers on individual road segments may differ considerably between different models. Only models with $\ln(\text{AADT})$ as the only predictor (in addition to segment length and number of years) are considerably weaker than models with most or all road characteristics as additional predictors.

Differences between model predictions and crash counts are for the most part only small and unsystematic. Only at high AADT (above ca. 10,000) the model predictions for numbers of injury crashes and slightly injured are systematically higher than the observed numbers. The total numbers of model predictions of crashes and injuries are at a maximum 2.5% higher (slightly injured) than the observed numbers.

Crash model spreadsheet

The attached spreadsheet Ulykkesmodeller.xlsx can be used to calculate:

- Model predictions for number of injury crashes, slightly injured, seriously injured, fatalities and KSI
- An overdispersion parameter for each dependent variable
- A statistical weight for each dependent variable that can be used to calculate the expected numbers of injury crashes, slightly injured, seriously injured, fatalities and KSI
- If observed numbers of crashes / injuries are entered, expected numbers of crashes / injuries are calculated as well.

It is also possible to convert the results to years between 1997 and 2020.

1 Bakgrunn

Rapporten beskriver ulykkesmodeller som er utviklet på oppdrag av Vegdirektoratet med ulykkesdata fra NVDB fra årene 2006-2011. Modellene kan brukes for å beregne det normale antall personskadeulykker (PSU), lett skadde (LS), hardt skadde (HS), drepte (D), samt hardt skadde eller drepte (D/HS) på riks- og fylkesveger i Norge. I denne rapporten brukes «ulykkestall» for lesbarhetens skyld synonymt med antall personskadeulykker, skadde eller drepte.

Rapporten bygger på tidligere rapporter om utvikling av ulykkesmodeller for det norske vegnettet (Ragnøy, Christensen & Elvik, 2002; Ragnøy & Elvik, 2003). Ragnøy et al. (2002) har utviklet ulykkesmodeller og en metode for å vekte sammen registrerte og normale ulykkestall til forventede ulykkestall. Basert på denne metoden har Ragnøy & Elvik (2003) gjort en trafikkikkerhetsanalyse av stamvegnettet. Modellene som er utviklet i den aktuelle rapporten bygger på mer aktuelle ulykkesdata og en videreutvikling av modellene og metodene som er brukt av Ragnøy et al. (2002). Modellutviklingen er også beskrevet i et arbeidsdokument som inneholder mer detaljert informasjon om de enkelte prediktorvariablene, samt en rekke testberegninger som er gjort for å velge prediktorvariabler og modellform (Høye, 2013).

Modellene er beregnet basert på data for riks- og fylkesvegnettet (unntatt etfeltveger, rundkjøringer og ramper). De fullstendige endelige modellene er vist i kapittel 3, samt i vedlagte regneark *Ulykkesmodeller.xlsx*. I det vedlagte regnearket kan man beregne normale og forventede antall PSU, LS, HS, D og D/HS, samt overspredningsparametere og vektorer for empiriske Bayes (EB)-beregninger. Det er også mulig å gjøre trendjusteringer for å omregne normale ulykkes- og skadetall, samt overspredningsparametere, til år mellom 1997 og 2020.

Normale ulykkestall: Ulykkesmodellene er statistiske modeller som kan brukes til å beregne normale ulykkestall som en funksjon av en rekke ulike vegegenskaper, bl.a. strekningslengde, trafikkmengde, fartsgrense og antall kjørefelt. For hver vegegenskap er det estimert en *koefisient* for hver skadegrad. Koeffisienten er et tall som ganges med vegegenskapen (eller en funksjon av vegegenskapen) før denne inngår i formelen for modellberegningen.

Et normalt ulykkestall er antall ulykker som skjer i gjennomsnitt på en veg av en gitt lengde og med de vegegenskapene som inngår i modellberegningen. Normale ulykkestall er som regel ikke nøyaktig identiske med registrerte ulykkestall på hver eneste vegstrekning, men i et større datamateriale kan man forvente at det kun er et lite avvik mellom det totale antall registrerte og normale ulykkestall. Ulykkesmodellene er, forenklet sagt, beregnet slik at avvikene mellom registrerte og normale ulykkestall er så små som mulig.

Forventede ulykkestall: Forventede ulykkestall er antall ulykker som man på lang sikt forventer at vil skje på en konkret vegstrekning. Antakelsen bak beregningen av forventede ulykkestall er at registrerte antall ulykker er påvirket av:

- Vegegenskapene som inngår i modellberegningen og dermed i beregningen av normale ulykkestall.

- Variasjon rundt det normale ulykkestall som skyldes delvis tilfeldigheter og delvis systematiske faktorer som ikke er fanget opp av ulykkesmodellen (Hauer et al., 2002).

Forventede ulykkestall beregnes som en funksjon av:

- Normale ulykkestall: Dermed er det tatt hensyn til den generelle sammenhengen mellom en rekke vegegenskaper og antall ulykker.
- Registrerte ulykkestall: Dermed er det tatt hensyn til at normale ulykkestall ikke har en perfekt sammenheng med faktiske ulykkestall, og at det faktiske ulykkestall kan være påvirket av faktorer som ikke er fanget opp av ulykkesmodellen som brukes for å beregne det normale ulykkestall.
- Overspredningsparameter: Denne brukes for å vekte sammen normale og registrerte ulykkestall:

$$\text{Forv. ul.} = a * \text{Reg. ul.} + (1 - a) * \text{Norm. ul.}$$

Vekten a i denne formelen beregnes som en funksjon av det predikerte ulykkestall og overspredningsparameteren:

$$a = \frac{1}{1 + \frac{\text{Predikert ulykkestall}}{\text{Overspredningsparameter}}}$$

Overspredningsparameteren er jo mindre desto større det normale ulykkestall er. Dermed får registrerte ulykkestall en større vekt på vegstrekninger med store normale ulykkestall.

Anvendelsesmulighet for forventede ulykkestall, empiriske Bayes (EB) studier:

Forventede ulykkestall kan bl.a. brukes i EB analyser for å kontrollere for regresjonseffekter. I en EB studie sammenligner man det registrerte ulykkestallet etter at et tiltak er satt inn på en veg med det forventede ulykkestallet. Det forventede ulykkestallet etter at tiltaket er satt inn beregnes som en funksjon av det forventede ulykkestallet før tiltaket ble satt inn (som beskrevet ovenfor) og den gjennomsnittlige endringen av antall ulykker på lignende veger. Ved å beregne det forventede ulykkestallet før tiltaket er satt inn tar man hensyn til at det registrerte antall ulykker før tiltaket ble satt inn kan ha vært tilfeldig høyt. Dermed kontrollerer man for regresjonseffekten, dvs. at antall ulykker mest sannsynlig går ned (er nærmere gjennomsnittet) etter å ha vært unormalt høyt i en periode.

2 Modellform og prediktorvariabler

Ulykkesmodellene er beregnet basert på data for riks- og fylkesvegnettet. Kun vanlige strekninger inngår i datagrunnlaget. Ettfeltstveger, rundkjøringer, ramper, tunneler og kommunale veger inngår ikke i datagrunnlaget. Segmenter hvor trafikkmengden (årsdøgntrafikk, ÅDT), antall år, lengde fartsgrense eller antall kjørefelt er uoppgitt inngår heller ikke i datagrunnlaget. Modellformen og de enkelte prediktorvariablene er beskrevet i de følgende avsnittene.

2.1 Modellform

Modellformen som er brukt for alle skadegrader er en negativ binomialmodell (NB modeller) med variabel overspredningsparameter. Modellformen er følgende:

$$E(n) = e^{\sum_i \text{Prediktor}_i * \text{Koeff}_i}$$

$E(n)$ er det predikerte (normale) ulykkestall (dvs. antall PSU, LS, HS, D eller D/HS), prediktorene er trafikkmengden og ulike vegegenskaper, i er subskript for prediktorene.

Negative binomialmodeller tar hensyn til overspredning i dataene, dvs. at spredningen er større enn man ville forvente hvis fordelingen av antall ulykker på enkelte vegsegmenter hadde fulgt en poissonfordeling. En negativ binomialfordeling er bestemt av to parametere, gjennomsnittet og overspredningsparameteren. En poissonfordeling derimot er bestemt av kun én parameter, gjennomsnittet. Legger man en poissonfordeling til grunn kan både koeffisientene og konfidensintervallene for koeffisientene være påvirket av overspredningen i dataene og dermed ikke være forventningsrette (skjeve).

En vanlig NB modell beregner en **fast overspredningsparameter**.

Overspredningsparameteren indikerer hvor mye de registrerte ulykkestallene varierer rundt de predikerte ulykkestallet (Hauer, 2001). Overspredningsparameteren kan brukes for å beregne vekten i EB analyser (se kapittel 1). Et problem med en fast overspredningsparameter er at predikerte ulykkestall er avhengige av i hvor lange delstrekninger man deler opp en strekning. Et annet problem er at man i EB analyser ikke har mulighet for å ta hensyn til at overspredningen varierer som en funksjon av strekningslengden, trafikkmengden mv. (Hauer, 2001).

Ved å beregne en **variabel overspredningsparameter** (istedenfor en fast overspredningsparameter som i vanlig NB regresjon) tar man hensyn til at overspredningen varierer som en funksjon av strekningslengde, antall år og trafikkmengde (Geedipally, Lord & Park, 2009). Ved bruk av en variabel overspredningsparameter vil problemene som er knyttet til en fast overspredningsparameter i mindre grad oppstå (Hauer, 2001; Lord & Park, 2008; Miaou & Lord, 2003).

De fleste prediktorvariablene i modellene er **dummyvariabler** (dvs. at variabelen har en verdi på enten null eller én). For fartsgrenser er det for eksempel definert syv dummyvariabler, én for hver fartsgrense, unntatt 80 km/t (fartsgrense 80 km/t er referansekategori og har ingen koeffisient). Grunnen til at mange variabler inngår i modellene som et sett med dummyvariabler er at man på denne måten bedre kan fange opp ikke-lineære sammenhenger mellom variabelen og antall ulykker eller skadde enn om man hadde definert én tallvariabel (for eksempel fartsgrense som én variabel som kan ha verdiene 30, 40, 50 mv.).

For dummyvariabler kan koeffisientene direkte brukes til å beregne endringen av den avhengige variabelen som en funksjon av prediktorvariabelen. Det relative antall ulykker på veger med fartsgrense X i forhold til veger med referansefartsgrensen (80 km/t) beregnes som:

$$\text{Rel. ant. ulykker} = e^{\text{Koeff.}}$$

Koeff. er koeffisienten for fartsgrense X. Eksempelvis har fartsgrense 60 km/t en koeffisient på 0,298 for PSU. Veger med fartsgrense 60 km/t har følgelig $e^{0,298} = 1,347$ ganger så mange PSU som veger med fartsgrense 80 km/t (hvis alt annet er likt).

De fleste øvrige variablene er definert som **tallvariabler** og inngår som $\text{Ln}(\text{antall} + 1)$ i modellene. Dette gjelder variablene kryss, rundkjøringer, ramper, kurver og stigninger. Den naturlige logaritmen brukes for å unngå at det predikerte antall ulykker stiger eksponentielt som en funksjon av prediktorvariabelen. Antall +1 brukes for å unngå å ta logaritmen av null. Det relative antall ulykker på en veg med X enheter (for eksempel kryss) i forhold til en veg med Y enheter kan beregnes som en funksjon av X, Y og koeffisienten:

$$\text{Rel. ant. ulykker} = \frac{e^{\text{Ln}(X+1)*\text{Koeff.}}}{e^{\text{Ln}(Y+1)*\text{Koeff.}}}$$

Trafikkmengden inngår med prediktorene $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$ i modellen (dette er nærmere forklart i avsnitt 2.3). Det relative antall ulykker på en veg med trafikkmengde X i forhold til en veg med trafikkmengde Y kan man beregne som en funksjon av trafikkmengdene X og Y, samt koeffisientene for $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$:

$$\text{Rel. ant. ulykker} = \frac{e^{\text{Ln}(X)*\text{Koeff.a} + \text{Ln}(X)^2*\text{Koeff.b}}}{e^{\text{Ln}(Y)*\text{Koeff.a} + \text{Ln}(Y)^2*\text{Koeff.b}}}$$

Koeff.a og Koeff.b er koeffisientene for henholdsvis $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$.

2.2 Ulykker og skadde

Følgende variabler brukes som avhengige variabler i modellberegningene: Antall PSU, antall LS, antall HS (meget alvorlig eller alvorlig skadde), antall D og antall D/HS. Det beregnes egne modeller for hver av disse variablene. Tabell 2.2.1 viser deskriptiv statistikk for alle fem variablene. De fleste segmentene har ingen PSU, skadde eller drepte. Det forholdsvis store standardavviket for antall LS kan være en følge av tilfeldig høye antall i noen segmenter.

Tabell 2.2.1: Deskriptiv statistikk for antall PSU, LS, HS, D og D/HS.

	PSU	LS	HS	D	D/HS
Alle år (2006-2011)					
Sum (totalt antall)	25 211	32 915	3 257	898	4 155
Gjennomsnitt	0,345	0,450	0,045	0,012	0,057
Standardavvik	0,985	1,462	0,256	0,125	0,304
Min.	0	0	0	0	0
Maks.	33	48	7	5	10
Antall segmenter med null	58 531	59 675	70 540	72 372	69 977
Andel segmenter med null	80 %	82 %	96 %	99 %	96 %
Per år					
Sum (antall per år)	4 586	6 008	581	158	739
Gjennomsnitt	0,063	0,082	0,008	0,002	0,010
Standardavvik	0,190	0,280	0,048	0,023	0,056
Min.	0,0	0,0	0,0	0,0	0,0
Maks.	16,2	16,2	2,0	1,1	2,0
Antall per mill. kjøretøy-km	0,149	0,195	0,019	0,005	0,024

2.3 Segmentlengde og antall år

Produktet av segmentlengde og antall år inngår i alle modellene som eksponeringsvariabel. Prediktorvariabelen i modellene er den naturlige logaritmen av produktet av segmentlengde og antall år og koeffisienten for denne variabelen settes i alle modellene lik 1 (i statistikkprogrammet som ble brukt til modellberegningene, Stata, er det kun mulig å definere én eksponeringsvariabel, derfor er produktet brukt istedenfor hver variabel for seg).

Hadde man estimert koeffisienter for segmentlengde og antall år sammen med koeffisientene for de øvrige prediktorvariablene, hadde antall predikert ulykker / skader ikke nødvendigvis variert proporsjonalt med segmentlengde og antall år. Dette forutsetter at segmentlengde og antall år i seg selv er relevante prediktorvariabler.

Segmentlengden kan ifølge Hauer (2001) være korrelert med bl.a. fart og avkjørselstetthet. Hvis man estimerer koeffisienter for segmentlengde uten at disse settes lik 1, mister man imidlertid muligheten for å generalisere resultatene fra modellberegningene på vegstrekninger som er lengre enn segmentene i ulykkesdatafilen. Koeffisientene vil også være avhengige av hvordan vegnettet er delt opp i segmenter og vil følgelig kunne endre seg ved en annen oppdeling. Det er liten grunn til å tro at segmentlengden i seg selv er en relevant prediktor (Elvik, 2012).

Antall år kan i det foreliggende datamaterialet ha sammenheng med vegegenskaper og med andre faktorer som påvirker ulykkesrisikoen. Ved å bruke antall år som en prediktor og ved å ikke sette koeffisienten lik 1, vil man i en viss grad kunne kontrollere for en slik sammenheng mellom periodelengde og antall ulykker / skader. Det har imidlertid vist seg at man trolig "overkontrollerer" trenden, dvs. at virkningen av en del variabler som har endret seg over tid fanges opp av prediktoren antall år istedenfor av de aktuelle variablene; dette viser seg bl.a. i usannsynlig store koeffisienter for $\ln(\text{antall år})$ hvis disse ikke er satt lik 1.

Deskriptiv statistikk for segmentlengde og periodelengde er vist i tabell 2.3.1. Datafilen som brukes til utviklingen av ulykkesmodeller inneholder 73 170 segmenter (rader) og data for til sammen 51 358 km veg. Den opprinnelige datafilen (med armer, ramper, rundkjøringer, tunneler, segmenter med $\text{ÅDT} = 1$ og lignende) inneholder data for 54 824 km veg. Dette betyr at 3 466 km veg, eller 6,3% av veglengden i den opprinnelige datafilen, er blitt «renset bort». Det som er blitt renset bort er segmenter som armer, ramper, rundkjøringer, tunneler og segmenter med $\text{ÅDT}=0$ eller antall år=0 eller manglende verdier på antall år, ÅDT , lengde, fartsgrense eller antall kjørefelt.

Tabell 2.3.1: Segmentlengde og periodelengde; periodelengde (vektet) er vektet med antall kjøretøykilometer per år.

	Segmentlengde (m)	Periodelengde (år)
Antall segmenter	73 170	73 170
Sum	51 358 315	
Gjennomsnitt	702	5,564
Standardavvik	377	1,162
Min.	1	0,008
Maks.	1 199	6

At den vektete gjennomsnittlige periodelengden er 5,227 år betyr at segmentene i gjennomsnitt omfatter trafikkmengde fra årene 2006 til mars 2011. Midten av denne perioden er i midten av 2008 og resultatene gjelder derfor i gjennomsnitt året 2008.

2.4 Trafikkmengde

Trafikkmengden inngår med to prediktorvariablene i modellene: $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$. Med $\text{Ln}(\text{ÅDT})^2$ i tillegg til $\text{Ln}(\text{ÅDT})$, som i alle modellene har en negativ koeffisient, blir modellene noe bedre og det predikerte antall ulykker/skader øker mindre ved høy ÅDT enn i modellene med kun $\text{Ln}(\text{ÅDT})$.

Det er gjort mange forsøk med alternative trafikkmengde-variabler, hvorav ingen har ført til en vesentlig forbedring av modellene (bl.a. dummyvariabler for høy/middels/lav ÅDT , flere ÅDT -variabler for høy, middels og lav ÅDT , $1/\text{ÅDT}$ og interaksjonseffekter mellom $\text{Ln}(\text{ÅDT})$ og vegtype / vegklasse). Av de alternative trafikkmengde-prediktorene ga interaksjonsvariablene $\text{Ln}(\text{ÅDT}) * \text{vegtype/-klasse}$ de mest lovende resultater. Her var det beregnet én variabel for hver vegtype/-klasse som er lik $\text{Ln}(\text{ÅDT})$ på segmenter i den respektive vegklassen / av den respektive vegtypen, og null ellers. Modellene ser ut til å predikere antall PSU og LS noe bedre enn modellene med $\text{Ln}(\text{ÅDT})$ og $\text{Ln}(\text{ÅDT})^2$. Forskjellene er imidlertid verken store eller konsistente mellom ulike skadegrader og modellvarianter. Ulempene med interaksjonsvariablene er at koeffisientene blir tilsynelatende ulogiske og vanskelige å tolke og at modellene trolig er mindre robuste ved endringer av trafikkmengden (resultatene gjelder kun ved den fordelingen av trafikkmengden på hver av vegtypene / -klassene som i det datasettet som ble brukt til å utvikle modellene).

Deskriptiv statistikk for trafikkmengden er vist i tabell 2.4.1. Tabell 2.4.2 viser hyppighetsfordelingen for ÅDT. Gjennomsnittlig ÅDT når man ikke tar hensyn til segmentlengden er 2690, mens den er 1643 når man tar hensyn til segmentlengden. Forskjellen skyldes at de lengste segmentene har lavest ÅDT. Det samlede årlige antall kjøretøykilometer er 30799 millioner. Dette gjelder datamaterialet som er brukt i modellberegningene. Det totale antall kjøretøykilometer på alle vegene i Norge (inkludert kommunale veger, rundkjøringer, ramper mv.) er anslagsvis 44000 millioner. Tabell 2.4.2 viser at over halvparten (57,2%) av segmentene har en ÅDT på under 1000 og at 93% av segmentene har en ÅDT på under 10000.

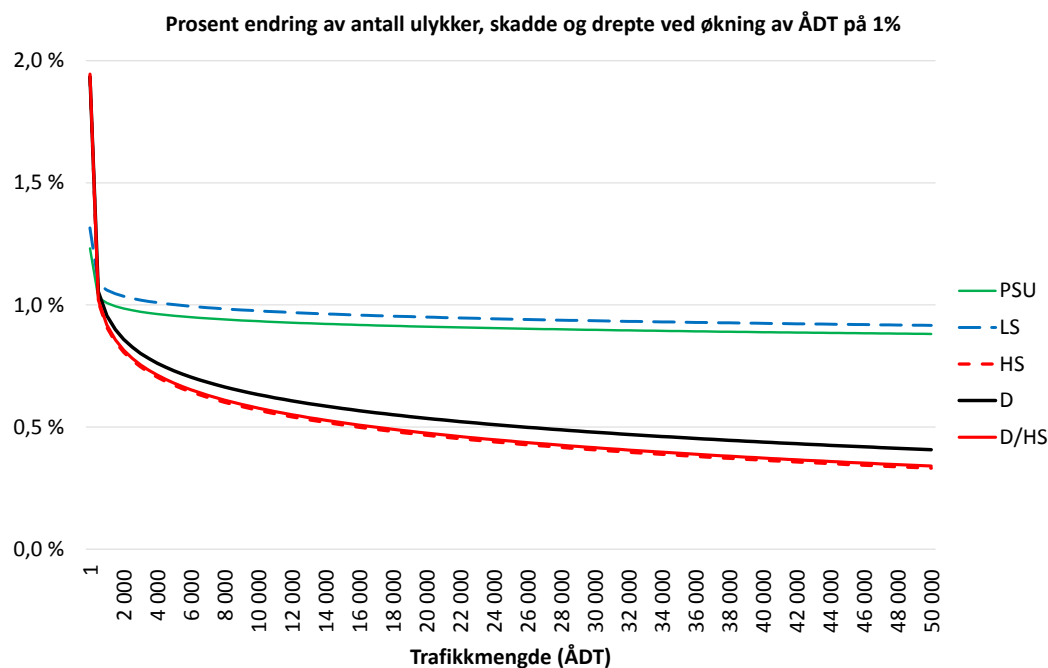
Tabell 2.4.1: Deskriptiv statistikk for ÅDT.

	ÅDT
Antall	73 170
Gjennomsnitt (uvektet)	2 690
Standardavvik	5 726
Gjennomsnitt (vektet med segmentlengde)	1 643
Min	1
Maks	92 957

Tabell 2.4.2: Hyppighetsfordeling av ÅDT med samlet segmentlengde.

ÅDT	Antall	Andel	Kum. andel	Gjsn. lengde (m)	Samlet lengde (km)	Andel av hele veglengden	Mill. kjøt-km per år	Andel
< 100	5 554	7,6 %	7,6 %	869	4 828	9,4 %	131	0,4%
100-499	24 645	33,7 %	41,3 %	836	20 607	40,1 %	2 042	6,6%
500-999	11 648	15,9 %	57,2 %	766	8 925	17,4 %	2 359	7,7%
1 000-1 999	10 455	14,3 %	71,5 %	706	7 384	14,4 %	3 866	12,6%
2 000-2 999	5 490	7,5 %	79,0 %	629	3 454	6,7 %	3 071	10,0%
3 000-3 999	2 923	4,0 %	83,0 %	552	1 615	3,1 %	2 051	6,7%
4 000-4 999	1 938	2,6 %	85,6 %	505	980	1,9 %	1 599	5,2%
5 000-9 999	5 419	7,4 %	93,0 %	387	2 097	4,1 %	5 320	17,3%
10 000-19 999	3 825	5,2 %	98,3 %	278	1 062	2,1 %	5 181	16,8%
20 000-39 999	935	1,3 %	99,5 %	303	283	0,6 %	2 685	8,7%
40 000-59 999	221	0,3 %	99,8 %	379	84	0,2 %	1 467	4,8%
60 000-79 999	96	0,1 %	100,0 %	299	29	0,1 %	709	2,3%
80 000-	21	0,0 %	100,0 %	487	10	0,0 %	318	1,0%
Total	73 170			538	51 358		30 799	

Sammenhengen mellom trafikkmengde og den predikerte økningen av antall ulykker, skadde og drepte er vist i figur 2.4.1. Figuren viser at antall ulykker, skadde og drepte øker med mindre enn 1 % for hver prosent økning av trafikkmengden og at økningen er jo mindre desto høyere trafikkmengden er, unntatt ved lav ÅDT. Ved ÅDT på 1200 eller mindre øker antall PSU mer enn trafikkmengden. For LS går grensen ved ÅDT 5000 og for HS, D og D/HS ved 500.



Figur 2.4.1: Sammenhengen mellom trafikkmengde og den predikerte økningen av antall ulykker, skadde og drepte.

2.5 Fartsgrense

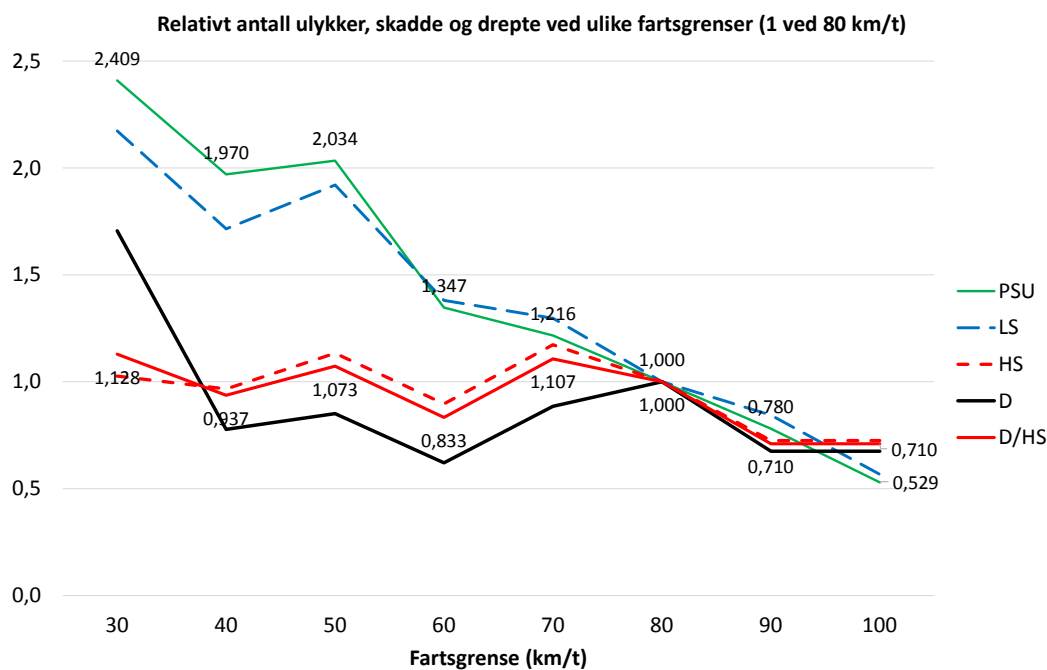
Fartsgrensen inngår i modellene med et sett dummyvariabler som prediktorvariabler. For hver fartsgrense er det definert en dummyvariabel som er 1 ved den respektive fartsgrensen og 0 ellers. Fartsgrensen 80 er referansekategorien (ikke inkludert i modellene). I modellene for D, HS og D/HS er fartsgrensene 90 og 100 slått sammen til én dummyvariabel. Det er så få HS og D ved disse fartsgrensene at separate dummyvariabler ikke gir meningsfulle resultater.

Dummyvariabler er valgt framfor en tallvariabel fordi det kan være generelle forskjeller mellom veger med ulike fartsgrenser som gjør at man ikke kan forvente at antall ulykker endrer seg monotont (eller til og med lineært) med fartsgrensen. Bl.a. er veger med ulike fartsgrenser ofte i ulike trafikkmiljøer. F.eks. er veger med fartsgrense 50 km/t eller lavere oftere enn andre veger i tettbygde strøk og har mer fotgjengertrafikk. En annen forklaring er at fartsgrensen ofte er satt ned på veger med spesielt mange ulykker (Strathman et al., 2001).

Deskriptiv statistikk for fartsgrensen er vist i tabell 2.5.1. Sammenhengen mellom fartsgrense og det predikerte relative antall ulykker, skadde og drepte er vist i figur 2.5.1. Figuren viser at sammenhengen ikke er lineær, især for drepte og HS.

Tabell 2.5.1: Hyppighetsfordeling av fartsgrense, med lengde, gjennomsnitts-ÅDT (vektet med segmentlengde) og antall kjøretøykilometer.

Fartsgrense	Antall	Andel	Samlet lengde (m)	Andel av hele veglengde	ÅDT	Mill. kjøt-km per år	Andel av alle kjøtkm
30 km/t	970	1,3%	256 669	0,5%	2 012	188	0,6%
40 km/t	2 523	3,4%	693 066	1,3%	3 068	776	2,5%
50 km/t	10 926	14,9%	4 786 462	9,3%	2 271	3 968	12,9%
60 km/t	13 634	18,6%	7 939 796	15,5%	1 954	5 663	18,4%
70 km/t	3 317	4,5%	1 732 884	3,4%	5 188	3 282	10,7%
80 km/t	40 146	54,9%	34 806 297	67,8%	1 032	13 115	42,6%
90 km/t	1 169	1,6%	901 300	1,8%	4 981	1 639	5,3%
100 km/t	485	0,7%	241 841	0,5%	24 566	2 169	7,0%
Total	73 170		51 358 315		1 643	30 799	



Figur 3.5.1: Sammenhengen mellom fartsgrense og det relative antall ulykker, skadde og drepte (relative antall skadde / drepte = 1 ved 80 km/t).

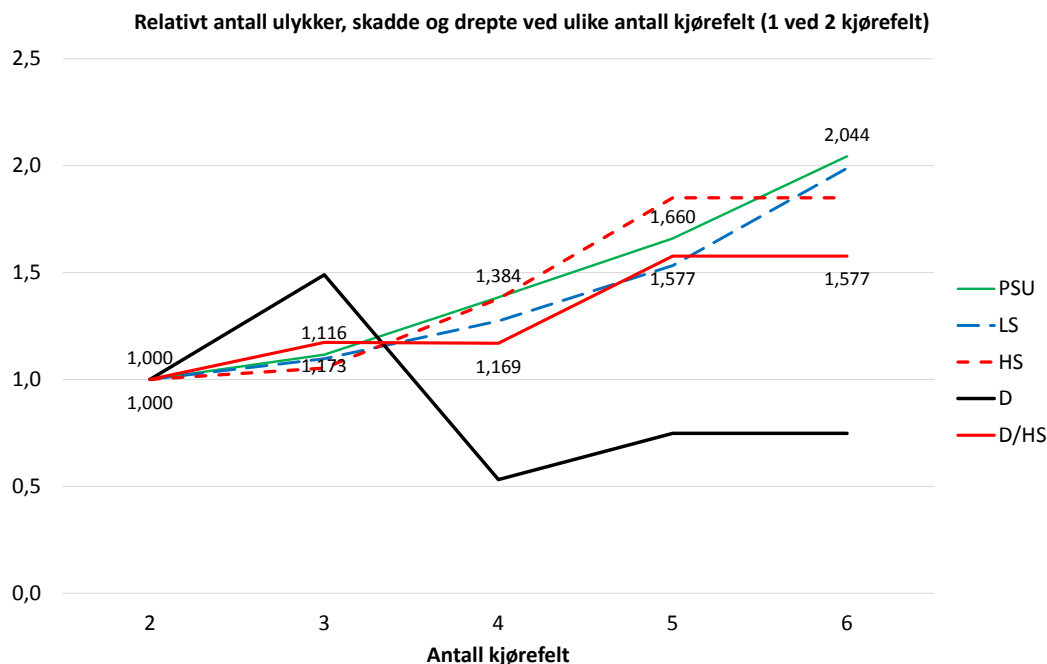
2.6 Antall kjørefelt

Prediktorvariabler for antall kjørefelt er definert på samme måte som for fartsgrense. For antall kjørefelt er det definert dummyvariabler. Segmenter med kun ett kjørefelt er utelatt fra datasettet fordi dette er en liten og svært heterogen gruppe av helt ulike typer vegger og dermed ikke noen meningsfylt kategori. Også ramper er utelatt fra datasettet. I modellene for PSU og LS er alle segmenter med seks eller flere kjørefelt slått sammen i én dummyvariabel og i modellene for HS og D er alle segmenter med fem eller flere kjørefelt slått sammen i én dummyvariabel.

Deskriptiv statistikk for antall kjørefelt er vist i tabell 2.6.1. Sammenhengen mellom antall kjørefelt og det predikerte relative antall ulykker, skadde og drepte er vist i figur 2.6.1. Figuren viser at det predikerte antall ulykker, skadde og drepte øker med økende antall kjørefelt (hvis alt annet er likt, inklusive trafikkmengde) og at sammenhengen ikke er lineær. For antall drepte avviker sammenhengen mellom fartsgrense og predikerte antall ulykkes, skadde og drepte forholdsvis mye de øvrige skadegradene. Forklaringen er at det er kun få drepte på vegger med fem eller flere kjørefelt og dermed stor tilfeldig variasjon i de registrerte antallene.

Tabell 2.6.2: Hyppighetsfordeling av antall kjørefelt, med lengde, gjennomsnitts-ÅDT (vektet med segmentlengde) og antall kjøretøykilometer.

Antall kjørefelt	Antall	Andel	Samlet lengde (m)	Andel av hele veglengden	ÅDT	Mill. kjøretøy-km per år	Andel av alle kjtkm
2	70 672	96,6 %	50 650 854	98,6 %	1 348	24 924	80,9%
3	825	1,1 %	207 721	0,4 %	9 607	728	2,4%
4	1 418	1,9 %	445 660	0,9 %	25 256	4 108	13,3%
5	124	0,2 %	22 048	0,0 %	40 470	326	1,1%
6+	131	0,2 %	32 032	0,1 %	60 983	713	2,3%
Total	73 170		51 358 315		1 643	30 799	



Figur 2.6.1: Sammenhengen mellom antall kjørefelt og det predikerte relative antall ulykker, skadde og drepte.

2.7 Kryss, rundkjøringer og ramper

Prediktorvariabler for X-kryss, T-kryss, rundkjøringer, av-ramper, på-ramper og uspesifiserte ramper er den naturlige logaritmen av antallet kryss/rundkjøringer/ramper pluss én (pluss én for å unngå å ta logaritmen av null). Slike variabler har vist seg å gi mer logiske resultater enn andre typer variabler (f.eks. utransformert eller kvadrert antall). Deskriptiv statistikk er vist i tabell 2.7.1 for plankryss og rundkjøringer og i tabell 2.7.2 for planskilte kryss (ramper).

Plankryss: X-kryss er firearmede plankryss og T-kryss er trearmede plankryss. En del av T-kryssene i datamaterialet kan være avkjørsler. Det skiller ikke mellom ulike typer kryssregulering. Kryss er i datamaterialet alltid lagt til hovedvegen i krysset. Det betyr at sidevegene i datamaterialet ikke har noen kryss. Dette medfører en viss underestimering av antall ulykker på sidevegene, men ulykkene er som regel uansett stedfestet på hovedvegen. Ulempen ved en slik underestimering er vurdert som mindre enn hvis alle kryss hadde vært lagt til både hoved- og sidevegene fordi da hadde alle ulykkene i kryss inngått to ganger i datamaterialet og modellberegningen.

Rundkjøringer: Rundkjøringer omfatter alle typer rundkjøringer. Rundkjøringene inngår ikke i datamaterialet, men det er gjort en opptelling av antall rundkjøringer som ligger på hvert segment. På samme måte som for plankryss er rundkjøringene alltid lagt til hovedvegen. Ulykkene som skjer i rundkjøringer er også lagt til segmentet på hovedvegen.

Planskilte kryss: Av- og påramper er ramper til og fra hovedveger med planskilte kryss. Uspesifiserte ramper er ramper hvor det i datafilen ikke er spesifisert om det er av- eller påramper (uspesifiserte ramper er ikke relevante for brukere av modellen).

Sammenhengen av plankryss og rundkjøringer med antall ulykker kan være forskjellig avhengig av bl.a. fartsgrense, vegtype, om vegen har skille mellom kjøreretningene eller ikke og kurver og stigninger. Slike interaksjonseffekter er imidlertid ikke tatt hensyn til i modellene.

Tabell 2.7.1: Deskriptiv statistikk og hyppighetsfordeling for kryss, rundkjøringer og ramper.

	X-kryss		T-kryss		Rundkjøringer	
	Antall	Andel	Antall	Andel	Antall	Andel
Antall kryss / rundkjøringer						
Totalt antall	2 034		33 358		1 112	
Gjennomsnitt	0,028		0,456		0,015	
Standardavvik	0,200		0,971		0,123	
Antall per km veg	0,040		0,650		0,022	
Segmenter med ... kryss / rundkjøringer						
0	71 416	97,60 %	52 359	71,6 %	72 067	98,5 %
1	1 572	2,15 %	14 144	19,3 %	1 094	1,5 %
2	128	0,17 %	3 892	5,3 %	9	0,01 %
3	32	0,04 %	1 363	1,9 %		
4	15	0,02 %	673	0,9 %		
5	1	0,001 %	331	0,5 %		
6	2	0,003 %	163	0,2 %		
7			113	0,2 %		
8	3	0,004 %	57	0,1 %		
9	1	0,001 %	35	0,05 %		
10-18			40	0,1 %		
Alle	73 170	100 %	73 170	100 %		

Tabell 2.7.2: Deskriptiv statistikk og hyppighetsfordeling for toplankryss (ramper).

	Av-ramper		På-ramper		Uspesifiserte ramper	
	Antall	Andel	Antall	Andel	Antall	Andel
Antall ramper						
Totalt antall	672		643		372	
Gjennomsnitt	0,0092		0,0088		0,0051	
Standardavvik	0,1078		0,1047		0,0957	
Antall per km veg	0,013		0,013		0,007	
Segmenter med ... ramper						
0	72 585	99,2 %	72 603	99,2 %	72 924	99,7 %
1	503	0,687 %	496	0,678 %	141	0,193 %
2	77	0,105 %	67	0,092 %	87	0,119 %
3	5	0,007 %	3	0,004 %	15	0,021 %
4			1	0,001 %	3	0,004 %

2.8 Kurver

Hver del av vegstrekningen på 50 m lengde med kurveradius under 300 m (uansett kurveretning) regnes som én **kurve**. Hvis en kurve oppfyller kriteriene, men er fordelt over to segmenter tilhører kurven det segmentet hvor mesteparten av kurven ligger. Eksempel: Av en 50 m lang kurve ligger 27 m på segment A og 23 m på segment B. Da har segment A én kurve og segment B har ingen kurve. Retningsendringer inngår ikke i definisjonen av kurver, dvs. at antall kurver er uavhengig av hvor mange ganger kurven endrer sin retning og man kan ikke telle antall kurver ved å telle retningsendringer.

For å ta hensyn til at kurver kan ha ulik sammenheng med antall ulykker / skader ved ulike fartsgrenser er det definert én kurvevariabel per fartsgrense. Følgende prediktorvariabler for kurver er definert:

- I modellene for PSU og LS: $\ln(\text{antall kurver} + 1)$ på veg med fartsgrense 30 km/t, 40 km/t, 50 km/t, 60 km/t, 70 km/t, 80 km/t og 90 eller 100 km/t
- I modellene for HS, drepte og D/HS: $\ln(\text{antall kurver} + 1)$ på veg med fartsgrense 30 eller 40 km/t, 50 km/t, 60 km/t, 70 km/t, 80 km/t og 90 eller 100 km/t

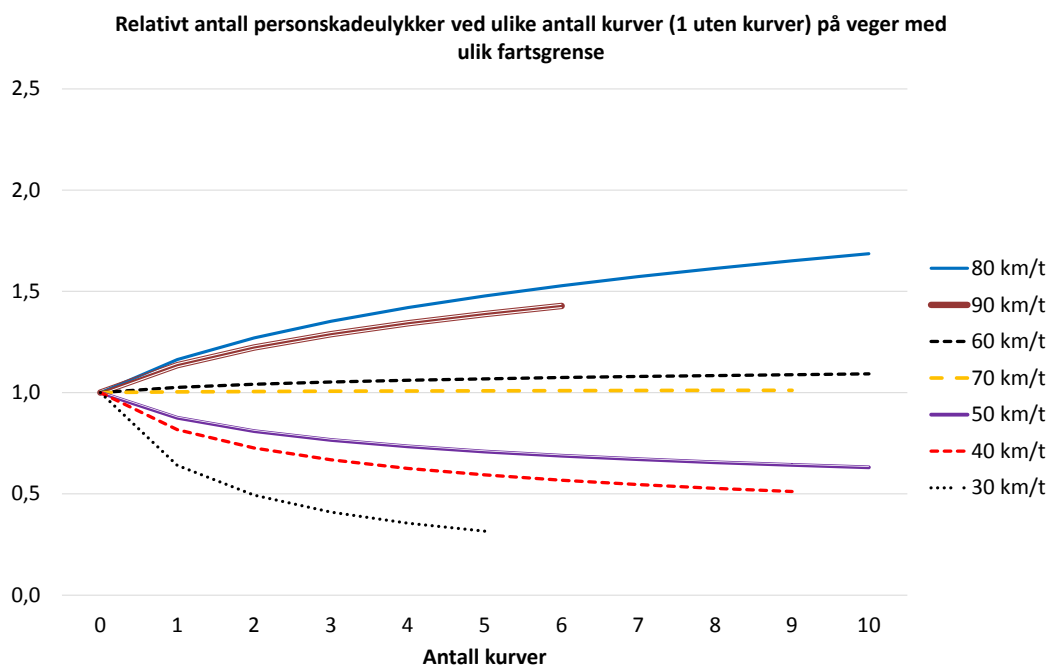
Kurvevariablene sier lite om veggeometrien fordi det ikke er mulig å ta hensyn til retningsendringer. En strekning med syv "kurver" etter denne definisjonen kan ha alt fra én langstrakt kurve til syv korte krappe kurver med til sammen åtte retningsendringer. Antall kurver sier følgelig mer om hvor mange meter veg i ett segment som ligger i en kurve / i kurver enn hvor mange retningsendringer segmentet har.

Det ble gjort forsøk med å skille mellom enslige kurver og kurver på kurverike strekninger (enslige kurver har som regel høyere risiko enn kurver som har flere andre kurver i nærheten). Dette ga imidlertid ikke meningsfulle resultater, trolig fordi kurvevariablene ikke tar hensyn til retningsendringer. Mange strekninger vil derfor ha «mange kurver», selv om det kun finnes én eller få lange kurver uten retningsendring.

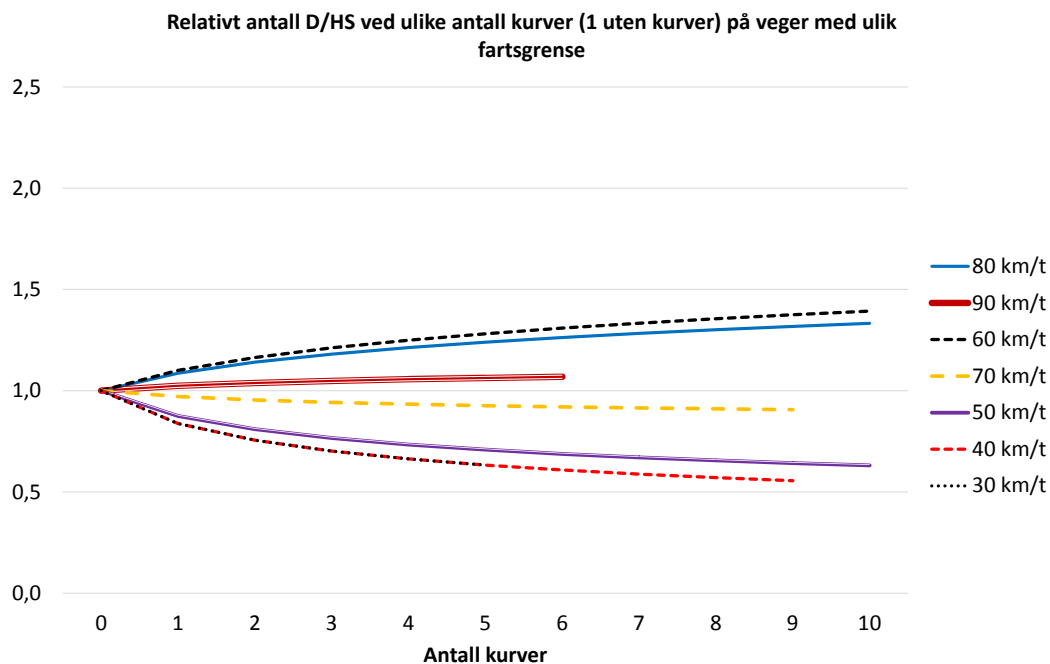
Deskriptiv statistikk for kurver er vist i tabell 2.8.1. De fleste kurvene ligger på veger med fartsgrense 80 km/t og på disse vegene er også andelen segmenter med minst én kurve størst. Sammenhengen mellom antall kurver og predikert antall PSU og D/HS på veger med ulik fartsgrense er vist i henholdsvis figur 2.8.1 og figur 2.8.2. Figurene viser at både antall PSU og antall D/HS øker med økende antall kurver på veger med fartsgrense 80 km/t eller høyere og at antall PSU og antall D/HS synker med økende antall kurver på veger med fartsgrense 50 km/t eller lavere. På veger med fartsgrense 60 eller 70 km/t er det kun en svak sammenheng. Sammenhengen mellom antall kurver og antall LS ligner på sammenhengen med antall PSU og sammenhengen mellom antall kurver og antall drepte og HS ligner på sammenhengen med antall D/HS.

Tabell 2.8.1: Deskriptiv statistikk for kurver.

Fartsgrense	Totalt antall	Kurver per segment	Antall med null kurver	Andel med null kurver	Kurver per kilometer
30 km/t	604	0,623	598	62%	2,353
40 km/t	1 849	0,733	1 550	61%	2,668
50 km/t	12 343	1,130	5 239	48%	2,579
60 km/t	17 032	1,249	6 123	45%	2,145
70 km/t	2 405	0,725	2 162	65%	1,388
80 km/t	70 190	1,748	13 374	33%	2,017
90 / 100 km/t	174	0,105	1 537	93%	0,152
Alle	104 598	1,430	30 583	42%	2,037



Figur 2.8.1: Sammenhengen mellom antall kurver og predikert antall PSU på vegger med ulike fartsgrenser.



Figur 2.8.2: Sammenhengen mellom antall kurver og predikert antall D/HS på veger med ulike fartsgrense.

Dersom antall kurver er ukjent er det mulig å oppgi null kurver. Alle segmentene sett under ett er avvikene mellom registrerte og normale ulykkestall ikke vesentlig større når man oppgir null kurver for alle segmentene i beregningen av normale ulykkestall enn når man oppgir det faktiske antall kurver. Det samme gjelder når man oppgir det gjennomsnittlige antall kurver per fartsgrense.

På strekninger hvor man vet at det finnes kurver, men hvor man ikke vet hvor mange det er etter den definisjonen som ligger til grunn for kurvevariabelen, kan man også oppgi et estimert antall kurver. Dermed vil det beregnede normale ulykkestall være høyere eller lavere (avhengig av fartsgrensen) enn hvis man hadde oppgitt null kurver. På veger med fartsgrense 70 km/t vil det imidlertid ikke gi noe stor utslag hvor mange kurver man oppgir.

2.9 Stigninger

Hver del av vegstrekningen på 200 m lengde med en stigning på minst 4% regnes som én stigning. For å ta hensyn til at stigninger kan ha ulike sammenheng med antall ulykker / skader er det definert én stigningsvariabel per fartsgrense. Følgende prediktorvariabler for stigninger er definert:

- I modellene for PSU og LS: $\ln(\text{antall stigninger} + 1)$ på veg med fartsgrense 30 eller 40 km/t, 50 km/t, 60 km/t, 70 km/t, 80 km/t og 90 eller 100 km/t.
- I modellene for HS, D og D/HS: $\ln(\text{antall stigninger} + 1)$.

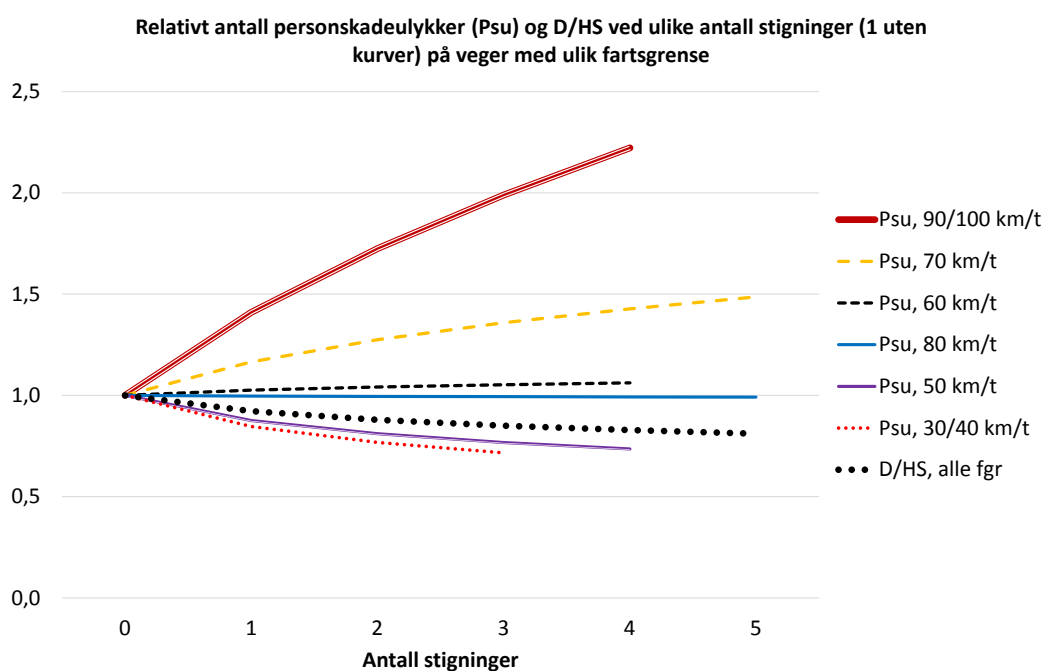
I modellene for HS og D brukes kun én stigningsvariabel (på veger med alle fartsgrenser) fordi det er for få HS / D på segmenter med stigninger og spesifikke fartsgrenser.

Problemene med stigningsvariablene er de samme som med kurvevariabelen, det er ikke tatt hensyn til retningsendringer (høy-/lavbrekk).

Deskriptiv statistikk for stigninger er vist i tabell 2.9.1. De fleste stigninger ligger på veger med fartsgrense 80 km/t. Sammenhengen mellom antall stigninger og predikert antall PSU og D/HS på veger med ulik fartsgrense er vist i henholdsvis figur 2.9.1. Figurene viser at antall PSU øker med økende antall stigninger på veger med fartsgrense 70, 90 og 100 km/t og at antall PSU og antall D/HS synker med økende antall kurver på veger med fartsgrense 50 km/t eller lavere. På veger med fartsgrense 60 eller 80 km/t er det kun en svak sammenheng. Antall D/HS synker ved økende antall stigninger.

Tabell 2.9.1: Deskriptiv statistikk for stigninger.

Fartsgrense	Totalt antall	Stigninger per segment	Antall med null stigninger	Andel med null stigninger	Stigninger per kilometer
30 / 40 km/t	473	0,135	3 063	88%	0,498
50 km/t	2 265	0,207	9 017	83%	0,473
60 km/t	3 111	0,228	11 056	81%	0,392
70 km/t	499	0,150	2 893	87%	0,288
80 km/t	16 395	0,408	27 442	68%	0,471
90 / 100 km/t	118	0,071	1 553	94%	0,103
Alle	22 861	0,312	55 024	75%	0,445



Figur 2.7.1: Sammenhengen mellom antall stigninger og predikert antall PSU og D/HS på veger med ulik fartsgrense.

2.10 Vegklasse

Det er definert fem dummyvariabler for ulike vegklasser:

1. Motorveg
2. Tofeltsveg med planskilte kryss (motortrafikkveg i tidligere terminologi)
3. TEN-T-veg (ikke motorveg eller tofeltsveg med planskilte kryss)
4. Øvrig europa-/riksveg (ikke motorveg eller tofeltsveg med planskilte kryss eller TEN-T veg)
5. Fylkesveg (referansekategori)

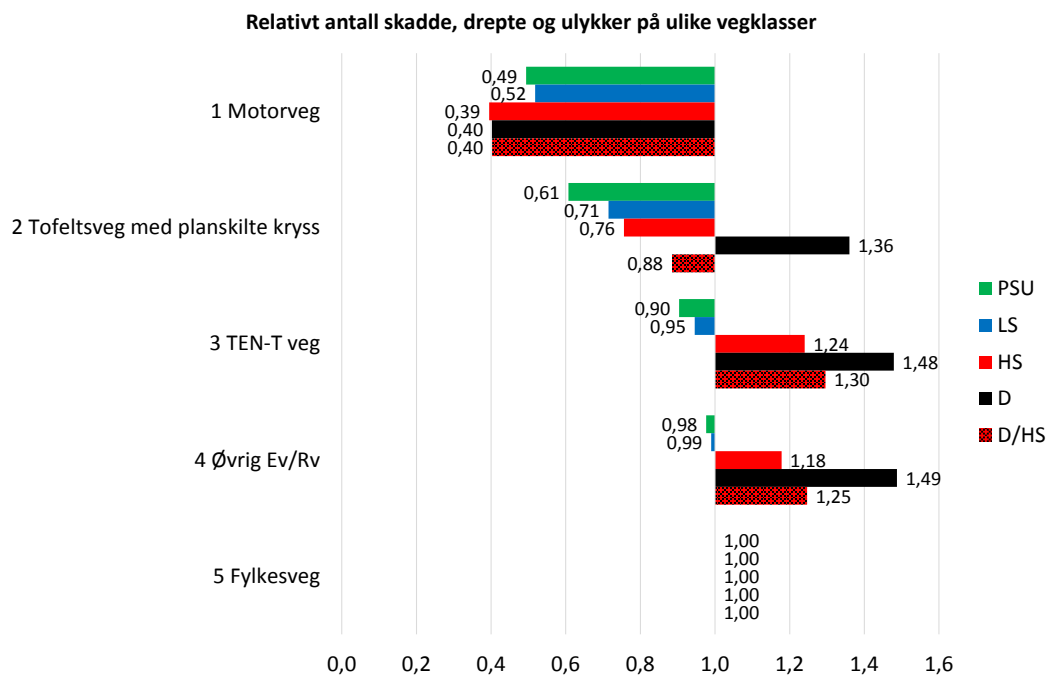
Disse vegklassene er valgt fordi det finnes generelle forskjeller i vegstandarden (bl.a. kjørefelt- og skulderbredde) mellom klassene. Eksempelvis har alle motorveger midtrekkverk og TEN-T veger har generelt en noe høyere standard enn øvrige Europa- og riksveger.

Deskriptiv statistikk for vegklasse er vist i tabell 2.10.1. Sammenhengen mellom vegklasse og relative ulykkestall er vist i figur 2.10.1. Motorveger har betydelig færre skadde, drepte og ulykker enn andre vegklasser. Dette gjelder selv om det er kontrollert for om vegene har midtdeler eller midtrekkverk (disse er egne prediktorvariabler i modellene).

Tabell 2.10.1: *Vegklasse, sammenheng med segmentlengde og ÅDT (vektet med segmentlengde).*

Vegklasse	N	Lengde (km)		ÅDT	Mill. kjøretøy-km per år	
		Gjennomsnitt	Sum		Andel av hele veglengden	N
1 Motorveg	777	440	341,9	0,7%	28 172	3 515 11,4%
2 Tofeltsveg med planskilte kryss	776	477	370,3	0,7%	11 548	1 561 5,1%
3 TEN-T	5 639	643	3 627,8	7,1%	3 857	5 107 16,6%
4 Øvrig Ev/Rv	7 247	632	4 581,8	8,9%	3 167	5 296 17,2%
5 Fylkesveg	58 731	723	42 436,5	82,6%	989	15 319 49,7%
Alle	73 170	702	51 358,3	100,0%	1 643	30 799

* Andel av alle kjøretøy-km



Figur 2.10.1: Relativt antall skadde, drepte og ulykker på ulike vegklasser (relativt antall på fylkesveger = 1).

2.11 Midtdeler og midtrekkverk

For midtdeler og midtrekkverk er det definert fire dummyvariabler som er prediktorvariabler i modellene:

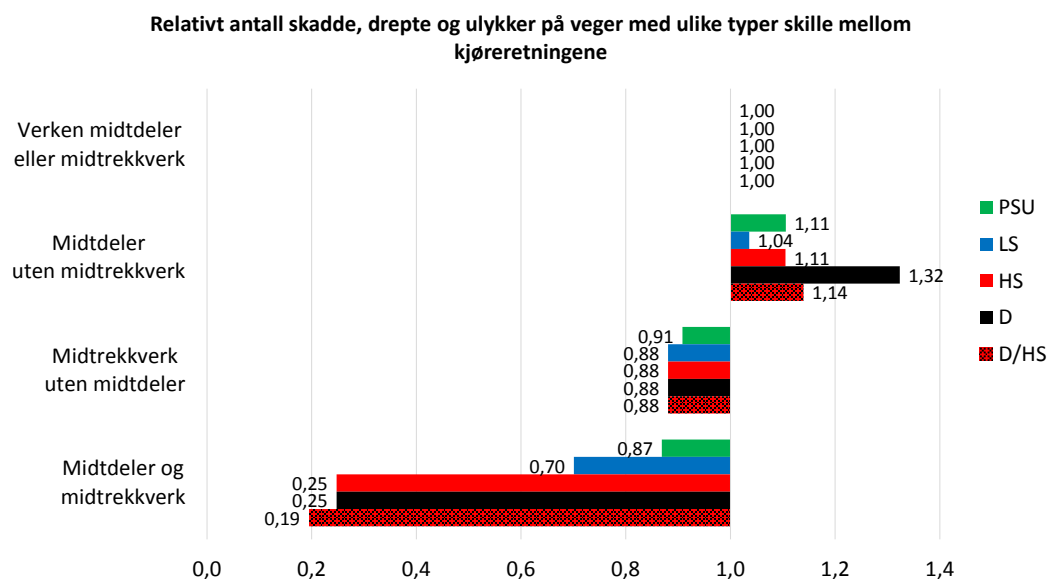
- Verken midtdeler eller midtrekkverk (referansekategori)
- Midtdeler, ikke midtrekkverk
- Midtrekkverk, ikke midtdeler
- Midtdeler med rekkverk (midtdeler og midtrekkverk)

Det er definert fire dummyvariabler istedenfor én for med/uten midtdeler og én for med/uten midtrekkverk fordi det kan være interaksjonseffekter mellom midtdeler og midtrekkverk, dvs. at virkningen av midtrekkverk kan være forskjellig på veger med og uten midtdeler og at virkningen av midtdeler kan være forskjellig på veger med og uten midtrekkverk.

Deskriptiv statistikk for midtdeler og midtrekkverk er vist i tabell 2.11.1. Figur 2.11.1 viser predikerte relative antall ulykker, skadde og drepte på veger med og uten midtdeler / midtrekkverk. På veger med midtdeler og midtrekkverk er det langt færre skadde og drepte enn på veger uten midtdeler eller midtrekkverk. For midtdeler eller midtrekkverk alene er det kun forholdsvis små effekter og disse er ikke statistisk signifikante. At motorvegene i datafilen har «Verken midtdeler eller –rekkverk» har ingen betydning for resultatene, virkningen av midtdeler og –rekkverk på motorveg er i sin helhet fanget opp av vegklassevariabelen.

Tabell 2.11.1: Middeler og midtrekkverk, deskriptiv statistikk og sammenheng med ÅDT (uten motorveger).

	Lengde (km)				ÅDT	Mill. kjt-km per år	
	N	Gjennomsnitt	Sum	Andel av hele veglengden		Gjennomsnitt	N
					Middeler med rekkverk		245
Kun middeler	2 510	145	362,7	0,71 %	15 418	2 041	7,05 %
Kun midtrekkverk	51	307	15,7	0,03 %	13 236	76	0,26 %
Verken middeler eller –rekkverk	70 364	723	50 548,9	98,42 %	1 331	24 566	84,90 %
Alle	72 393	51 016,4	99,33 %	1 456	27 283	94,29 %	



Figur 2.11.1: Relativt antall skadde, drepte og ulykker på vegger med ulike former for skille mellom kjøretretningene (relativt antall på vegger uten middeler eller midtrekkverk = 1).

2.12 Forsterket midtoppmerking

Forsterket midtoppmerking er kombinasjonen av vanlig oppmerking og rumleriller. For forsterket midtoppmerking er det definert to dummyvariabler som er prediktorvariabler i modellene:

- Bred forsterket midtoppmerking (over 1 m bred)
- Smal forsterket midtoppmerking (opptil 1 m bred)

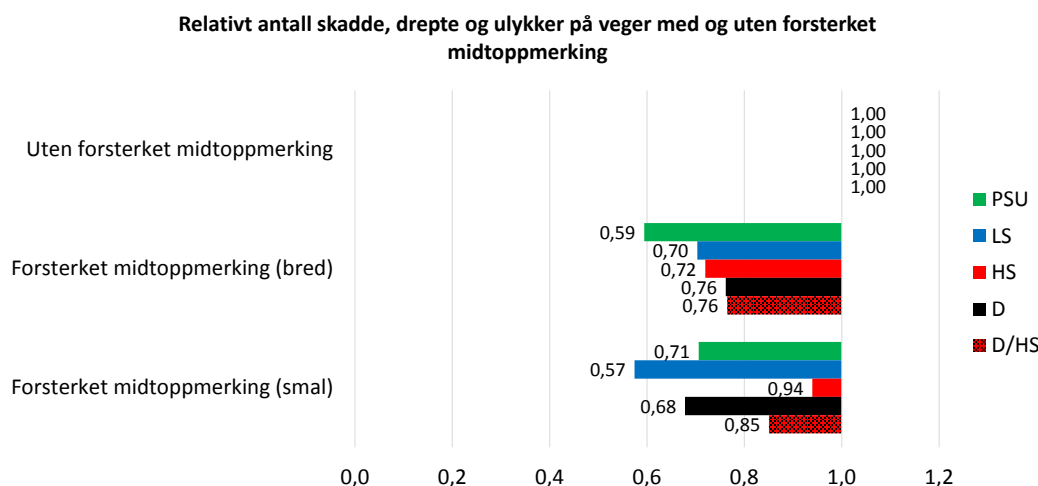
Veger uten forsterket midtoppmerking er referansekategori.

Deskriptiv statistikk forsterket midtoppmerking er vist i tabell 2.12.1. Figur 2.12.1 viser predikerte relative ulykestall på veger med og uten forsterket midtoppmerking. Forsterket midtoppmerking medfører reduserte ulykestall for alle skadegrader. Virkningene på antall HS og D er ikke statistisk signifikante. Dette betyr ikke nødvendigvis at virkningene er mindre, men skyldes trolig i hovedsak små antall D og HS på veger med forsterket midtoppmerking.

Tabell 2.12.1: Forsterket midtoppmerking, deskriptiv statistikk og sammenheng med ÅDT (vektet med sementlengde).

Forst. midtoppmerking	N	Lengde (km)		Andel av hele veglengden	ÅDT Gjennomsnitt	Mill. kjt-km per år	
		Gjennomsnitt	Sum			N	%*
Bred	200	556	111 291	0,2%	11 050	449	1,5%
Smal	228	617	140 700	0,3%	7 244	372	1,2%
Uten	72 742	703	51 106 324	99,5%	1 607	29 978	97,3%
Alle	73 170		51 358 315		1 643	30 799	

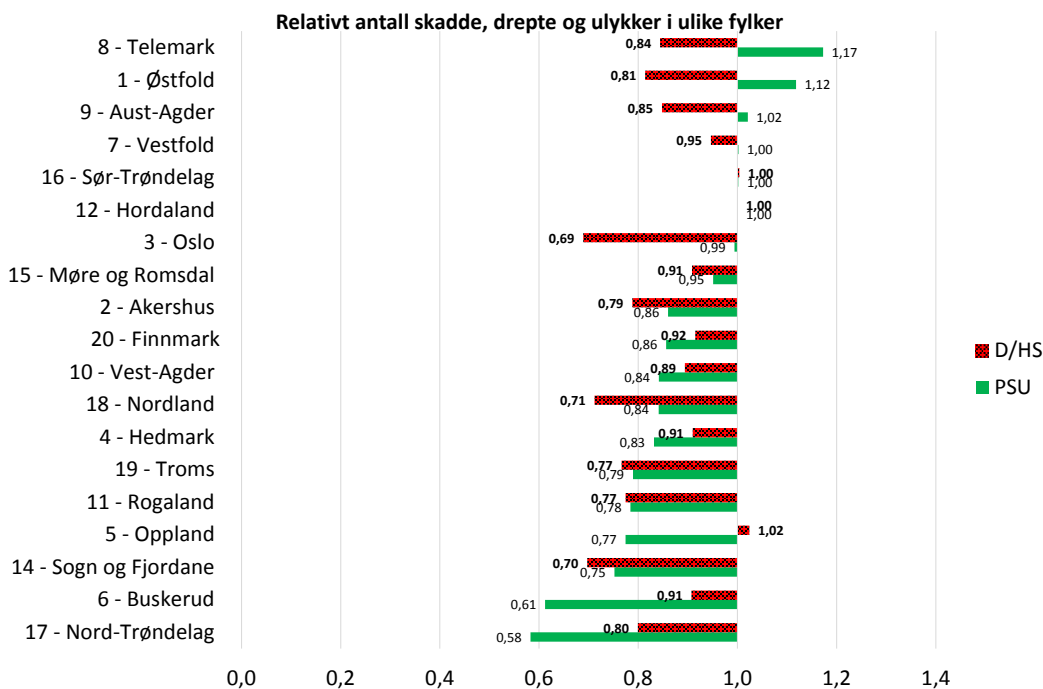
* Andel av alle kjt-km



Figur 2.12.1: Relativt antall skadde, drepte og ulykker på veger med og uten forsterket midtoppmerking (relativt antall på veger uten forsterket midtoppmerking = 1).

2.13 Fylke

For hvert fylke er det definert én dummyvariabel. Fylkesvariablene skal fange opp generelle forskjeller mellom fylkene som f.eks. forskjeller i topografi, vær og befolkningstetthet. Relative antall PSU og D/HS i de ulike fylkene er vist i figur 2.13.1. Det relative antall PSU og D/HS i Hordaland er lik 1 (dette er referansekategori). Fylkene er sortert etter relative antall PSU i figur 2.13.1. Relative antall LS i fylkene ligner på de relative antallene PSU og relative antall D og HS ligner på de relative antallene D/HS.



Figur 2.13.1: Relative antall PSU og D/HS i ulike fylker (relative antall i Hordaland = 1; sortert etter relativt antall PSU).

2.14 Konstanterterm

Alle modellene inneholder en konstanterterm. Denne er beregnet automatisk av statistikkprogrammet.

2.15 Sammenheng mellom ulike vegegenskaper

De følgende tabellene viser samlet veglengde (tabell 2.15.1) og antall kjøretøykilometer (tabell 2.15.2) etter fartsgrense, antall kjørefelt og skille mellom kjøreretningene for veger som ikke er motorveger. Samlet veglengde og kjøretøykilometer etter fartsgrense og antall kjørefelt for motorveger vises i tabell 2.15.3. Motorveger vises separat fordi det på de fleste motorvegstrekingene mangler informasjon om midtdeler / midtrekkverk.

Tabell 2.15.1: Samlet veglengde (km) for segmenter med fartsgrense 80 km/t eller høyere (uten motorveger), etter fartsgrense, antall kjørefelt og skille mellom kjøreretningene.

Fgr.	Felt	Ingen	Middeler, ikke rekkverk	Rekkverk, ikke midtdeler	Middeler med rekkverk	Forsterket midt-oppmerking	Alle
80	2	34 329,6	58,5	5,6	11,2	198,0	34 602,9
	3	70,2	10,7	4,0	2,5	9,0	96,3
	4	16,8	29,6	2,2	9,7	0,1	58,4
	5	1,5	3,3	-	1,3	-	6,1
	6	0,5	7,5	0,4	1,4	-	9,8
	Total	34 418,7	109,6	12,1	26,1	207,0	34 773,5
90	2	732,7	14,4	-	28,7	15,9	791,6
	3	8,4	2,7	0,1	20,4	1,0	32,7
	4	-	11,6	0,3	6,7	-	18,6
	Total	741,1	28,8	0,4	55,8	16,8	842,9
100	2	1,0	-	-	-	-	1,0
	4	-	0,2	-	0,0	-	0,2
	Total	1,0	0,2	-	0,0	-	1,1
Total	2	35 063,3	72,9	5,6	39,9	213,8	35 395,5
	3	78,6	13,4	4,1	22,9	9,9	129,0
	4	16,8	41,4	2,5	16,4	0,1	77,2
	5	1,5	3,3	-	1,3	-	6,1
	6	0,5	7,5	0,4	1,4	-	9,8
	Total Total	35 160,7	138,5	12,5	81,9	223,9	35 617,6

Tabell 2.15.2: Samlet årlig antall kjøretøykilometer (mill.) for segmenter med fartsgrense 80 km/t eller høyere (uten motorveger), etter fartsgrense, antall kjørefelt og skille mellom kjøreretningene.

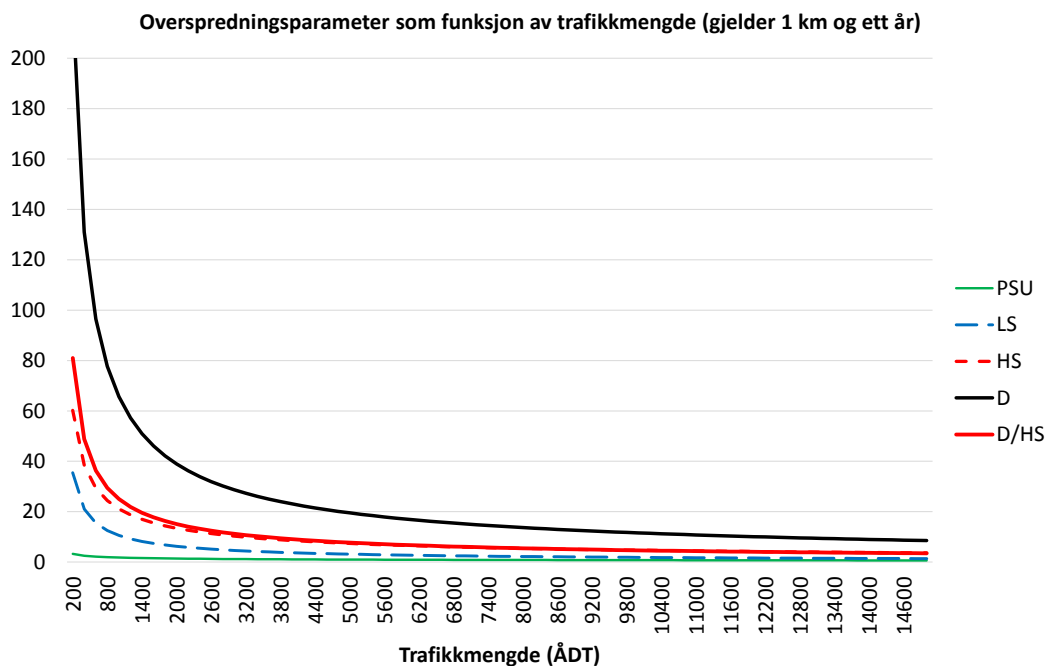
Fgr.	Felt	Ingen	Middeler, ikke rekkverk	Rekkverk, ikke midtdeler	Middeler med rekkverk	Forsterket midt-oppmerking	Alle
80	2	10 376	213	24	67	623	11 303
	3	158	35	11	8	39	251
	4	150	462	15	115	1	743
	5	34	53	0	23	0	110
	6	8	205	7	25	0	245
	Total	10 726	969	56	238	663	12 652
90	2	335	84	0	158	33	610
	3	16	7	0	90	2	115
	4	0	40	1	32	0	73
	Total	351	131	1	280	35	798
100	2	3	0	0	0	0	3
	4	0	1	0	0	0	1
	Total	3	1	0	0	0	4
Total	2	10 714	297	24	224	656	11 915
	3	174	42	11	98	41	366
	4	150	504	16	147	1	818
	5	34	53	0	23	0	110
	6	8	205	7	25	0	245
	Total Total	11 080	1 101	57	517	698	13 453

Tabell 2.15.3: Samlet veglengde (km) og årlig antall kjøretøykilometer (mill.) for segmenter med fartsgrense 80 km/t eller høyere (kun motorveger), etter fartsgrense og antall kjørefelt.

Fartsgrense	Felt	Lengde (km)	Mill, kjøretøykilometer per år
80	2	4,3	40,0
	3	0,2	2,0
	4	21,9	228,0
	6	6,4	194,0
	Total	32,8	464,0
90	2	4,6	30,0
	3	0,7	6,0
	4	43,4	577,0
	5	4,8	98,0
	6	4,9	130,0
	Total	58,4	841,0
100	2	8,7	37,0
	4	221,4	1 945,0
	5	6,7	88,0
	6	3,9	94,0
	Total	240,7	2 164,0
Total	2	17,5	107,0
	3	1,0	8,0
	4	286,7	2 750,0
	5	11,6	186,0
	6	15,1	418,0
	Total	331,9	3 469,0

2.16 Overspredningsparameter

Alle modellene inneholder koeffisienter for å beregne overspredningsparameteren som en funksjon av segmentlengde, antall år og trafikkmengde. Dermed tar man hensyn til at overspredningen synker med økende segmentlengde, antall år og trafikkmengde. Figur 2.16.1 viser overspredningsparameteren som en funksjon av trafikkmengden. Resultatene gjelder et vegsegment på 1 km lengde og ett år.



Figur 2.16.1: Overspredningsparameter som en funksjon av trafikkmengden (gjelder 1 km og ett år).

Overspredningsparameteren kan beregnes med hjelp av koeffisientene som er vist i neste avsnitt (tabell 3.1.1). For PSU beregnes overspredningsparameteren for eksempel etter følgende formel:

$$\text{Overspredningsparameter} = e^{8,529 + \ln(\text{lengde} * \text{år}) * -0,778 + \ln(\text{ÅDT}) * -0,371}$$

3 Modellene

3.1 Koeffisientene og beregningseksempel

Tabell 3.1.1 viser koeffisientene for alle prediktorvariablene i modellene for alle skadegrader. For enkelte variabler er koeffisientene for HS, D og / eller D/HS satt lik koeffisienten for den nest lavere skadegrad fordi statistikkprogrammet ikke klarte å beregne koeffisientene på grunn av for små antall skadde / drepte i enkelte kategorier av variablene.

For alle dummyvariablene er alltid en av kategoriene referansekategori, dvs. at det i modellen ikke er beregnet noen koeffisient og at alle koeffisientene for de øvrige kategoriene for denne variabelen viser forholdet mellom antall ulykker i denne kategorien og i referansekategorien.

For å vise hvordan modellen brukes for å beregne normale ulykkestall er det lagt et beregningseksempel. Eksempelen viser beregningen av det normale antall ulykker i løpet av ett år for en vegstrekning med følgende egenskaper:

Lengde (meter):	5 000
Trafikkmengde:	10 000
Fartsgrense:	70
Antall kjørefelt:	2
Antall X-kryss:	1
Antall kurver:	2
Vegklasse:	Riksveg, ikke motorveg og ikke del av TEN-T vegnettet
Fylke:	Hedmark (fylke 4)

Ved å sette inn de respektive verdiene for de enkelte prediktorvariablene og koeffisientene i modellformelen beregnes normale antall ulykker og skadde. Modellformelen er

$$E(n) = e^{\sum_i \text{Prediktor}_i * \text{Koeff}_i}$$

Det normale antall PSU er følgende

PSU

$$= e^{-17,62 + \ln(5000*1) + \ln(10000)*1,230 + \ln(1000)^2 * -0,016 + 0,196 + \ln(1+1)*0,314 + \ln(2+1)*0,005 - 0,024 - 0,184}$$

$$= 2,9186$$

Og det normale antall D/HS er

DHS

$$= e^{-21,05 + \ln(5000*1) + \ln(10000)*1,937 + \ln(1000)^2 * -0,074 + 0,101 + \ln(1+1)*0,176 + \ln(2+1)*-0,043 + 0,220 - 0,095}$$

$$= 0,5263$$

Siden vegen har to kjørefelt, som er referansekategorien for antall kjørefelt, settes det ikke inn noen koeffisient for antall kjørefelt. Antall kryss, rundkjøringer mv. som er null, settes heller ikke inn i formelen. Av antall kurver og antall X-kryss settes den naturlige logaritmen av antallet pluss én inn i formelen. For antall kurver brukes koeffisienten for antall kurver på veger med fartsgrense 70 km/t.

Tabell 3.1.1: Koeffisienter i de endelige modellene for personskadenulykker (PSU), lett skadde (LS), hardt skadde (HS), drepte og drepte eller hardt skadde (D/HS).

	PSU		LS		HS		Drepte		D/HS	
	Koeff.	p	Koeff.	p	Koeff.	p	Koeff.	p	Koeff.	p
Konstanterterm										
Konstanterterm	-17,62	0,000	-17,84	0,000	-21,25	0,000	-22,63	0,000	-21,05	0,000
Ln(segmentlengde * år)										
Ln (lengde * år)	1,000		1,000		1,000		1,000		1,000	
Trafikkmengde										
Ln(ÅDT)	1,230	0,000	1,314	0,000	1,928	0,000	1,925	0,000	1,937	0,000
Ln(ÅDT) ²	-0,016	0,000	-0,018	0,000	-0,074	0,000	-0,070	0,002	-0,074	0,000
Fartsgrense (dummyvariabler)										
30 km/t	0,879	0,000	0,776	0,000	0,026	0,928	0,534	0,294	0,121	0,642
40 km/t	0,678	0,000	0,539	0,000	-0,035	0,855	-0,253	0,521	-0,066	0,710
50 km/t	0,710	0,000	0,652	0,000	0,125	0,218	-0,161	0,415	0,070	0,457
60 km/t	0,298	0,000	0,322	0,000	-0,109	0,214	-0,477	0,004	-0,183	0,023
70 km/t	0,196	0,000	0,260	0,000	0,159	0,102	-0,123	0,468	0,101	0,252
80 km/t	(referanse)		(referanse)		(referanse)		(referanse)		(referanse)	
90 km/t	-0,248	0,000	-0,170	0,018	-0,322	0,072	-0,393	0,187	-0,343	0,032
100 km/t	-0,637	0,000	-0,568	0,000	-0,322	0,072	-0,393	0,187	-0,343	0,032
Antall kjørefelt (dummyvariabler)										
2 felt	(referanse)		(referanse)		(referanse)		(referanse)		(referanse)	
3 felt	0,110	0,076	0,092	0,186	0,051	0,778	0,398	0,137	0,160	0,310
4 felt	0,325	0,000	0,241	0,000	0,320	0,079	-0,632	0,119	0,156	0,352
5 felt	0,507	0,000	0,427	0,001	0,615	0,022	-0,291	0,621	0,456	0,066
6 og flere felt	0,715	0,000	0,687	0,000	0,615	0,022	-0,291	0,621	0,456	0,066
Kryss, rundkjøringer og ramper (Ln(antall + 1))										
T-kryss	0,138	0,000	0,143	0,000	0,153	0,001	-0,058	0,515	0,111	0,010
X-kryss	0,314	0,000	0,274	0,000	0,236	0,066	-0,166	0,567	0,176	0,145
Rundkjøringer	0,447	0,000	0,402	0,000	0,403	0,041	-0,111	0,805	0,307	0,098
Av-ramper	-0,433	0,000	-0,327	0,000	-0,302	0,204	0,110	0,808	-0,214	0,319
På-ramper	-0,216	0,007	-0,203	0,014	-0,160	0,510	-0,472	0,338	-0,229	0,302
Uspes. ramper	0,534	0,000	0,490	0,000	0,244	0,482	-0,296	0,765	0,183	0,573
Kurver (Ln(antall kurver + 1) på segmenter med fartsgrense X)										
Kurver (30 km/t)	-0,643	0,000	-0,781	0,000	-0,206	0,313	-0,639	0,210	-0,255	0,186
Kurver (40 km/t)	-0,291	0,000	-0,282	0,004	-0,206	0,313	-0,639	0,210	-0,255	0,186
Kurver (50 km/t)	-0,192	0,000	-0,186	0,000	-0,187	0,037	-0,248	0,196	-0,192	0,022
Kurver (60 km/t)	0,037	0,158	0,025	0,435	0,135	0,066	0,141	0,340	0,138	0,042
Kurver (70 km/t)	0,005	0,907	-0,052	0,308	-0,127	0,269	0,236	0,198	-0,043	0,677
Kurver (80 km/t)	0,218	0,000	0,205	0,000	0,112	0,020	0,146	0,071	0,120	0,006
Kurver (90 / 100 km/t)	0,183	0,465	0,156	0,582	0,382	0,501	0,382		0,034	0,952
Stigninger (Ln(antall stigninger + 1) på segmenter med fartsgrense X)										
Stig. (30/40 km/t)	-0,241	0,084	-0,232	0,191	-0,064	0,319	-0,334	0,006	-0,117	0,048
Stigninger (50 km/t)	-0,191	0,001	-0,232	0,002	-0,064	0,319	-0,334	0,006	-0,117	0,048
Stigninger (60 km/t)	0,037	0,458	0,024	0,696	-0,064	0,319	-0,334	0,006	-0,117	0,048
Stigninger (70 km/t)	0,221	0,006	0,309	0,001	-0,064	0,319	-0,334	0,006	-0,117	0,048
Stigninger (80 km/t)	-0,005	0,888	-0,047	0,245	-0,064	0,319	-0,334	0,006	-0,117	0,048
Stigninger (90 / 100 km/t)	0,496	0,033	0,435	0,076	-0,064	0,319	-0,334	0,006	-0,117	0,048

Tabell 3.1.1 (forts.)

	PSU		LS		HS		Drepte		D/HS	
	Koeff.	p	Koeff.	p	Koeff.	p	Koeff.	p	Koeff.	p
Vegklasser										
1 Motorveg	-0,705	0,000	-0,658	0,000	-0,929	0,003	-0,910	0,199	-0,910	0,002
2 Tofelts veg med planskilte kryss	-0,499	0,000	-0,336	0,000	-0,280	0,145	0,307	0,276	-0,123	0,452
3 TEN-T (øvrige)	-0,101	0,000	-0,056	0,075	0,215	0,002	0,391	0,001	0,259	0,000
4 Øvrig Ev/Rv	-0,024	0,307	-0,010	0,716	0,164	0,010	0,397	0,000	0,220	0,000
5 Fylkesveg	(referanse)		(referanse)		(referanse)		(referanse)		(referanse)	
Midtdeler, rekkverk og forsterket midtoppmerking										
Midtdeler & rekkv.	-0,140	0,299	-0,355	0,010	-1,396	0,023	-1,396		-1,635	0,007
Midtdeler u/ rekkv.	0,100	0,026	0,035	0,479	0,100	0,465	0,280	0,238	0,131	0,286
Midtrekkv. u/ midtdeler	-0,096	0,755	-0,127	0,671	-0,127		-0,127		-0,127	
Ingen midtdeler / - rekkverk	(referanse)		(referanse)		(referanse)		(referanse)		(referanse)	
Forst. midtoppm. > 1m	-0,521	0,001	-0,352	0,016	-0,329	0,371	-0,272	0,582	-0,269	0,378
Forst. midtoppm. < 1m	-0,348	0,039	-0,555	0,001	-0,062	0,866	-0,388	0,546	-0,162	0,626
Fylke (fylkesnr. se fotnote)										
Fylke 1	0,112	0,003	0,102	0,031	-0,227	0,049	-0,127	0,537	-0,206	0,050
Fylke 2	-0,151	0,000	-0,256	0,000	-0,232	0,024	-0,290	0,137	-0,238	0,012
Fylke 3	-0,006	0,933	-0,120	0,112	-0,326	0,157	-0,534	0,289	-0,373	0,079
Fylke 4	-0,184	0,000	-0,237	0,000	-0,150	0,154	0,058	0,744	-0,095	0,321
Fylke 5	-0,256	0,000	-0,265	0,000	0,054	0,599	-0,111	0,561	0,024	0,804
Fylke 6	-0,491	0,000	-0,512	0,000	-0,154	0,186	0,089	0,654	-0,097	0,357
Fylke 7	0,002	0,960	-0,009	0,862	0,020	0,865	-0,450	0,073	-0,055	0,615
Fylke 8	0,159	0,000	0,204	0,000	-0,108	0,348	-0,461	0,042	-0,170	0,115
Fylke 9	0,021	0,657	0,050	0,376	-0,142	0,283	-0,262	0,287	-0,165	0,176
Fylke 10	-0,173	0,000	-0,278	0,000	-0,273	0,030	0,344	0,075	-0,112	0,311
Fylke 11	-0,243	0,000	-0,290	0,000	-0,285	0,005	-0,151	0,411	-0,256	0,006
Fylke 12	(referanse)		(referanse)		(referanse)		(referanse)		(referanse)	
Fylke 14	-0,285	0,000	-0,329	0,000	-0,357	0,008	-0,374	0,120	-0,361	0,004
Fylke 15	-0,050	0,189	-0,022	0,638	-0,131	0,215	0,024	0,895	-0,096	0,320
Fylke 16	0,002	0,962	-0,005	0,916	0,088	0,392	-0,386	0,065	0,004	0,970
Fylke 17	-0,540	0,000	-0,632	0,000	-0,300	0,016	0,030	0,884	-0,224	0,045
Fylke 18	-0,173	0,000	-0,177	0,000	-0,392	0,000	-0,158	0,394	-0,340	0,001
Fylke 19	-0,237	0,000	-0,232	0,000	-0,344	0,006	-0,026	0,899	-0,266	0,019
Fylke 20	-0,155	0,009	-0,115	0,106	-0,119	0,419	0,002	0,993	-0,089	0,511
Modell for beregning av overspredningsparameteren										
Ln(lengde * år)	-0,778	0,000	-0,742	0,000	-0,710	0,000	-1,045	0,000	-0,787	0,000
Ln(ÅDT)	-0,371	0,000	-0,752	0,000	-0,652	0,000	-0,754	0,000	-0,730	0,000
Kosstantterm	8,529	0,000	12,676	0,000	12,457	0,000	16,608	0,000	13,701	0,000

Fylke: 1 - Østfold; 2 - Akershus; 3 - Oslo; 4 - Hedmark; 5 - Oppland; 6 - Buskerud; 7 - Vestfold; 8 - Telemark; 9 - Aust-Agder; 10 - Vest-Agder; 11 - Rogaland; 12 - Hordaland; 14 - Sogn og Fjordane; 15 - Møre og Romsdal; 16 - Sør-Trøndelag; 17 - Nord-Trøndelag; 18 - Nordland; 19 - Troms; 20 - Finnmark

3.2 Hvor gode er modellene?

Tabell 3.2.1 viser ulike goodness-of-fit (GOF) statistikker som beskriver hvor gode modellene predikerer registrerte ulykkestall:

- **Pseudo- R^2 :** Denne statistikken er beregnet av statistikkprogrammet, er basert på loglikelihood og indikerer hvor "sannsynlig" et sett med modellparametre er, gitt de registrerte ulykkestallene. Jo større Pseudo- R^2 , desto bedre er modellen. Denne statistikken anvendes av statistikkprogrammet (Stata) som kriterium i modellberegningen, dvs. at koeffisientene settes slik at R^2 blir maksimert.
- **MSPE (mean square prediction error):** Dette er det gjennomsnittlige kvadrerte avviket mellom registrerte og predikerte antall ulykker / skadde / drepte. Jo mindre MSPE, desto bedre er modellen.
- **Elvik-indeks:** Elvik indeks er en funksjon av overspredningen i de predikerte og registrerte ulykkestallene og indikerer hvor stor andel av variansen i de registrerte ulykkestallene som er forklart av modellen.

Når man ser på Pseudo- R^2 og MSPE, er den endelige modellen for PSU den beste, mens modellene for HS og D er forholdsvis svake.

I tillegg er det lagd **Cure-plots** for alle skadegrader. Cure står for cumulative residuals. En CURE-plot viser hvor godt modeller predikerer registrerte ulykkestall ved ulike verdier av en eller flere prediktorvariabler (Caliendo et al., 2007), i dette tilfelle trafikkmengde. For å lage en Cure plot sorteres dataene etter den aktuelle prediktorvariabelen (her trafikkmengde) og man beregner kumulative residualer, dvs. at man for hver datapunkt beregner summen av alle avvik mellom registrerte og predikerte ulykkestall for alle datapunktene fra den minste trafikkmengden til trafikkmengden på det aktuelle datapunktet. Cure-plots for alle skadegradene er vist i figur 3.2.1. Når de kumulative residualene (de brede linjene) ligger innenfor \pm to standardavvik (de smale linjene) anser man som regel modellene som «gode» dvs. at de predikerte tallene ikke viser store systematiske avvik. De kumulative residualene begynner alltid på null (ved den minste trafikkmengden) og dersom det totale antall predikerte er lik det totale antall registrerte ulykker ender linjen som viser de kumulative residualene også på null (ved den høyeste trafikkmengden). Dette er stort sett tilfelle, de brede linjene ligger for det meste godt innenfor området mellom de smale linjene. Kun for PSU og LS ved høy ÅDT er de kumulative residualene forholdsvis høye og for det meste større enn to standardavvik, især for LS. Dette viser at det predikerte antall PSU og LS i gjennomsnitt er høyere enn det registrerte ved høy ÅDT.

Tabell 3.2.1 viser i tillegg til GOF indikatorene deskriptiv statistikk for registrerte og normale ulykkestall samt avvik mellom normale og registrerte ulykkestall for hele datamaterialet (alle år). For antall normale og registrerte ulykkestall per år vises summene og avvikene mellom normale og registrerte ulykkestall. Det prosentvise avviket i de årlige ulykkestallene er større enn i de totale ulykkestallene. Dette skyldes at normale ulykkestall for segmenter som inngår med flere år i datafilen er mer nøyaktige enn normale ulykkestall for segmenter med færre år med data. Det prosentvise avviket for de årlige ulykkestallene er i de fleste tilfellene positivt, dvs. at de normale ulykkestallene er noe høyere enn de registrerte (unntatt i enkelte trafikkmengdegrupper for HS, D og D/HS).

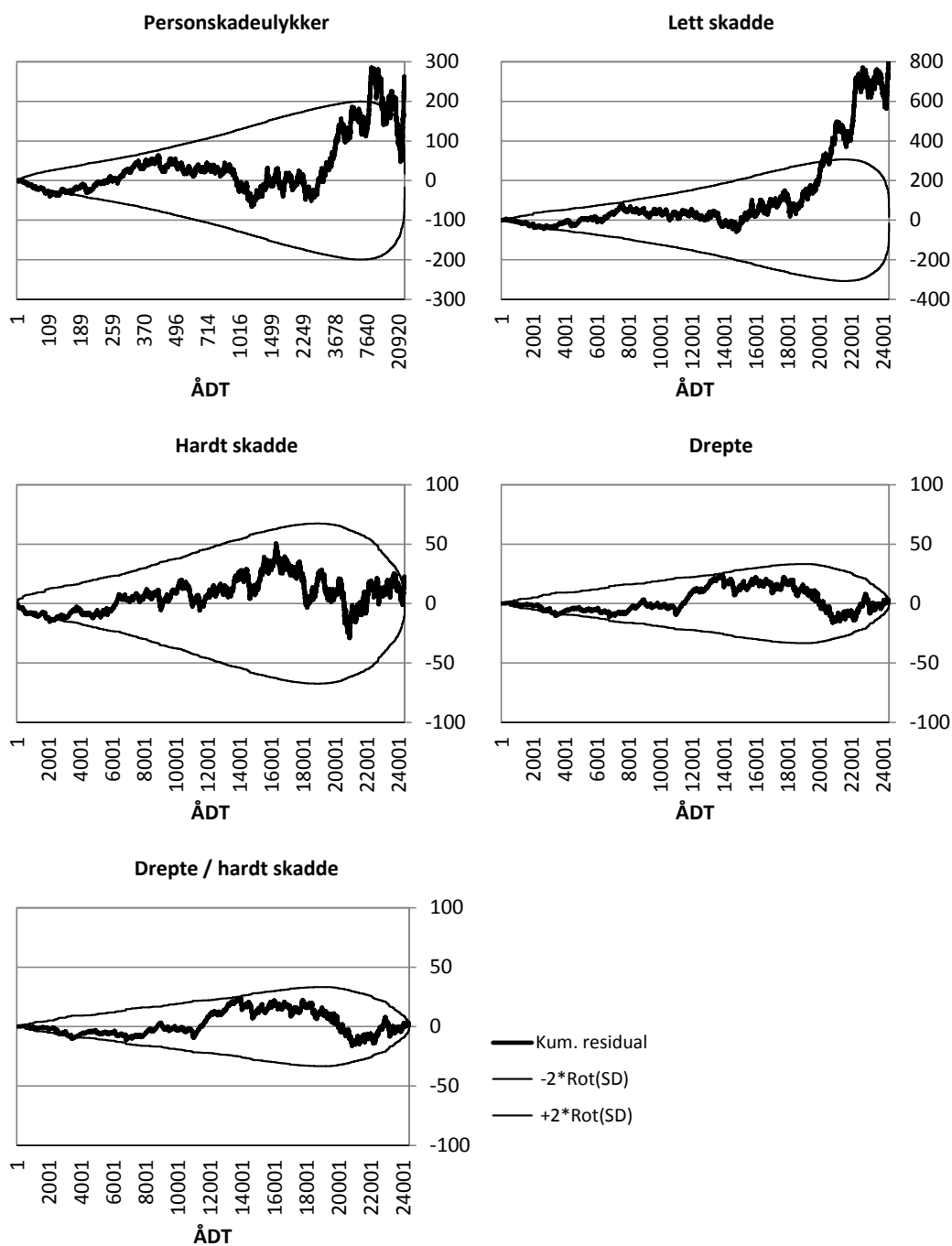
Avvikene (både for totale og årlige ulykkestall) er større i modellene for D og HS enn i modellen for D/HS. I analyser som gjelder både D og HS kan det derfor være en fordel å bruke modellen for D/HS istedenfor modellene for de enkelte skadegradene.

Den nederste delen av tabellen viser at avvikene for årlige ulykkestall varierer mellom segmenter med ulik ÅDT. Forklaringen er at det er færre ulykker, skadde eller drepte i enkelte ÅDT-grupper enn i hele datamaterialet og dermed større usikkerhet (og større avvik).

Tabell 3.2.1: GOF for de endelige modellene og avvik mellom normale og registrerte ulykkestall.

	km veg	PSU	LS	HS	D	D/HS
Pseudo R²		0,200	0,144	0,075	0,074	0,073
MSPE¹		0,544	1,286	0,062	0,015	0,086
Elvik-indeks		0,600	0,605	2,895	4,559	2,393
Totalt ulykkestall (alle år): Deskriptiv statistikk for registrerte og normale ulykkestall						
Registrerte (sum)	51 358 km	25 211	32 915	3 257	898	4 155
Registrerte (maks.)	51 358 km	33	48	7	5	10
Registrerte (varians)	51 358 km	0,970	2,138	0,065	0,016	0,092
Registrerte (overspred.)	51 358 km	5,267	8,342	10,561	22,048	11,020
Normale (sum)	51 358 km	25 464	33 722	3 278	901	4 162
Normale (maks.)	51 358 km	32,52	42,27	1,55	0,47	1,33
Normale (varians)	51 358 km	0,603	1,160	0,005	0,000	0,007
Normale (overspred.)	51 358 km	2,105	3,293	-20,013	-78,471	-15,349
Totalt ulykkestall (alle år): Avvik normale minus registrerte ulykkestall						
Avvik (antall)	51 358 km	252,8	807,4	21,5	2,5	7,2
Avvik (%)	51 358 km	1,00%	2,45%	0,66%	0,28%	0,17%
Per år: Antall normale og registrerte ulykkestall						
Registrerte (sum)	51 358 km	4 586	6 008	581	158	739
Predikerte (sum)	51 358 km	4 701	6 236	597	164	757
Per år: Avvik normale minus registrerte ulykkestall						
Avvik (antall)	51 358 km	115,2	228,4	16,0	6,5	18,5
Avvik (%)	51 358 km	2,51%	3,80%	2,75%	4,11%	2,50%
Prosentvise avvik mellom normale og registrerte antall per år på veger med ulik ÅDT						
ÅDT ≤ 1000	34 363 km	0,28 %	0,40 %	4,95 %	14,08 %	6,97 %
1,000 < ÅDT ≤ 1,500	4 461 km	1,77 %	3,36 %	4,16 %	0,24 %	3,41 %
1,500 < ÅDT ≤ 2,000	2 929 km	1,33 %	3,86 %	1,73 %	-5,53 %	0,02 %
2,000 < ÅDT ≤ 5,000	6 041 km	3,86 %	5,65 %	-2,89 %	-6,36 %	-3,86 %
5,000 < ÅDT ≤ 10,000	2 095 km	3,70 %	5,21 %	7,58 %	15,46 %	8,34 %
10,000 ≤ ÅDT	1 469 km	2,23 %	3,09 %	5,03 %	10,96 %	3,67 %

¹Ikke sammenlignbar mellom modeller for ulike skadegrader



Figur 3.2.1: Cure plots for alle skadegrader.

Testberegninger som er beskrevet i Høye (2013) for å finne de beste definisjonene for enkelte prediktorvariablene viste at modellenes GOF nesten ikke påvirkes av hvordan man definerer enkelte prediktorer eller om enkelte prediktorer er med eller ikke. Når man sammenligner ulike modeller (med ulike prediktorer) viser GOF indikatorene kun svært små forskjeller mellom ulike modellvarianter. De eneste modellene som skiller seg ut er modellene med kun $\text{Ln}(\text{ÅDT})$ som prediktor; disse er klart dårligere enn modellene med øvrige prediktorer. Dette betyr i praksis at modellene ikke vil endre seg i stor grad hvis man endrer eller utelater enkelte prediktorer (for normale ulykestall på enkelte segmenter eller vegstrekninger kan det likevel ha betydning hvilke variabler som er med i modellen).

Det er to problemer med modellresultatene som det ikke er funnet noen løsning på:

- Alle modellene beregner i sum for høye normale ulykkestall, især modellen for LS (2,5% flere enn registrert når man ser på data for alle år). Avvikene er størst på veger med høy trafikkmengde. Mulige forklaringer er: A) At det finnes relevante prediktorvariabler som ikke er inkludert i modellen (for eksempel forholdet mellom kapasitet og trafikkmengde); B) At antall LS i større grad er underrapportert på veger med høy trafikkmengde enn på andre veger og C) at segmenter med høy ÅDT i gjennomsnitt er representert med data fra færre år slik at usikkerheten (og dermed avvikene) her er større enn på segmenter med lavere ÅDT.
- Variansen i de predikerte antallene HS, D og D/HS er mindre enn gjennomsnittet, noe som pleier å være omvendt. Av denne grunnen er Elvik-indeks større enn én for D, HS og D/HS. Forklaringen er trolig at det finnes svært mange segmenter med null D, HS og D/HS.

For øvrig ser modellene generelt logiske ut i den forstand at de aller fleste koeffisientene har omtrent forventet størrelse og fortegn. Det er kun enkelte koeffisienter i modellene for HS og D som er justert manuelt fordi koeffisientene var absurd store, noe som trolig skyldes små antall HS og D i segmenter med de respektive variablene.

4 Trendfaktor

Normale ulykkestall som er beregnet med ulykkesmodellene kan tolkes som resultater for 2008. For å omregne resultatene til andre år må man justere to tall: Det normale ulykkestall og overspredningsparameteren. For å trendjustere normale ulykkestall er det utviklet en trendfaktor som er basert på informasjon om transportarbeid i Norge i 1997 til 2011 med alle typer motorkjøretøy (Vågane, 2012) og antall PSU, LS, HS, D og D/HS i 1997 til 2011 (SSB). Trendjusteringen av overspredningsparameteren gjøres ut fra sammenhengen mellom normale ulykkestall og overspredningsparameteren. Trendjusteringen av normale ulykkestall er beskrevet i avsnitt 4.1, trendjusteringen av overspredningsparameteren i avsnitt 4.2.

Når man trendjusterer normale ulykkestall som er beregnet med ulykkesmodellen må man i modellberegningen oppgi trafikkmengden for det året det skal trendjusteres til. Når man kun har trafikkmengden for et annet år kan man beregne trafikkmengden for det året som skal trendjusteres til som beskrevet i avsnitt 4.3.

4.1 Trendjustering av predikerte (normale) ulykkestall

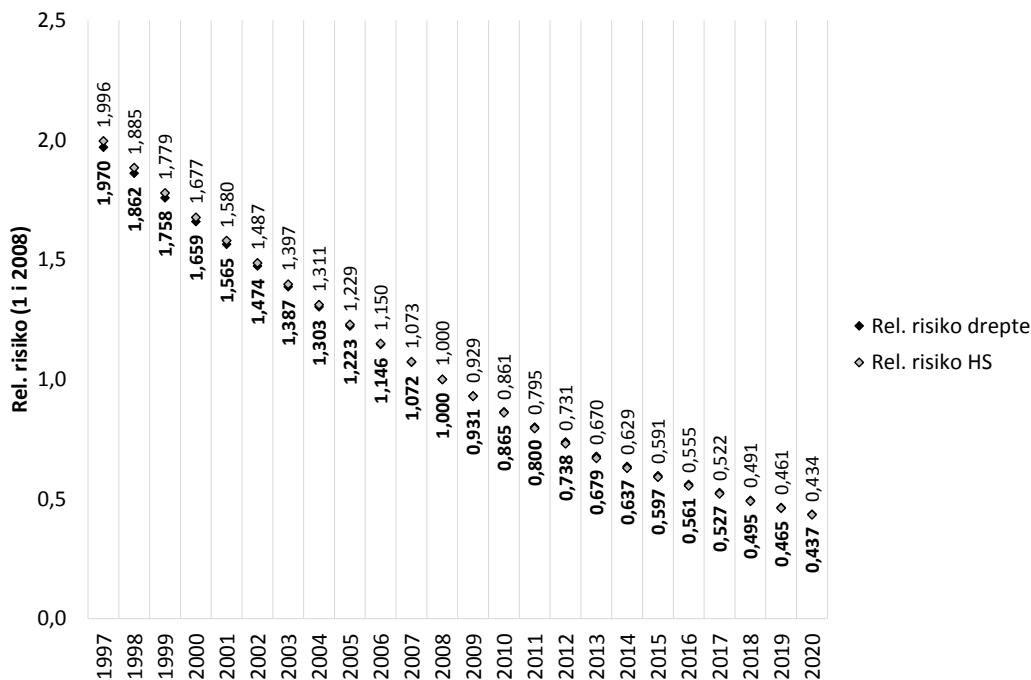
For å trendjustere normale ulykkestall til et år X ganges disse med det relative risikotall for år X som er vist i figur 4.1.1 for D og HS, i figur 4.1.2 for D/HS og i figur 4.1.3 for LS og PSU. De relative risikotallene for hvert år i perioden 1997-2020 viser risikoen i år X i forhold til risikoen i år 2008. Risikoen for hvert år er beregnet som antall skadde / drepte som er beregnet med en trendfunksjon delt på den totale trafikkmengden. Dette betyr at de relative risikotallene viser hvor mange ulykker det hadde vært i et år X , i forhold til 2008, hvis alt annet inkludert trafikkmengden hadde vært likt. Når man trendjusterer normale ulykkestall forutsettes dermed at man har oppgitt trafikkmengden for det året man trendjusterer til.

Risikotallene for 1997-2020 er beregnet med to forskjellige trendfunksjoner for antall skadde / drepte:

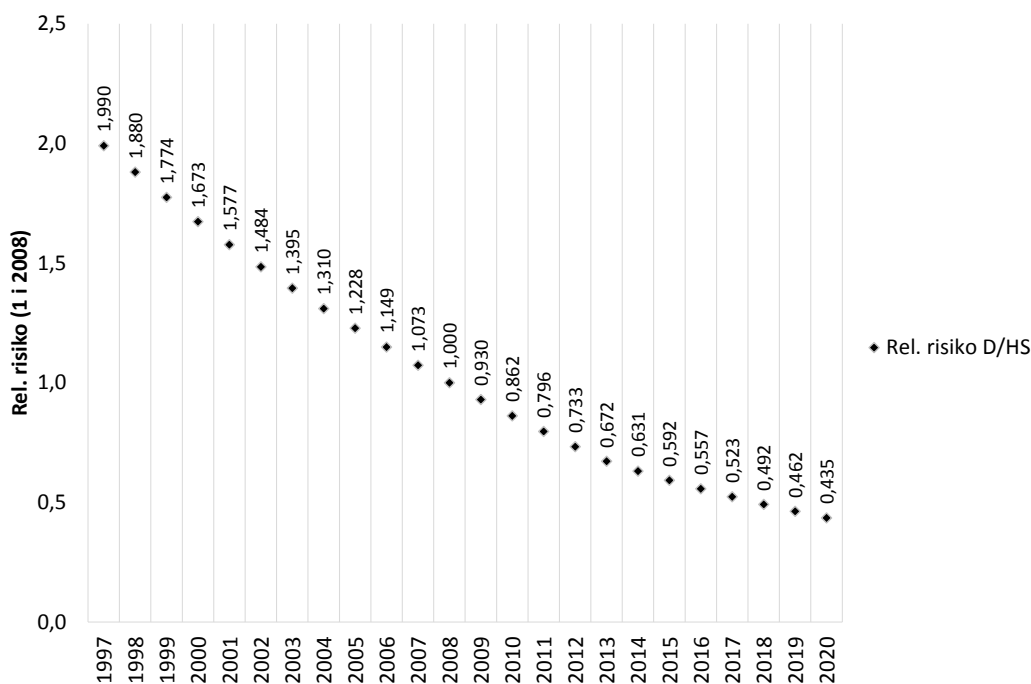
- A. For årene 1997-2012 brukes for hver skadegrad den trendfunksjonen som beskriver de faktiske antallene skadde / drepte i årene 1997-2011 best (lineær trend for D og HS, polynom trend for LS og PSU).
- B. For årene 2013-2020 brukes for alle skadegradene en eksponentiell trendfunksjon (som er basert på faktiske risikotall i 1997-2011); dermed unngår man å predikere en urealistisk stor nedgang av ulykkestallene i fremtidige år (med den første trendfunksjonen ville den relative risikoen krysse null-linjen allerede få år etter 2020).

For trafikkmengden er det beregnet en lineær trendfunksjon basert på transportytelsene ifølge Vågane (2012). Risikotallene for D, HS og D/HS følger ikke en lineær trend, selv om de er beregnet med en lineær trend for antall drepte, HS og D/HS).

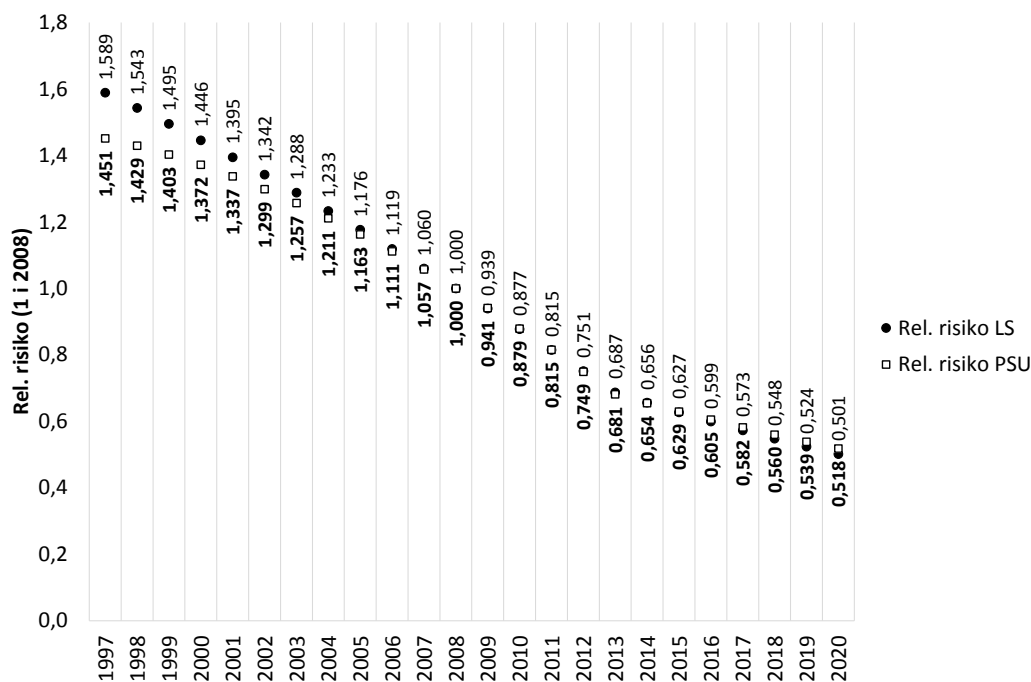
Referanseåret for den relative risikoen er 2008, dvs. at den relative risikoen for hver skadegrad er satt lik 1 i 2008. 2008 er det «gjennomsnittlige» året i datafilen som er lagt til grunn for beregningen av ulykkesmodellene. Når man vekter antall år som inngår i dataene for hvert segment med det årlige antall kjøretøykilometer er det gjennomsnittlige antall år for alle segmenter 5,227, dvs. at et gjennomsnittlig segment med gjennomsnittlig trafikkmengde har data fra 2006 til mars 2011. Midten av denne perioden er midten av 2008, og 2008 er derfor valgt som referanseår.



Figur 4.1.1: Relativ risiko for antall drepte og HS i 1997-2020 med trend A 1997-2012 og med trend B i 2013-2020 (relativ risiko i 2008 = 1; relativ risiko for drepte i fet skrift nederst).



Figur 4.1.2: Relativ risiko for antall D/HS i 1997-2020 med trend A 1997-2012 og med trend B i 2013-2020 (relativ risiko i 2008 = 1).



Figur 4.1.3: Relativ risiko for antall LS og PSU i 1997-2020 med trend A 1997-2012 og med trend B i 2013-2020 (relativ risiko i 2008 = 1; relativ risiko for PSU i fet skrift nederst).

Den store nedgangen av den relative risikoen i løpet av årene 1997 til 2011 (og videre til 2020) kan forklares med at det har vært en stor økning av trafikkarbeidet, samtidig som antall ulykker, skadde og drepte har gått ned:

- Trafikkarbeidet med alle typer motorkjøretøy har økt fra ca. 33,000 mill. kjøretøykilometer i 1997 til ca. 44,000 mill. kjøretøykilometer i 2011 (+33%)
- Antall PSU har gått ned fra 8,765 i 1997 til 6,154 i 2012 (-30%)
- Antall drepte har gått ned fra 303 i 1997 til 145 i 2012 (-52%)
- Antall HS har gått ned fra 1,341 i 1997 til 699 i 2012 (-48%)

4.2 Trendjustering av overspredningsparameteren

Når man trendjusterer normale ulykkestall må man, i en EB-analyse, i tillegg til de normale ulykkestallene også justere overspredningsparameteren. Den trendjusterte overspredningsparameteren beregnes som en funksjon av det normale ulykkestall for det året man har data for og det året det skal trendjusteres til, samt to koeffisienter, etter følgende formel:

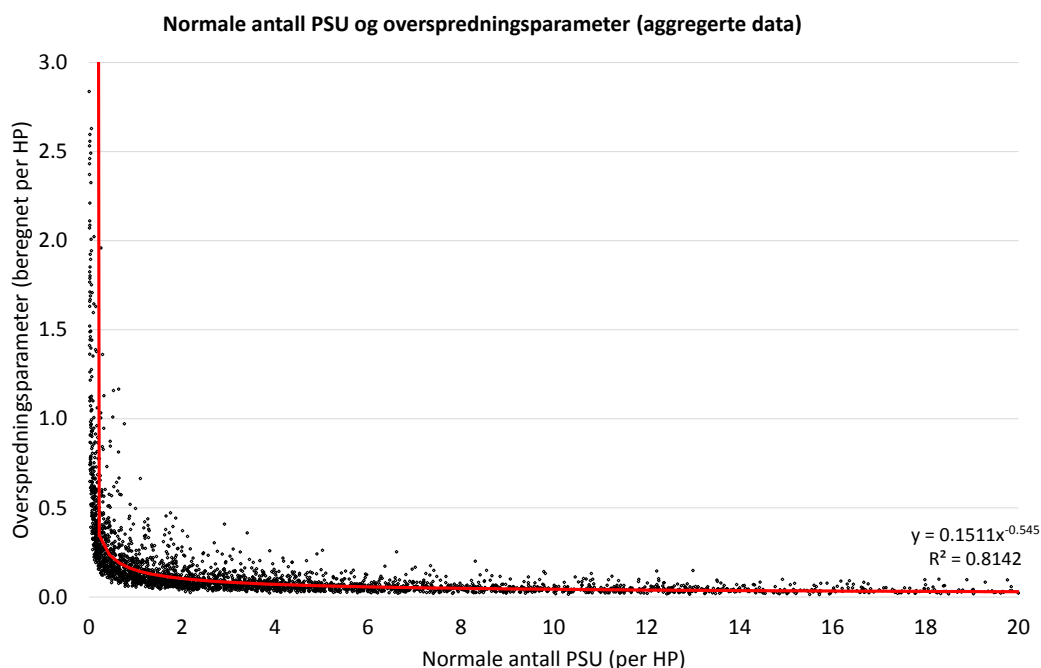
$$Faktor = \frac{a * x_1^b}{a * x_2^b}$$

Hvor x_1 er det normale ulykkestallet i året det skal justeres til og x_2 er det normale ulykkestallet som er beregnet med modellen (og som gjelder året 2008), a og b er koeffisienter. Hvilke koeffisienter som skal brukes avhenger av skadegraden og er vist i tabell 4.2.1. I utgangspunktet anbefales koeffisientene som står under «Anbefalte koeffisienter». De alternative koeffisientene kan ev. brukes for korte segmenter (under 1,2 km, se forklaring nedenfor).

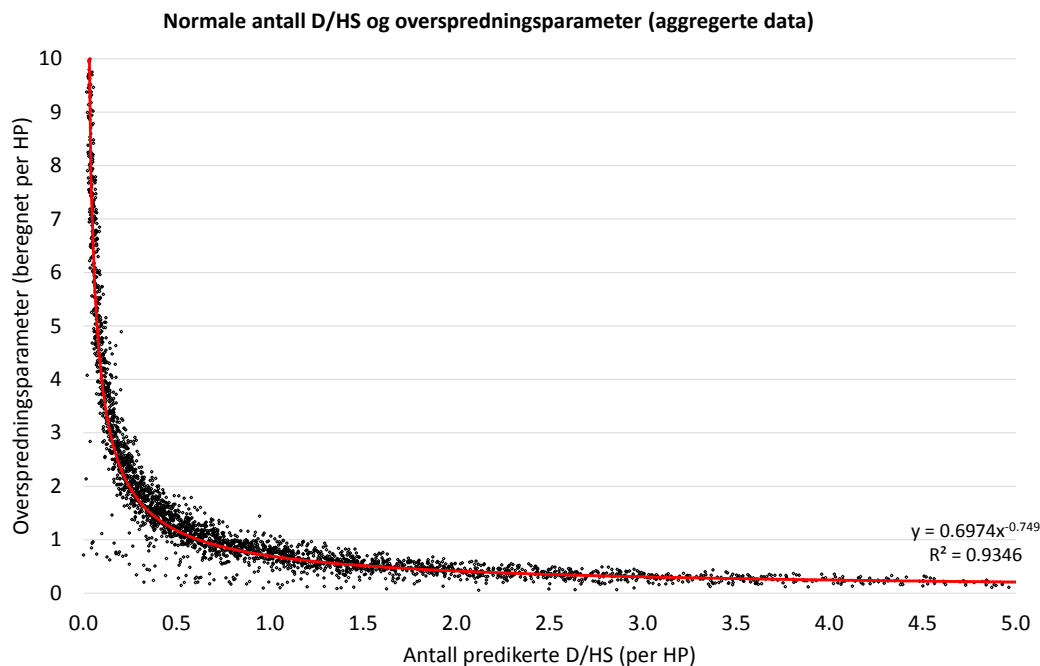
Tabell 4.2.1: Koeffisienter for beregning av overspredningsparameteren som en funksjon av normale ulykkes- og skadetall med en formel av typen $y = a * x^b$ hvor y er overspredningsparameteren, x er normalt ulykkestall (ikke trendjustert) og a og b er koeffisientene.

Koeffisient	PSU	LS	HS	D	D/HS
Anbefalte koeffisienter					
a	0,1511	1,3875	0,7055	0,1572	0,6974
b	-0,545	-0,717	-0,688	-0,861	-0,749
Alternative koeffisienter (for korte segmenter, < 1,2 km)					
a	0,3292	1,61	0,8736	0,3537	1,0126
b	-0,492	-0,688	-0,658	-0,814	-0,688

Trendjusteringen med de anbefalte koeffisientene er basert på sammenhengen mellom predikerte ulykkestall og overspredningsparameter i en aggregert datafil hvor én datapunkt er en hel hovedparsell på én veg i ett fylke. Sammenhengen for PSU er vist i figur 4.2.1 og sammenhengen for antall D/HS er vist i figur 4.2.2. De røde linjene viser de trendfunksjonene (potensfunksjon) som ligger til grunn for koeffisientene i tabell 4.2.1.



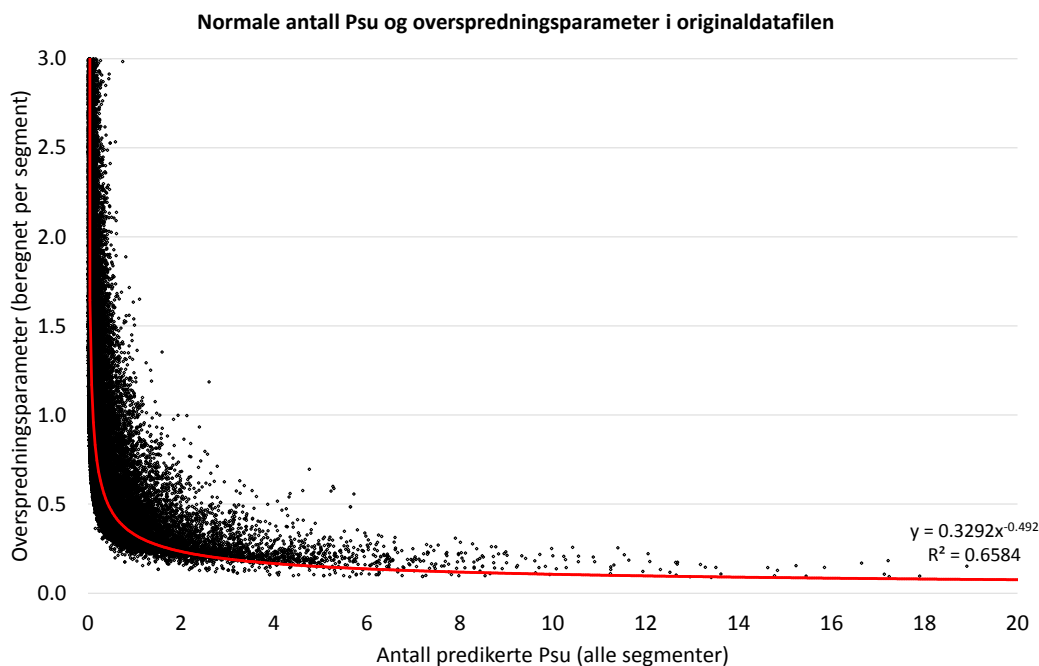
Figur 4.2.1: Sammenhengen mellom normale antall PSU og overspredningsparameter i aggregert datafil (segmentlengde i gjennomsnitt 15,5 km, maks. 185 km).



Figur 4.2.2: Sammenhengen mellom normale antall D/HS og overspredningsparameter i aggregert datafil (segmentlengde i gjennomsnitt 15,5 km, maks. 185 km).

En potensfunksjon av formen $y = a * x^b$ (y er overspredningsparameteren og x er normalt ulykkestall) er funksjonen som beskriver sammenhengen best. Figurene viser likevel at det er forholdsvis stor spredning rundt trendlinjene.

Beregner man slike funksjoner basert på originaldatafilen hvor alle segmentene er 1,2 km eller kortere, får man lignende trendfunksjoner. Disse har imidlertid mye større spredning og for det samme normale antall skadde / drepte er overspredningsparameteren i gjennomsnitt noe høyere (figur 4.2.3). Derfor anbefales i utgangspunktet å bruke koeffisientene som er beregnet med den aggregerte datafilen. Disse gjelder for forholdsvis lange segmenter; i den aggregerte datafilen er segmentene i gjennomsnitt 15,5 km lange (opptil 185 km). For segmenter under 1,2 km lengde kan man likevel bruke resultatene fra originaldatafilen. Dermed vil man unngå å underestimere overspredningsparameteren.



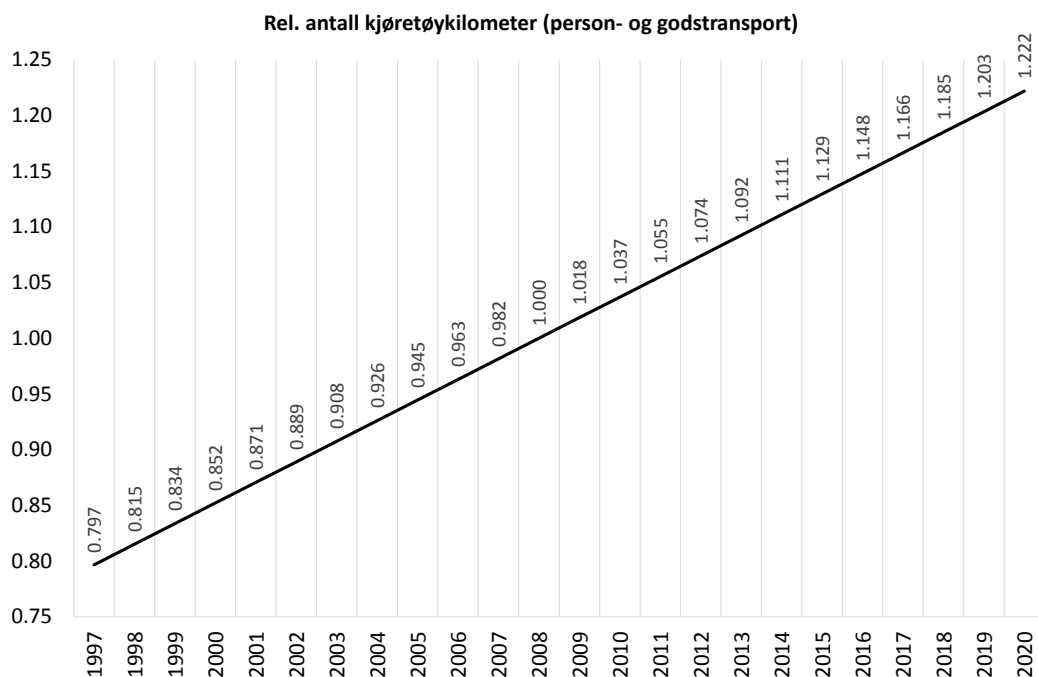
Figur 4.2.3: Sammenhengen mellom normale antall PSU og overspredningsparameter i original datafil (segmentlengde i gjennomsnitt 0,7 km, maks. 1,2 km).

Sammenhengen mellom normalt antall ulykker og overspredningsparameteren avhenger av hvilket datagrunnlag som brukes for å beregne sammenhengen. Dermed vil sammenhengen strengt tatt ikke være overførbart til andre datagrunnlag og trendjusteringen av overspredningsparameteren vil alltid være noe unøyaktig. Det finnes imidlertid ingen mulighet for å trendjustere overspredningsparameteren på en standardisert måte som ikke medfører noe unøyaktighet. Beregner man den trendjusterte overspredningsparameteren basert på originaldatafilen er unøyaktigheten større enn når den beregnes basert på den aggregerte datafilen, især for D, HS og D/HS, men man unngår å underestimere overspredningsparameteren. Totalt sett har det imidlertid forholdsvis liten betydning for den trendjusterte overspredningsparameteren og dermed for EB-vektene hvilken trendjustering som velges for overspredningsparameteren.

4.3 Trendjustering av trafikkmengden

Trendfaktoren som er beskrevet i forrige avsnitt tar ikke hensyn til endringer av trafikkmengden i den forstand at de relative risikotallene er beregnet som antall ulykker i forhold til den faktiske trafikkmengden i de respektive årene. Dersom man justerer et antall ulykker til et annet år med hjelp av trendfaktoren forutsettes derfor implisitt at trafikkmengden i det året man justerer til er nøyaktig det samme som i det året hvor ulykkestallet ble beregnet. Hvis man vil ta hensyn til at også trafikkmengden endrer seg over tid, kan man beregne ulykkestallet med den antatte trafikkmengden i det året det skal justeres til. Figur 4.3.1 viser den relative totale trafikkmengden (antall kjøretøykilometer med både person- og godstransport) i 1997 til 2020. Den relative trafikkmengden i 2008 er satt lik én. Dette er den samme relative trafikkmengden som ble lagt til grunn i beregningen av de relative risikotallene for trendfaktoren.

Hvis man for eksempel har trafikkmengden for året 2008 og vil justere denne til året 2014, ganger man trafikkmengden med 1,111. Hvis man har trafikkmengden for 2014 og vil justere denne til året 2018, ganger man trafikkmengden først med 1,185 (den relative trafikkmengden i 2018) og deler resultatet på 1,111 (den relative trafikkmengden i 2014).



Figur 4.3.1: Relativ trafikkmengde (antall kjøretøykilometer, person- og godstransport; relativt antall i 2008 = 1).

5 Brukerveiledning for Ulykkesmodeller.xlsx

I vedlagte xls-regneark Ulykkesmodeller.xlsx (11. april 2014) kan man beregne:

- **Normale antall** PSU, LS, HS, D og D/HS som en funksjon av prediktorvariablene som er beskrevet ovenfor
- **Overspredningsparameteren** for hver skadegrad
- **EB-vekten** for hver skadegrad som kan brukes til å beregne forventede ulykkes/-skadetall
- **Forventede antall** PSU, LS, HS, D og D/HS (hvis man i tillegg oppgir registrerte ulykkes- og skadetall)

Beregningene kan gjøres for én vegstrekning om gangen. Vegstrekningen har ideelt sett konstante vegegenskaper på hele strekningen (for eksempel samme fartsgrense, trafikkmengde mv.). Fremgangsmåten for heterogene vegstrekninger er beskrevet nedenfor. Alle tallene beregnes både uten trendjustering og med trendjustering til et valgfritt år mellom 1997-2020. Resultater uten trendjustering gjelder året 2008.

Brukeren må oppgi følgende informasjon i de gule cellene i regnearket:

- **Alle vegegenskapene som er prediktorer i ulykkesmodellene:**
 - **Trafikkmengde (ÅDT):** Må alltid oppgis; skal man bruke trendjustering oppgis trafikkmengden for det året man skal justere til
 - **Fartsgrense (km/t):** Må alltid oppgis
 - **Antall kjørefelt:** Må alltid oppgis
 - **Antall T-kryss:** Dersom antall T-kryss er ukjent kan man oppgi null T-kryss eller et estimert antall
 - **Antall X-kryss:** Må alltid oppgis
 - **Antall rundkjøringer:** Må alltid oppgis
 - **Antall av-ramper:** Må alltid oppgis
 - **Antall på-ramper:** Må alltid oppgis
 - **Antall uspesifiserte ramper:** Skal ikke oppgis
 - **Antall kurver:** Dersom antall kurver er ukjent kan man oppgi null kurver eller et estimert antall; på veger med fartsgrense 60 eller 70 km/t har antall kurver kun svært lite betydning for det normale antall ulykker
 - **Antall stigninger:** Dersom antall stigninger er ukjent kan man oppgi null stigninger eller et estimert antall; på veger med fartsgrense 60 eller 80 km/t har antall stigninger kun svært lite betydning for det normale antall ulykker
 - **Vegkategori (velges fra rullegardinmeny):** Må alltid oppgis

- **Midtdeler.** Må alltid oppgis, unntatt for motorveger¹
- **Midtrekkverk.** Må alltid oppgis, unntatt for motorveger²
- **Forsterket midtoppmerking.** Må alltid oppgis
- **Fylke (velges fra rullegardinmeny):** Må alltid oppgis; fylket har stor betydning for det normale antall ulykker
- **Lengde (meter):** Dette er strekningslengden i meter
- **Antall år:** Dette er antall år som man vil beregne normale ulykkestall for (obs: trendjustering kan kun gjøres for ett år om gangen, se nedenfor)
- **År for trendjustering:** Dette er året det skal trendjusteres til; år mellom 1997 og 2020 kan oppgis (eller utelates)
- **Justering av overspredningsparameteren:** Man kan velge om overspredningsparameteren skal trendjusteres for lange eller korte segmenter; det gis følgende anbefaling:
 - Generelt anbefales trendjustering for lange segmenter; denne er, især for D, HS og D/HS mer nøyaktig enn for korte segmenter
 - For korte strekninger som omfatter kun en liten andel av en hovedparsell kan trendjusteringen for korte segmenter brukes

Noen praktiske problemstillinger er beskrevet i følgende.

Ukjente antall kurver, stigninger eller T-kryss: Dersom en av disse variablene er ukjent har man følgende muligheter:

- a) Antallet oppgis som null. Totalt sett (hele vegnettet eller større deler av vegnettet sett under ett) vil avvikene mellom normale og registrerte ulykkestall ikke øke i vesentlig grad dersom man oppgir null kurver, stigninger eller T-kryss for alle segmentene.
- b) Hvis man vet at vegen har kurver, stigninger eller T-kryss, men ikke hvor mange, oppgis et estimert antall kurver, stigninger og T-kryss.
- c) Man kan også oppgi det gjennomsnittlige antall i hele vegnettet. Det gjennomsnittlige antallet per kilometer veg er 2,037 for kurver, 0,445 for stigninger (gjennomsnittlige antall kurver / stigninger ved ulike fartsgrenser er vist i avsnitt 2.8 og 2.9), samt 0,040 for X-kryss, 0,650 for T-kryss, 0,022 for rundkjøringer og 0,013 for av- og påramper.

¹ Virkningen av midtdeler og midtrekkverk på motorveg er fanget opp av vegklassevariabelen (vegklasse 1 motorveg); resultatene i regnearket er derfor uavhengige av hva som er oppgitt for midtdeler og midtrekkverk når vegtypen er motorveg.

² Se fotnote for midtdeler.

Heterogene strekninger (strekninger med varierende vegegenskaper): Ideelt sett har hele strekningen de samme vegegenskapene. For å beregne normale ulykkestall mv. for strekninger med varierende egenskaper kan man dele strekningen opp i homogene segmenter med konstante egenskaper og summere resultatene. Overspredningsparameter og EB-vekt lar seg ikke summere. Skal man beregne forventede ulykkestall for heterogene strekninger må man først beregne det forventede ulykkestall for hver (homogen) delstrekning (som funksjon av normalt ulykkestall, overspredningsparameter og registrerte ulykkestall for hver delstrekning). De forventede ulykkestallene kan man summere til det forventede ulykkestallet for hele strekningen.

Analysen for flere år: Skal man trendjustere resultatene må man gjøre én beregning for hvert år som skal inngå i analysen (trendjusteringen kan kun gjøres for ett år om gangen). Normale ulykkestall kan summeres for flere år. Forventede ulykkestall for flere år beregnes på samme måte som forventede ulykkestall for heterogene strekninger (se forrige punkt).

Trendjustering og trafikkmengde: Trendfaktoren (de relative risikotallene) justerer kun for endringer av ulykkesrisikoen over tid (dvs. antall ulykker / skadde / drepte per million kjøretøykilometer). Trafikkmengdeendringer over tid er ikke tatt hensyn til i trendjusteringen. Trafikkmengden må derfor alltid oppgis for det aktuelle året (året det skal trendjusteres til). Skal man for eksempel beregne det normale antall ulykker på en veg i 2014 oppgir man trafikkmengden på denne vegen i 2014. De beregnede normale ulykkestallene viser hvor mange ulykker man normalt kunne ha forventet på en veg med den gitte trafikkmengden i 2008. Ganger man dette tallet med 0,656 får man det normale antall ulykker på en veg med den gitte trafikkmengden i 2014.

6 Referanser

- Caliendo, C., Guida, M. & Parisi, A. (2007). A crash-prediction model for multilane roads. *Accident Analysis and Prevention*, 39, 657-670.
- Elvik, R. (2012). Oppdatering av ulykkesmodeller: Foreløpig analyse av data for Sør-Trøndelag. TØI-Arbeidsdokument 50195. Oslo: Transportøkonomisk institutt.
- Geedipally, S. R., Lord, D., & Park, B.-J. (2009). Analyzing different parameterizations of the varying dispersion parameter as a function of segment length. *Transportation Research Record*, 2103, 108-118.
- Hauer, E. (2001). Overdispersion in modelling accidents on road sections and in empirical Bayes estimation. *Accident Analysis and Prevention*, 33, 799-808.
- Hauer, E., Harwood, D. W., Council, F. M., & Griffith, M. S. (2002). Estimating safety by the empirical Bayes method: A tutorial. *Transportation Research Record*, 1784, 126-131.
- Høye, A. (2013). Ulykkesmodeller. TØI-Arbeidsdokument 50429. Oslo: Transportøkonomisk institutt.
- Lord, D., & Park, P. Y.-J. (2008). Investigating the effects of the fixed and varying dispersion parameters of Poisson-gamma models on empirical Bayes estimates. *Accident Analysis & Prevention*, 40(4), 1441-1457.
- Miaou, S.-P., & Lord, D. (2003). Modeling traffic crash-flow relationships for intersections. *Transportation Research Record*, 1840, 31-40.
- Ragnøy, A., Christensen, P. & Elvik, R. (2002). Skadegradstetthet - SGT. Et nytt mål på hvor farlig en vegstrekning er. TØI-rapport 618/2002. Oslo: Transportøkonomisk institutt.
- Ragnøy, A. & Elvik, R. (2002). Trafikksikkerhetsanalyse av stamvegnettet i Norge. TØI-rapport 649/2003. Oslo: Transportøkonomisk institutt.
- Strathman, J. G., Dueker, K. J., Zhang, J., & Williams, T. (2001). Analysis of design attributes and crashes on the Oregon highway system. Report FWHA-OR-RD-02-01. Center for Urban Studies. College of Urban and Public Affairs. Portland State University. Portland, Oregon.
- Vågane, L. (2012). Transportytelser i Norge 1946-2011. TØI-Rapport 1227/2012. Oslo: Transportøkonomisk institutt.

Flere relevante referanser

- Abdel-Aty, M., & Radwan, A. E. (2000). Modeling traffic accident occurrence and involvement. *Accident Analysis & Prevention*, 32, 633-642.
- Anastasopoulos, P. C., & Mannering, F. (2011). An empirical assessment of fixed and random parameter logit models using crash- and non-crash-specific injury data. *Accident Analysis & Prevention*, 43, 1140-1147.

- Anastasopoulos, P. C., Tarko, A. P., & Mannering, F. L. (2008). Tobit analysis of vehicle accident rates on interstate highways. *Accident Analysis & Prevention*, 40(2), 768-775.
- Bonneson, J. A., & Pratt, M. P. (2008). Procedure for developing accident modification factors from cross-sectional data. *Transportation Research Record*, 2083, 40-48.
- Cenek, P. D., Davies, R. B., & Henderson, R. J. (2012). Crash risk relationships for improved road safety management. NZ Transport Agency research report 488. Wellington, New Zealand.
- Couto, A., & Ferreira, S. (2011). A note on modeling road accident frequency: A flexible elasticity model. *Accident Analysis & Prevention*, 43(6), 2104-2111.
- Dietze, M. & Weller, G. (2011). Applying speed prediction models to define road sections and to develop accident prediction models: A German case study and a Portuguese exploratory study. Deliverable 6.2 of ERANET project RISMET. Dresden, Technical University of Dresden, 2011.
- Dinu, R. R., & Veeraragavan, A. (2011). Random parameter models for accident prediction on two-lane undivided highways in India. *Journal of Safety Research*, 42(1), 39-42.
- Findley, D. J., Hummer, J. E., Rasdorf, W., Zegeer, C. V., & Fowlere, T. J. (2012). Modeling the impact of spatial relationships on horizontal curve safety. *Accident Analysis & Prevention*, 45, 296-304.
- Fridstrøm, L., Ifver, J., Ingebrigtsen, S., Kulmala, R., & Krogsgård Thomsen, L. (1995). Measuring the contribution of randomness, exposure, weather, and daylight to the variation in road accident counts. *Accident Analysis & Prevention*, 27(1), 1-20.
- Garay, A. M., Hashimoto, E. M., Orteg, E. M. M., & Lachos, V. H. (2011). On estimation and influence diagnostics for zero-inflated negative binomial regression models. *Computational Statistics and Data Analysis*, 55, 1304-1318.
- Geedipally, S. R., Lord, D., & Park, B.-J. (2009). Analyzing different parameterizations of the varying dispersion parameter as a function of segment length. *Transportation Research Record*, 2103, 108-118.
- Geedipally, S. R., Lord, D., & Dhavala, S. S. (2012). The negative binomial-Lindley generalized linear model: Characteristics and application using crash data. *Accident Analysis & Prevention*, 45(0), 258-265.
- Hanley, K. A., Gibby, A., & Ferrara, T. (2000). Analysis of accident reduction factors on California state highways. *Transportation Research Record*, 1717, 37-45.
- Ivan, J. N., & O'Mara, P. J. (1997). Prediction of traffic accident rates using Poisson regression. Transportation Research Board 76th Annual Meeting, Washington DC.
- Ivan, J. N., Wang, C., & Bernardo, N. R. (2000). Explaining two-lane highway crash rates using land use and hourly exposure. *Accident Analysis & Prevention*, 32, 787-795.
- Kaminska, J., & Elvik, R. (2012). Choice of functional form and type of model to describe the relationship between independent variables and the dependent variable in accident prediction models. Unpublished Manuscript. Oslo, Institute of Transport Economics.
- Khattak, A. J., Khattak, A. J., & Council, F. M. (2002). Effects of work zone presence on injury and non-injury crashes. *Accident Analysis & Prevention*, 34, 19-29.
- Kim, D.-G., & Washington, S. (2006). The significance of endogeneity problems in crash models: An examination of left-turn lanes in intersection crash models. *Accident Analysis & Prevention*, 38(6), 1094-1100.

- Lord, D., & Mahlawat, M. (2009). Examining application of aggregated and disaggregated Poisson-gamma models subjected to low sample mean bias. *Transportation Research Record*, 2136, 1-10.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.
- Lord, D., & Persaud, B. (2000). Accident prediction models with and without trend. *Transportation Research Record*, 1717, 102-108.
- Lord, D., Washington, S., & Ivan, J. N. (2007). Further notes on the application of zero-inflated models in highway safety. *Accident Analysis & Prevention*, 39(1), 53-57.
- Lord, D., Washington, S. P., & Ivan, J. N. (2005). Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes: Balancing statistical fit and theory. *Accident Analysis & Prevention*, 37(1), 35-46.
- Lord, D., & Mannering, F. (2010). The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. *Transportation Research Part A: Policy and Practice*, 44(5), 291-305.
- Miaou, S.-P. (1996). Measuring the goodness of fit of accident prediction models. FHWA-RD-96-040.
- Miaou, S.-P., Lu, A., & Lum, H. S. (1996). Pitfalls of using R^2 to evaluate goodness of fit of accident prediction models. *Transportation Research Record*, 1542, 6-13.
- Milton, J., & Mannering, F. (1998). The relationship among highway geometries, traffic-related elements and motor-vehicle accident frequencies. *Transportation*, 25, 395-413.
- Mitra, S., & Washington, S. (2007). On the nature of over-dispersion in motor vehicle crash prediction models. *Accident Analysis & Prevention*, 39(3), 459-468.
- Noland, R. B., & Oh, L. (2004). The effect of infrastructure and demographic change on traffic-related fatalities and crashes: A case study of Illinois county-level data. *Accident Analysis & Prevention*, 36, 525-532.
- Oh, J., Lyon, C., Washington, S., Persaud, B., & Bared, J. (2003). Validation of FHWA crash models for rural intersections. *Transportation Research Record*, 1840, 41-49.
- Poch, M., & Mannering, F. (1996). Negative binomial analysis of intersection-accident frequencies. *Journal of Transportation Engineering*, 122(2), 105-113.
- Ragnøy, A., Christensen, P., & Elvik, R. (2002). Skadegradstetthet - SGT. Et nytt mål på hvor farlig en vegstrekning er. TØI-Rapport 618/2002. Oslo: Transportøkonomisk institutt.
- Sawalha, Z., & Sayed, T. (2001). Evaluating safety of urban arterial roadways. *Journal of Transportation Engineering*, 127(2), 151-158.
- Shankar, V. N., Albin, R. B., Milton, J. C., & Mannering, F. L. (1998). Evaluating median cross-over likelihoods with clustered accident counts: An empirical inquiry using random effects negative binomial model. *Transportation Research Record*(1635), 44-48.
- Shankar, V., Milton, J., & Mannering, F. (1997). Modeling accident frequencies as zero-altered probability processes: An empirical inquiry. *Accident Analysis & Prevention*, 29(6), 829-837.

- Shankar, V. N., Ulfarsson, G. F., Pendyala, R. M., & Nebergall, M. B. (2003). Modeling crashes involving pedestrians and motorized traffic. *Safety Science*, 41, 627-640.
- Vogt, A., & Bared, J. (1998). Accident models for two-lane rural roads: Segments and intersections. Report FHWA-RD-98-133.
- Wang, X., & Abdel-Aty, M. A. (2007). Investigation of signalized intersection right-angle crash occurrence at intersection, roadway, and approach levels. Paper presented at the 84th Annual Meeting of the TRB, Washington, D.C.
- Washington, S., Persaud, B., Lyon, C., & Oh, J. (2005). Validation of accident models for intersections. Federal Highway Administration, Final Report, FHWA-RD-03-037.
- Zhang, C., & Ivan, J. N. (2005). Effects of geometric characteristics on head-on crash incidence on two-lane roads in Connecticut. *Transportation Research Record*, 159-164.

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 70 høyt spesialiserte forskere.

Instituttet utgir tidsskriftet Samferdsel med 10 nummer i året 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
Gautstadalléen 21
NO-0349 Oslo

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