Modeling Driving Risk Using Naturalistic Driving Study Data

Youjia Fang

Dissertation submitted to the faculty of the Virginia Polytechnic Institute and State University in partial fulfillment of the requirements for the degree of

Doctor of Philosophy
in
Statistics

Feng Guo, Chair
Inyong Kim
Leanna L. House
Sheila G. Klauer

September 23, 2014
Blacksburg, Virginia

Keyword Naturalistic Driving Study, Driving Risk, Poisson Regression, Bayesian Model, Random Exposure, Hierarchical Model, Meta-Analysis

Modeling Driving Risk Using Naturalistic Driving Study Data

Youjia Fang

ABSTRACT

Motor vehicle crashes are one of the leading causes of death in the United States. Traffic safety research targets at understanding the cause of crash, preventing the crash, and mitigating crash severity. This dissertation focuses on the driver-related traffic safety issues, in particular, on developing and implementing contemporary statistical modeling techniques on driving risk research on Naturalistic Driving Study data. The dissertation includes 5 chapters. In Chapter 1, I introduced the backgrounds of traffic safety research and naturalistic driving study. In Chapter 2, the state-of-practice statistical methods were implemented on individual driver risk assessment using NDS data. The study showed that critical-incident events and driver demographic characteristics can serve as good predictors for identifying risky drivers. In Chapter 3, I developed and evaluated a novel Bayesian random exposure method for Poisson regression models to account for situations where the exposure information needs to be estimated. Simulation studies and real data analysis on Cellphone Pilot Analysis study data showed that, random exposure models have significantly better model fitting performances and higher parameter coverage probabilities as compared to traditional fixed exposure models. The advantage is more apparent when the values of Poisson regression coefficients are large. In Chapter 4, I performed comprehensive simulation-based performance analyses to investigate the type-I error, power and coverage probabilities on summary effect size in classical meta-analysis models. The results shed some light for reference on the prospective and retrospective performance analysis in meta-analysis research. In Chapter 5, I implemented classical- and Bayesian-approach multi-group hierarchical models on 100-Car data. Simulation-based retrospective performance analyses were used to investigate the powers and parameter coverage probabilities among different hierarchical models. The results showed that under fixed-effects model context, complex secondary tasks are associated with higher driving risk.

Acknowledgements

Five years came and went in the blink of an eye. I've had a great deal of fun and would like to send my sincerest thanks to all those who have helped me in these five years of PhD study.

First of all, Dr. Feng Guo, my PhD advisor, deserves a great deal of thanks for having provided me with so many understandings, supports, and guidance throughout the entire span of my graduate study and life. Thanks also go to my other committee members, Dr. Ingyong Kim, Dr. Leanna L. House, and Dr. Sheila Klauer for their patient support and help in my dissertation research and graduate work.

Thanks also go to all my team colleagues: Chen Chen, Dengfeng Zhang, Qing Li, Yi Liu, Yuming Shen. All of you always give me warmest support and inspiration. Thanks for all the good time and help you've given me over these years. My best wishes to all of you. Thanks are also due to all of the faculty and graduate student colleagues in Statistics Department I've asked for favors.

Finally, I would like to deliver my sincere thanks to my parents, my wife and my family. Thanks for your love, understanding and support in my entire study life. I wouldn't have the achievements without you.

Thank you all for everything you've done for me!

Table of Contents

List of Tables

List of Figures

Chapter 1 Introduction

1.1 Introduction of Traffic Safety Research

Motor vehicles are the most popular mode of transportation in the United States. Every year over 10 million motor vehicle crashes happen in the US and billions of dollars of damage are lost (U.S. Census Bureau 2012). Motor vehicle crashes are one of the leading causes of death in the US with over 30,000 fatalities annually (CDC 2010). With the goals of understanding the cause of crash, preventing the crash and mitigating crash severity, traffic safety research has become a widely recognized, highly invested, flourishing research area. Safety research outcomes influence the government, business, industry, and the everyday life.

Traffic safety research involves drivers, vehicles and driving environment. There is extensive research by traffic engineers on the safety impact of transportation infrastructure and traffic characteristics, e.g., the impacts of intersection design features, pavement conditions, weather, and traffic flow conditions (Hauer *et al.* 1988, Poch and Mannering 1996, Maze *et al.* 2006, Guo *et al.* 2010b, Lord and Mannering 2010). Crash occurrence is the primary risk measure for infrastructure-related safety impact evaluation, with Poisson and negative binomial (NB) models being the state-of-practice analysis tools. However, there is relatively limited research on individual driver risk in traffic and human factor engineering fields.

Compared to traffic engineers, the insurance and actuarial science industries have a longer history of research on classification of drivers according to risk level to facilitate underwriting and pricing. Estimation of the occurrence of claims based on the driver's age and other relevant variables has been a standard practice in actuarial research (Segovia-Gonzalez *et al.* 2009). For the insurance industry, quantified individual risk is directly related to the risk classification standards (Walters 1981). However, insurance data are in general proprietary and not available for public access.

1.2 Naturalistic Driving Study

Driving risk research relies heavily on the data collection process. There are three major types of data collection systems: Empirical Data Collection System, Epidemiological Data Collection System, and large-Scale Naturalistic Data Collection System (Dingus *et al.* 2006) (Table 1). Driver behavior plays a central role in driver risk, but it is difficult to measure in real-world driving situations. Recent developments in vehicle instrumentation techniques, such as in

Naturalistic Driving Study (NDS) (University of Michigan Transportation Research Institute 2005, Dingus *et al.* 2006, Guo and Hankey 2009) and the DriveCam system (Hickman *et al.* 2010) have made it both technologically possible and economically feasible to monitor individual driving behaviors and kinematic signatures on a large scale. These data collected through advanced in-vehicle instrumentation provide an opportunity to link driver behavior with risk at the individual driver level (Figure 1 from Dingus *et al.* (2006)). NDSs collect rich kinematic, Global Positioning System (GPS), radar, and video data at a high frequency, which provide an opportunity to detect abnormal driving situations.

Table 1 Traffic Safety Data Collection System

Data Collection	Example	Advantage	Disadvantage
Empirical collection	Test track Simulator	Proactive Ordinal crash risk	Imprecise Relies on unproven safety surrogate Experiment alters driver behavior
Epidemiological collection	Crash database	Precise crash risk Crash scenarios	Reactive Limited pre-crash info
Naturalistic collection	100-Car NDS	Natural behavior Pre-crash info Aggressive driving Drowsiness Driver errors Vehicle dynamics Surrogate validation	High economic and labor cost Long collection time Data processing and reduction Limited number of participants Strict participant privacy

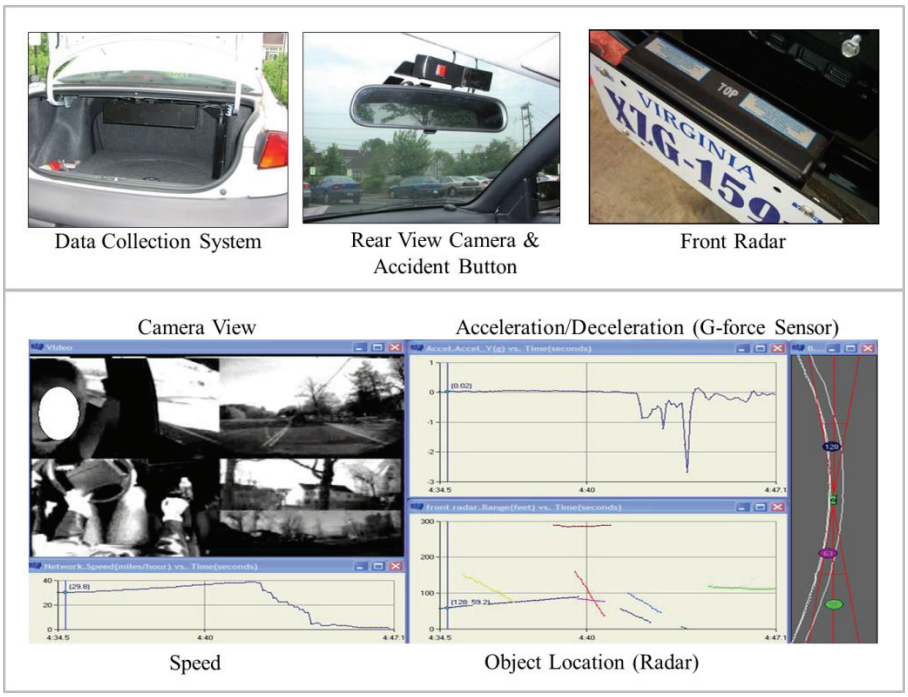

Figure 1 100-CarNDS Data Collection System (Dingus et al. 2006)

In a typical NDS, such as 100-Car NDS and Strategic Highway Research Program 2 (SHRP2), the NDS data collection system records each trip from the start to the end of the ignition, generating a time-series trip file. Upon a period of time, the data was transferred from the hard drive of data collection system to the database in research facilities. Different kinematic triggers - the mathematical algorithms that detect abnormal and abrupt changes (e.g. latitudinal/longitudinal g-force, yaw rate, radar information) - are applied simultaneously to determine when a certain type of Safety Critical Event (SCE) happened. There are three types of SCEs: Crash (C), where measurable contact is made between the subject vehicle and an object; Near-Crash (NC), where the rapid evasive maneuver is made to avoid a crash; and Critical Incident (CI), where a crash-relevant conflict is less severe than a Near-Crash, but more severe than normal driving (Fitch *et al.* 2013). Once a potential SCE is detected by the kinematic triggers, video reduction will be performed. The video reductionists open the file, visually examine the part of video record surrounding the event detected by the trigger, determine whether an event has really happened, and record necessary information on driver activities and environment conditions according to research protocol. The resolution of the video reduction can be high. For instance, the 100-Car study has videos record frequency of 10 Hz and every sync

equals to 0.1 second. Therefore, NDS data have high resolution and large data size, while also demand large human and financial resources.

1.3 Individual Driving Risk and Distracted Driving Risk

Individual driver risk can be affected by many factors. Demographic variables (e.g. age, gender, and education level), driver personality, and risky driving behaviors all play important roles in individual driving risk. Studies have shown that driver's age, physical and mental conditions are associated with different levels of individual driving risk (Antin *et al.* 2012, Guo and Fang 2013, Ouimet *et al.* 2014). Studies have also demonstrated that there is an association between driver personality characteristics and risky driving behavior (Jonah 1997, Jonah *et al.* 2001, Ulleberg and Rundmo 2003, Dahlen and White 2006, Machin and Sankey 2008, Guo and Fang 2013).

Distracted driving is a major contributing factor of traffic crashes (Stutts *et al.* 2001, Klauer *et al.* 2006, Olson *et al.* 2009, Klauer *et al.* 2010) . There are 3 major types of driving distractions: visual distraction (taking the eyes off road); manual distraction (taking the hands of the wheel), and cognitive distraction (taking the mind off the driving). Compared to cognitive distraction, visual and manual distractions are easier to capture by technology equipment such as video camera. Visual-Manual (VM) distractions are currently popular topics in traffic safety research. A good example of VM-distracted driving is cellphone use related driving. Cell phone use contributes to an estimated 25% of injury and property damage-only crashes (NSC 2009). Many studies have shown the connection between cellphone use related distracted driving and the elevated driving risk (Ishigami and Klein 2009, Council 2010, Fitch and Hanowski 2011, Fitch *et al.* 2013). While different cell phone use activities (dialing, talking, texting etc.) and different types of cellphone equipment (hand-held phone, Bluetooth headset, on-board calling system etc.) have different impacts on driving risk (Fitch *et al.* 2013). Other types of VM distractions, such as eating, drinking, reading, and looking at back seat, are also associated with higher driving risks (Klauer *et al.* 2006, Klauer *et al.* 2010).

The statistical analysis in driving risk research is based on classical epidemiological design methods. Some primary study design methods are case-crossover study (McEvoy *et al.* 2005, Klauer *et al.* 2010); case-control study (Fitch *et al.* 2013); and case-cohort study (Fitch *et al.* 2013, Klauer *et al.* 2014). Poisson and negative binomial regression models are state-of-

practice methods to more the SCE count data. Survival analysis is used to model the time-to-event data.

1.4 Dissertation Structure

This dissertation focused on developing and implementing statistical methods in NDS driving risk research. In Chapter 2, I presented the study results on identifying individual driver's driving risk using 100-Car NDS data, which is a published in the journal of Accident Analysis and Prevention (Guo and Fang 2013). In Chapter 3, I developed a Bayesian random exposure method for Poisson regression models to accommodate the situations where exposure information needs to be estimated. In Chapter 4, I performed comprehensive simulation studies to investigate the performance of classical meta-analysis models. In Chapter 5, I implemented and evaluated the Bayesian hierarchical models on multi-group driving risk using 100-Car NDS distraction data.

Chapter 2 Individual Driver Risk Assessment using NDS Data[1]

2.1 Introduction

The substantial variation in individual driving risk has been documented in many studies (Deery and Fildes 1999, Ulleberg 2001, Dingus *et al.* 2006). Identifying factors associated with individual driving risk and predicting high-risk drivers will enable proper driver-behavior intervention and safety countermeasures to reduce the crash likelihood of high-risk groups and improve overall driving safety.

NDSs collect rich kinematic, Global Positioning System (GPS), radar, and video data at a high frequency, which provides an opportunity to detect abnormal driving situations. In particular, I am interested in whether critical-incident events (CIEs) – non-crash safety events marked by a high acceleration/deceleration rate or other kinematic signatures – can be used to predict high-risk drivers. The premise is that critical incidents are caused by driver behaviors similar to that of CNCs. Since critical incidents happen at a much higher frequency (100 times the frequency of crashes and 10 times the frequency of near-crashes), this provides an opportunity to identify high-risk drivers before crashes actually happen. This will allow designing and implementing proactive safety countermeasures to improve the safety of the high-risk drivers.

The objectives of this study (Guo and Fang 2013) are twofold. The first objective is to investigate the risk factors associated with individual driving risk. The second objective is to build up a model to predict high-risk drivers, which includes two steps: identification using cluster analysis, and prediction using a logistic regression model. The 100-Car Naturalistic Driving Study was used for methodology development and application.

2.2 Materials and Methods

2.2.1 The 100-Car naturalistic driving study data

The 100-Car Naturalistic Driving Study is the first large-scale NDS conducted in the United States (Dingus *et al.* 2006). The study included 102 primary drivers in northern Virginia. In order to catch as much safety critical events as possible, the samples lean towards young drivers

[1] This chapter is published on Accident Analysis and Prevention, vol. 61, pp. 3-9. I acknowledge publisher Elsevier for the acceptance of the paper and the permission for inclusion of the paper in this chapter of dissertation.

and high mileage drivers. The vehicles of the participants were instrumented with advanced data acquisition systems. The system included five camera views (forward, driver face, over the shoulder, left and right mirror), GPS, speedometer, three-dimension accelerometer, and radar, etc. Driving data were collected continuously for 12 months. The study collected data for approximately 2,000,000 vehicle miles and almost 43,000 hours of data.

The data were reduced based on the kinematic and video records. Three types of safety-related events were identified: crashes, near-crashes, and safety-critical events (Dingus *et al.* 2006, Klauer *et al.* 2006). A crash is defined as an event with "any contact between the subject vehicle and another vehicle, fixed object, pedestrian, pedacyclist, or animal" (Dingus *et al.* 2006, p. xvii). The crash involves kinetic energy transfer or dissipation. A near-crash is "a conflict situation that requires a rapid, severe evasive maneuver to avoid a crash. The rapid, evasive maneuver involves conducting maneuvers that involve steering, braking, accelerating, or any combination of control inputs that approaches the limits of the vehicle capabilities" (Dingus *et al.* 2006, p. xvii).

The CIE is a conflict less severe than the near-crash. CIEs were detected by three approaches (Dingus *et al.* 2006): 1) flagging events where the car sensors exceeded a specified value (e.g., brake response of >0.6 g); 2) when the driver pressed an incident pushbutton located on the data acquisition system; 3) through analysts' judgments when reviewing the video. A rigorous data reduction was implemented by using different threshold values for the kinematic threshold values and visual confirmation.

Although not a safety concern by itself, the CIE can be regarded as a measure of driving aggressiveness. The hypothesis is that a relatively safe driver, based on his/her driving skills and safety consciousness, will try to avoid evasive maneuvers that could lead to a hazardous scenario, including a CIE. A high rate of CIEs reflects the lack of such skills and safety consciousness; thus, the rate of CIEs is an indicator of driving aggressiveness. If the above hypothesis holds, the rate of CIEs will be a good predictor for individual driver risk.

Other factors that may be associated with different driving risks include age, gender, and personality. The 100-Car Naturalistic Driving Study included a survey that measures personalities based on the NEO Five-Factor Inventory, which includes the following five aspects: Neuroticism (N), Extroversion (E), Openness to Experience (O), Agreeableness (A), and Conscientiousness (C) (Costa and McCrea 1992, Klauer *et al.* 2006). A number of research

studies have been conducted to evaluate the relationship between the NEO five factors with driving safety (Shaw and Sichel 1971, Loo 1979, Arthur and Graziano 1996, Klauer *et al.* 2006).

Due to the relatively small number of crashes, near-crashes are commonly used as a crash surrogate. Several research studies in risk assessment using NDS used near-crashes in conjunction with crashes for risk assessment (Klauer *et al.* 2006, Guo *et al.* 2010b, Klauer *et al.* 2010). Guo *et al.* (2010b) concluded that the near-crash is a valid crash surrogate for risk assessment purposes. Based on the research cited above, a combination of crash and near-crash events was used as a risk metric for individual driving risk.

2.2.2 Statistical methods

The study was designed to evaluate two objectives: assess risk factors and predict high-risk drivers. For the first objective, a state-of-the-practice Negative Binomial (NB) model was used to assess the relationship between the CNC risk and potential risk factors. There are two steps for the prediction of high-risk drivers. First, a K-mean cluster analysis was used to identify high-risk driver groups. Logistic regression models were then developed to predict the high-risk drivers using the risk factors identified in the first objective. The prediction performance of the logistic regression model was evaluated by the receiver operating characteristics curve (ROC). The details of the models and the analysis techniques are discussed in this section.

2.2.2.1 Negative Binomial model for evaluating risk factors

The NB regression model is state-of-the-practice for traffic safety modeling (Lord and Mannering 2010). The model assumes that the observed frequency of crashes and near-crashes for driver *i*, Y_i, follows an NB distribution,

$$Y_i \sim NB(E_i\lambda_i, \gamma)$$

where λ_i is the expected CNC rate for driver *i*, as measured by the number of CNCs per 1,000 miles; E_i is the miles traveled by driver *i* (per 1,000 miles); and γ is the NB over-dispersion parameter. A log link function connects λ_i with a set of covariates,

$$\log(\lambda_i) = X_i\beta,$$

where $\boldsymbol{X_i}$ is the matrix of covariates for driver *i* and β is the vector of regression parameters. In this study, the age, gender, and personality score based on the NEO five-factor inventory, and the critical incident were used as covariates.

2.2.2.2 Cluster analysis for identifying high-risk drivers

The main criterion for evaluating the overall risk of individual drivers is by the CNC rate. The cluster analysis provides an objective approach to classify drivers into different risk levels and has been used in traffic safety research (Donmez *et al.* 2010). A K-mean cluster method was adopted to classify primary drivers into different risk groups based on CNC rate. The K-mean cluster partitions the observations into *k* clusters with a predetermined number of clusters (Tan *et al.* 2005). An observation is assigned to the cluster whose mean is closest to its value. The K-mean method minimizes the within-cluster sum of squares:

$$\arg\min_{s} \sum_{i=1}^{k} \sum_{x_j \in S_i} \left\| X_j - \mu_i \right\|^2,$$

where (X_1 , X_2 , … X_n) are the observed data which are the CNC rates in the context of this chapter; $\boldsymbol{S} = (S_1 , \ldots S_k)$ is the set of *k* clusters; and μ_i is the mean of the observations in set S_i. Each driver was classified into one of three clusters (high-, moderate-, and low-risk groups). Drivers in the clusters with the highest mean CNC rate were considered to be high-risk drivers.

2.2.2.3 Logistic regression models for predicting high-risk drivers

After risk groups were identified through cluster analysis, two logistic regression models were developed to model the probability of being a high-risk driver. The first model evaluates the probability of high-risk drivers only, while the second model evaluates the probability of high- or moderate-risk drivers. The two models could support the interest of researchers with different perspectives. The model setup is as follows. Define

$$\mathrm{Y_i} = \begin{cases} 1 & \text{If driver i is a high risk driver (or a moderate risk driver)} \\ 0 & \text{Otherwise} \end{cases}$$

Let p_i be the probability of being a risky driver for drive *i*. The observed Y_i is assumed to follow a Bernoulli distribution, i.e. $Y_i \sim Bernoulli(p_i)$. The key parameter is the probability of being a high/moderate risk driver, p_i. This probability is associated with a set of covariates by a logit link function,

$$\mathrm{logit}(p_i) = X_i\beta$$

where X_i is the vector of predictors for individual *i*, and β is the vector of j regression parameters. The exponential of regression parameter, $\exp(\beta_j)$, is the odds ratio (OR) for the j^{th} variable. The CIE rate, age group, and personality score were used as driver characteristics. The logistic regression will estimate the probability of being a risky driver based on predictors. A driver will be predicted as a risky driver if this probability is greater than a predefined threshold value p_0 .

The predictive performance of the logistic models was evaluated by the ROC curve (Agresti 2002), which measures model sensitivity and specificity. In the context of this study, the sensitivity is the probability of correctly predicting a risky driver, and the specificity is the probability of correctly predicting a safe driver, as shown in the following formula, i.e.,

Sensitivity = Probability (Classified as risky driver | the driver is risky)

Specificity = Probability (Classified as safe driver | the drive is safe)

Both measures were related to the threshold value p_0 and there is a tradeoff between sensitivity and specificity. The ROC curve is a plot of sensitivity versus false positive rate; i.e., (1 – Specificity), for all possible thresholds p_0's. The performance of the prediction model can be measured by the area under the curve (AUC): a higher AUC value indicates better prediction power for the logistic regression model. A perfect prediction method would yield the maximum AUC of 1. A completely random guess would give a diagonal line in the ROC space with AUC of 0.5.

2.3 Results

2.3.1 Exploratory data analysis

The 100-Car Study data include 60 crashes, 675 near-crashes, and 7,394 critical incidents from primary drivers. The event rate was calculated as number of events per 1,000 miles traveled:

$$\text{Event Rate} = \frac{\text{Number of events}}{\text{Miles travelled (1,000 miles)}}$$

Based on overall risk by age and sample size considerations, three age groups were defined: younger than 25 years, between 25 and 55 years, and older than 55 years. The summary statistics stratified by age and gender are shown in Table 2.

Table 2 Summary Statistics by Age and Gender

Variables	Age <25		Age 25-55		Age >55	
	Male	Female	Male	Female	Male	Female
Number of Drivers	16	18	39	16	8	5
Total Number of CIEs	1234	2209	2490	930	490	41
Total Number of CNCs	163	224	174	105	61	8
Subject Miles (KMiles)	160.7	204.2	525.2	142.9	105.2	192.0
Mean CIE rate[a]	8.2	11.37	4.861	7.63	4.57	2.579
Mean CNC rate[a]	1.11	1.27	0.38	0.73	0.58	1.10
Mean CIE rate[a]	9.88		5.67		3.81	
Mean CNC rate[a]	1.20		0.48		0.78	

[a] unit of rate is number of events per 1000 miles traveled

Drivers under the age of 25 had the highest CIE and CNC rates among all the age groups. Drivers between 25 and 55 had a higher CIE rate than did drivers older than 55 but had a lower CNC rate. The CNC and CIE rates also vary by gender and age group. Male drivers have lower CIE and CNC rates than female drivers in the <25 and 25-55 age groups. The gender difference is not consistent for drivers older than 55. Male drivers over 55 had a lower CNC rate but a higher CIE rate as compared to female drivers over 55.

The NEO five-factor personality inventory represents various aspects of personality using five variables. The data analysis indicated that the Extroversion, Agreeableness, and Conscientiousness factors have strong correlations with the response CNC rate, as shown in Table 3. However, the five variables themselves are highly correlated. Including all factors in the same model will lead to multicollinearity issues and biased inference. Choosing a subset of variables could mitigate the multicollinearity issues but would lead to insufficient use of information. The principal component analysis (PCA) was adopted to address the multicollinearity and maintain the maximum information from the five variables.

Table 3 Pearson Correlation Coefficients between NEO-5 Scores and Response

	N[a]	O[a]	E[a]	A[a]	C[a]	CIE Rate	CNCRate
N	1.00	0.64[b]	0.62	0.59	0.35	0.03	-0.17
		<.0001[c]	<.0001	<.0001	0.0004	0.7851	0.0955
O		1.00	0.62	0.58	0.43	0.03	-0.12
			<.0001	<.0001	<.0001	0.7762	0.2049
E			1.00	0.70	0.63	-0.13	-0.20
				<.0001	<.0001	0.2111	0.0473
A				1.00	0.65	-0.20	-0.26
					<.0001	0.0456	0.0093
C					1.00	-0.14	-0.21
						0.1706	0.0381

[a] N - Neuroticism ,O - Openness to Experience, E - Extroversion, A - Agreeableness, and C- Conscientiousness

[b] Correction Coefficients

[c] P-value, two sided test under the null hypothesis of zero correlation

The PCA uses an orthogonal transformation to convert correlated variables into a set of uncorrelated variables called principal components (Jolliffe 2002). The first principal component has the highest variance and accounts for the largest portion of the variability in the data. Each succeeding component in turn has the highest variance possible under the constraint of orthogonality (uncorrelated) with the preceding components. A principal component is a linear combination of optimally weighted observed variables.

The first step of the principal component analysis is to identify significant principal components for the set of correlated observed variables. The eigenvalue-one criterion was used to choose the significant component, which states that a component contains substantial information if the corresponding eigenvalue is greater than 1. As can be seen from Table 4, the eigenvalue of the first component is 3.365 (much larger than 1) and all four of the other components have eigenvalues smaller than 1. The first component could contribute to 67.3% of the variability in the data. Therefore, the first component from the PCA was used to represent the personality scores. A sensitivity analysis was also conducted using the first two components and

the results indicate essentially identical prediction power. Therefore, one component is considered sufficient in risk modeling.

Table 4 Eigenvalues for Principal Components

Component	Eigenvalue	Difference	Proportion	Cumulative
1	3.365	2.679	0.673	0.673
2	0.686	0.311	0.137	0.810
3	0.374	0.067	0.074	0.885
4	0.306	0.03	0.061	0.946
5	0.267	-	0.053	1.000

The results of the PCA provide the following formula for the personality score:

$$\text{PCA personality score}_i = 0.232N_i^* + 0.261E_i^* + 0.237O_i^* + 0.256A_i^* + 0.231C_i^*$$

where the N_i^* to C_i^* are standarized values for driver i. For example, $N_i^* = \frac{N_i - \overline{N}}{std.dev.(N)}$, with the $\overline{N}$ and *std.dev.(N)* being the mean and standard deviation of the observed variable N, respectively.

2.3.2 Negative Binomial models for risk factor evaluation

The NB regression model was fitted using the CNC frequency as the response variable and the miles travelled as exposure (per 1,000 miles). Three covariates were included: the CIE rate as measured by the number of CIEs per thousand miles, the age group, and the personality score as computed by PCA. The gender variable does not have significant impacts on the CNC rate and was excluded from the NB model and subsequent analysis. The model fitting results are shown in Table 5. As can be seen, all three factors are highly significant. The over-dispersion parameter is quite small (0.262), which indicated the presence of over-dispersion and justified the use of NB regression. The point estimate for the critical incident parameter is 0.091, which implies that for every one unit of increase in critical incident rate, the CNC rate will increase to a multiplicative factor of *exp*(0.091)=1.09. That is an approximately 10% increase in CNC rate for every one unit of increase in CIE rate.

The 25-55 age group showed the lowest crash and near-crash rates. The CNC rate ratio between drivers younger than 25 years and drivers between 25 and 55 years is *exp*(0.541)=1.72; for drivers older than 55 and drivers between 25 and 55 years it is *exp*(0.481)=1.62.

Table 5 Parameter Estimation for Negative Binomial Models

Parameters	Estimate	Standard Error	Wald 95% Confidence Limits		P-Value
Intercept	-1.557	0.125	-1.802	1.312	<.0001
CI Rate	0.091	0.011	0.070	0.112	<.0001
Age: <25 vs. 25-55	0.541	0.157	0.234	0.848	0.0006
Age: >55 vs. 25-55	0.481	0.234	0.021	0.940	0.0402
Personality	-0.317	0.082	-0.477	-0.157	0.0001
Dispersion	0.262	0.069	0.141	0.424	

2.3.3 Identify driver risk groups

The K-mean cluster method was applied to the 102 primary drivers based on CNC rate. The number of clusters was predefined to be three, to represent the high, moderate, and low driver risk groups. The number of clusters is determined based on sample size and ease of interpretation for the modelling results.

The output of the cluster analysis is illustrated in Figure 2. A relatively small number of drivers were in the risk groups (6 drivers in the high-risk group and 12 drivers in the moderate-risk group). As the goal of the study is to identify risky drivers, the relatively small numbers in these two groups fit the context well. The within-cluster variations for the low-, moderate-, and high-risk groups were 0.31, 0.44, and 0.44, respectively.

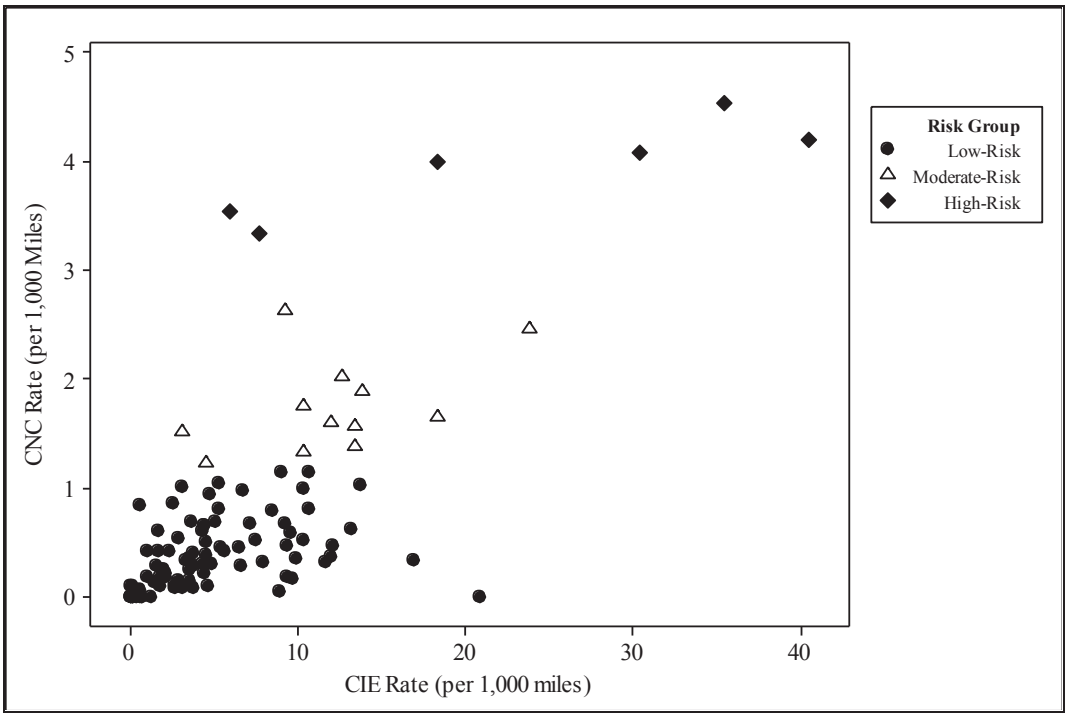

Figure 2 Cluster Analysis Results

The characteristics of the three risk groups are summarized in Table 6. As can be seen, the CNC rate of the high-risk group is 10 times that of the safe (i.e., low-risk) group, and the rate of the moderate-risk group is more than 4 times that of the safe group. The average age of the safe group is substantially higher (38.1). The overall pattern of the NEO personality factors suggests that the low-risk group has relatively high values in all five factors and the high-risk group has relatively low values in the NEO five-factors.

Table 6 Characteristics of Driver Risk Groups

Risk Groups	Number of drivers	Mean CNC rate	% of males in each group	Mean age	Means of the NEO personality factors				
					A	E	O	N	C
Low-Risk Group	84	0.39	65.5	38.1	38.0	36.7	35.1	25.4	37.2
Moderate-Risk Group	12	1.75	58.3	28.7	33.3	33.5	31.7	22.5	32.2
High-Risk Group	6	3.95	16.7	30.0	29.7	31.7	34.7	21.3	32.0

2.3.4 Logistic prediction models results for predicting high-risk drivers

The cluster analysis identified the high-, moderate-, and low-risk groups. The key question is

whether the high-risk drivers can be predicted by the driver characteristics. Depending on specific research questions, it could be of interest to predict extremely high-risk drivers or moderate- to high-risk drivers. Therefore, two logistic prediction models were developed. The first model predicted high-risk drivers against moderate-risk/safe drivers. The second model predicted high/moderate-risk drivers against the safe drivers. The risk factors identified in the NB regression model were used in logistic regression; i.e. the critical incident rate, age group, and PCA component based on the NEO-5 personality score. The model outputs are summarized in Table 7.

Table 7 Logistic Regression Model Output

Models	Effect	Parameter Estimate	P-Value	Odds Ratio	95% Odds Ratio Confidence Limits	
Model 1: High vs. Moderate/Low Risk	Intercept	-11.1	0.023	-	-	-
	CIE	**0.346**	**0.019**	**1.414**	**1.058**	**1.889**
	Age: <25 vs. 25-55	4.37	0.175	79.42	0.142	>999
	Age: >55 vs. 25-55	6.46	0.108	638.2	0.242	>999
	Personality	-0.900	0.432	0.407	0.043	3.838
Model 2: High/Moderate vs. Low Risk	Intercept	-5.62	<.0001	-	-	-
	CIE	**0.296**	**<0.001**	**1.345**	**1.150**	**1.571**
	Age: <25 vs. 25-55	**2.13**	**0.021**	**8.456**	**1.378**	**51.90**
	Age: >55 vs. 25-55	2.36	0.059	10.64	0.918	123.3
	Personality	-0.48	0.368	0.616	0.214	1.770

In both models the CIE rate had a significant impact on the probability of being a risky driver. The OR was calculated to quantitatively evaluate the impacts of each variable The OR represents the relative odds of being a risky driver for every one unit increase in a continuous variable (critical incident rate and personality score), or relative risk between two levels of a categorical covariate (the age group variable). The results indicated that, for every one unit increase in CIE rate, the relative odds of being a high-risk driver will increase by 41%

(OR=1.414). Based on Model 2, every one unit increase in CIE rate will increase the relative odds of being a moderate/high-risk driver by 35% (OR=1.345).

The personality score variable, differing from the NB models, did not show significant results in the logistic regression models. The age group variables are not significant in the prediction model for high-risk drivers. However, the model for predicting moderate/high-risk drivers indicates a significant difference between the young driver group (<25) and the middle age group (25-55). One potential cause for the discrepancy between factors identified in the NB model and the logistic regression model is that the cluster process masks the CNC rate difference among drivers within the same group.

The predictive models are as follows,

Model 1: Probability (high-risk driver)

$$= \frac{\exp(-11.1 + 0.346 \times CIE\ Rate + 4.37 \times AgeL25 + 6.46 \times AgeG55 - 0.900 \times PER)}{1 + \exp(-11.1 + 0.346 \times CIE\ Rate + 4.37 \times AgeL25 + 6.46 \times AgeG55 - 0.900 \times PER)},$$

Model 2: Probability (high- or moderate-risk driver)

$$= \frac{\exp(-5.62 + 0.296 \times CIE\ Rate + 2.13 \times AgeL25 + 2.36 \times AgeG55 - 0.48 \times PER)}{1 + \exp(-5.62 + 0.296 \times CIE\ Rate + 2.13 \times AgeL25 + 2.36 \times AgeG55 - 0.48 \times PER)},$$

where *PER* is the standardized personality score, *AgeL25* is an indicator variable on whether the driver age is less than 25; and *AgeG55* is an indicator variable on whether driver age is greater than 55.

To evaluate the prediction performance, ROCs for both models were generated as shown in Figure 3. The solid bold lines are the ROC curves, and the straight diagonal dashed lines are the reference lines. Both models showed high predictive power. The AUC is 0.938 for Model 1 and 0.930 for Model 2, both close to the perfect AUC value of 1.

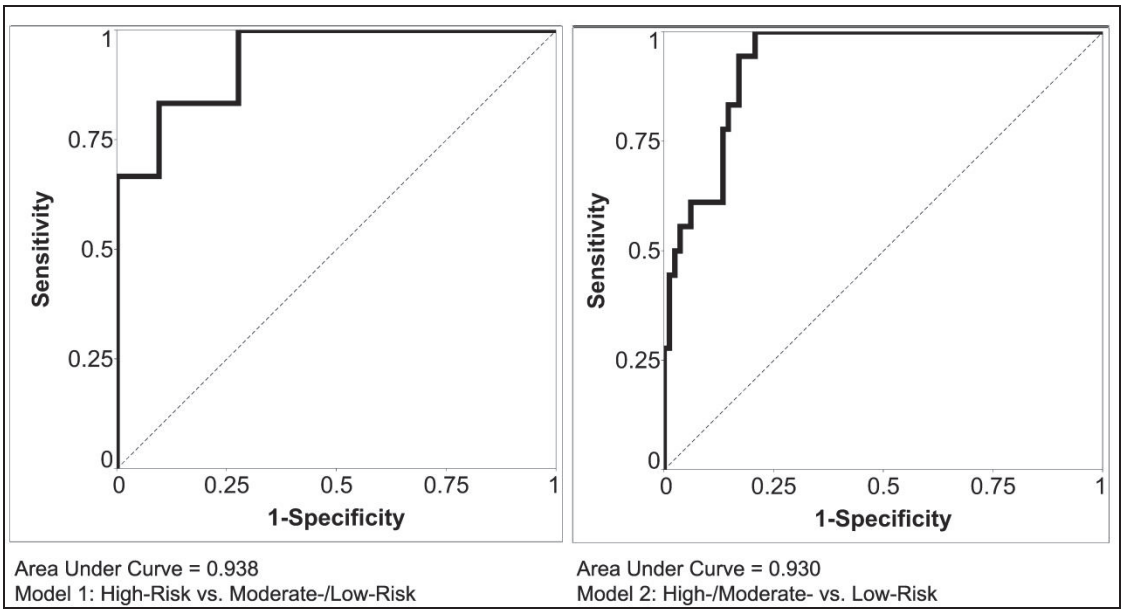

Figure 3 The ROC Curves

2.4 Discussion

The NDSs collects rich real-life driving data for an extended period of time. The continuous data collection approach provides the opportunity to evaluate not only crash risk, but also non-crash driving behavior. This study evaluated the individual driving risk and the risk factors associated with high-risk drivers using the 100-Car Naturalistic Driving Study data, the first large-scale NDS conducted in the United States.

Driving risk varies substantially among drivers. The cluster analysis indicated that about 6% of the drivers had a substantially higher risk (10 times higher than low-risk groups) and about 12% of the drivers showed moderate to high risk (4 times higher than the low-risk group). This result is consistent with the substantial variation in driving risk observed from previous NDS, epidemiological, self-reporting, and simulator studies (Deery and Fildes 1999, Ulleberg 2001, Dingus *et al.* 2006, Donmez *et al.* 2010). Although high-risk drivers only account for a small proportion of the driver population, they have a substantial impact on overall traffic safety. The ability to identify high-risk drivers will provide a valuable reference for developing safety education programs, regulations, and proactive safety countermeasures.

The NB regression model indicated that driver personality, age, and CIE rate had significant impacts on the CNC risk for individual drivers. It is well known that young and elderly drivers have a higher risk as compared to other age groups (National Highway Traffic

Safety Administration 2008). The relative risk between age groups from this analysis is consistent with the national data.

The relationship between driver personality and driving risk has been evaluated in previous studies (Ulleberg and Rundmo 2003, Dahlen and White 2006, Machin and Sankey 2008). It is surprising to observe high correlations among the five factors, which were designed to measure different aspects of personality. The PCA was used to utilize maximum information from all five factors without inducing the multicollinearity issue. This study confirmed that the NEO five-factor inventory did associate with CNC risk for individual drivers.

A primary focus of this study is whether CIE rate, a measure of driving characteristics, can be used to assess driving risk and predict high/moderate-risk drivers. The results confirmed that the CIE rate has a strong relationship with individual driving risk. Furthermore, a logistic prediction model using critical incident rates can successfully identify high- and moderate-risk drivers. The CIE rate had a statistically significant influence on the prediction. For every one-unit increase in CIE rate, the relative probability of being a high-risk or moderate-to-high risk driver increased by approximately 41% and 35%, respectively.

The strong association between the CIE rate and individual driver risk can have significant implications on the individual risk assessment and safety interventions. Since the number of crashes for individual drivers is often limited, predicting risky drivers using past crashes history may be inefficient. The CIEs occur at a much higher frequency than crashes (one hundredfold higher) and near-crashes (tenfold higher). This makes it possible to proactively identify the high-risk population. This is particularly important for developing proactive safety countermeasures and for improving the safety of high-risk drivers. There have been studies that not only monitored driver behaviors but also attempted to improve safety by providing feedback to alter driver behaviors (Donmez *et al.* 2010). The insurance industry has also begun incorporating driving characteristics into pricing; e.g., the SnapShot® program by the Progressive Casualty Insurance Company.

Traffic crashes are rare events, thus, surrogates are needed when there is not a sufficient number of crahes for safety assessment (Tarko *et al.* 2009). Traffic conflict is one of the most widely used surrogates (Williams 1981, Hauer and Garder 1986, Tiwari *et al.* 1998). Surrogates are especially critical for NDSs which usually observe a limited number of crashes but with high resolution. Guo et al. (2010b) concluded that the near-crash could provide valuable information

on driving risk and can serve as a crash surrogate for risk-assessment purposes, especially for time-variant risk factors such as distraction. The strong association between the CIE rate and CNC risk at the individual driver level, as well as the strong prediction power for high-risk, drivers implies that CIEs could be a valid safety surrogate for driver risk analyses.

Although this study is limited by the voluntary and relatively small sample size, a significant relationship was observed between individual driving risk and factors such as demographic, personality, and driving characteristics. More importantly, it demonstrates that naturalistic data can provide opportunities to identify individual driver risk differences from less severe safety events, that when observed over time can provide a more complete profile of unsafe driving behavior. This study is a first step and additional studies with larger and more representative data are clearly needed. Large scale naturalistic studies currently underway in the US (Committee for the Strategic Highway Research Program 2: Implementatation 2009) and in other countries, such as the EuroFOT program (Benmimoun *et al.* 2009), hold promise for providing the data needed to expand upon the research described in this chapter. The outcomes of this research can help designers, educators, and researchers developing better proactive safety countermeasures and safety programs to improve the driving safety.

2.5 Conclusion

Using the 100-Car Naturalistic Driving Study data, this study showed that individual drivers' driving risk varies substantially with three distinct risk groups. The cluster analysis identified approximately 6% of drivers as high-risk drivers, which accounted for 24% of total crash and near-crash events; and 12% of drivers as high/moderate-risk drivers, which accounted for 45% of total crash and near-crash events. The NB regression model indicated that age, personality, and CIE rate had significant impacts on individual drivers' CNC risk. The logistic prediction models for high- and moderate-risk groups had high predictive powers using the CIE rate. This study concluded that CNC risk for individual drivers is associated with CIE rate, age, and personality characteristics. Furthermore, the CIE rate is an effective predictor for high-risk drivers.

Chapter 3 Bayesian Random Exposure Method for Poisson Regression Model

3.1 Introduction

Driver distraction has been a major contributing factor for traffic safety. Recent studies have illustrated that visual-manual (VM) tasks, such as texting and dialing, could significant increase driving risk (Klauer *et al.* 2014). The previous studies are based on a case-cohort approach which relies on random sample baselines to estimate the cellphone use exposure during normal driving conditions (Guo *et al.* 2010a). Recently, the National Highway Traffic Safety Administration (NHTSA) sponsored a naturalistic driving study, the Cellphone Pilot Analysis (CPA) study, (Fitch *et al.* 2013), to evaluate the cellphone use risk. Compared with previous NDS studies, the CPA collected detailed cellphone record information from carriers. This allows the study to use a risk-rate approach based on Poisson regression models to assess the risk of VM tasks. The core part of this approach is to combine the cellphone records and video reduction to estimate the duration of VM tasks. The Fitch et al 2013 report treated estimated exposure as fixed value, which does not account for the uncertainty associated with the estimation. This study developed a random exposure Poisson exposure model to incorporate this uncertainty.

Poisson and negative binomial regression models are the most widely used modeling technique for count data. In traffic safety research, these two models are the state-of-practice in modeling crash and traffic safety event count. Poisson regression is one type of Generalized Linear Regression (GLM) (Agresti 2002), which models the relationship between response variables from an exponential distribution family and a linear combination of covariates. Poisson regression models the expected event rate as a linear combination of covariates via a link function. A typical model formation of Poisson regression model with log-link function follows as:

$$Y_i \sim Poisson\left(E_i\lambda_i\right),$$

$$\text{with } \log\left(E_i\lambda_i\right) = \log\left(E_i\right) + \beta_0 + \sum_{j=1}^{p}\beta_j X_{ij},$$

$$\text{RateRatio}\left(\beta_j\right) = \exp\left(\beta_j\right),$$

where E_i is the exposure for observation *i*; λ_i is the mean event rate; $\text{RateRatio}\left(\beta_j\right)$ is the event rate ratio given one unit change of covariate X_{ij}. The Exposure E_i is a critical component of the

Poisson model as it is the base for normalizing raw event frequencies to event rates. The exposure is a quantitative measure of the length of opportunity when events could happen. In driving safety research, the exposure can be driving mileage or time the driver engaged in certain distraction. Inaccurate exposure information could lead to biased estimation of crash risk.

Conventionally, the exposure term in Poisson regression model is considered as fixed and obtained by either direct observation or estimation from the study. When the exposure can be directly observed and recorded (e.g. the number of miles driven, the number of hours driven), treating exposure as fixed in the Poisson model is appropriate. However, in many situations the exposure is difficult to be observed directly and needs to be estimated. For the CPA project the research question is whether VM tasks for cell phone is associated with increased driving risk (Fitch *et al.* 2013). The response variable is the number of Safety Critical Events (SCEs), and the exposure is the total duration of driving with cell phone VM and talking tasks (talking, dialling, text messaging etc.). Extracting the total time of VM tasks from all videos is costly prohibitive. The strategy used in Fitch *et al.* (2013) is to sample a portion of cell phone use from phone record, make detailed video reduction for all activities, and estimate the durations for each subtasks based on cell phone record and data reduction results. Due to uncertainty associated with the estimation process, the estimated exposure should be considered as a random quantity rather than a fixed value. Failure to incorporate the uncertainty associated with the estimation of exposure could lead to biased estimation and jeopardize the validity of statistical inference.

The goal of this chapter is to develop a random exposure model to accommodate the uncertainty associated with the estimation of exposure. Under Bayesian framework, I treated the estimated exposure as random to incorporate the uncertainty. The posterior distribution of the exposure reflects the randomness associated with exposure, and the posterior distributions of regression parameters will inherently incorporate the uncertainty of the exposure.

The structure of this chapter is as follows. In Section 3.2, I briefly introduced CPA study data. In Section 3.3, I examined the theoretical properties of expectation and variance of the Maximum Likelihood Estimate (MLE) of Poisson regression model parameters using fixed and random exposure. Then I focused on the situation where the base exposure segment follows an Exponential distribution. Two Gamma-distributed priors were proposed for the total exposure. The mean and variance of the Poisson model using different priors were evaluated. In Section 3.4, simulation studies were conducted to examine the performances of proposed models under

different data settings. In Section 3.5, the proposed random exposure models were applied to CPA data to evaluate texting related driving risk. Section 3.6 concluded this chapter.

3.2 Overview of Cellphone Pilot Analysis Study Data

Cellphone Pilot Analysis (CPA) study is a naturalistic driving study with data acquisition systems installed in the participants' own vehicles captured their in situ behavior by continuously recording video from multiple cameras and various kinematic data, GPS, and vehicle CAN information. The study collected data from 204 drivers who each participated in the study for approximate 31 days from February, 2011 to November, 2011. Drivers were recruited from Northern Virginia, Southern Virginia, and the Raleigh-Durham region of North Carolina. Only drivers that reported talking on a cellphone while driving at least once per day were recruited in order to increase the probability of observing cellphone use. A total of 129 females and 75 males participated in the study, with average age of 41 years old (range from 18 to 84). A key feature of this study is that drivers provided their cell phone records. The cell phone records allowed the determination of when drivers used their cell phone and the direction of the call/text message, while the video data allowed the determination of the type of cell phone used, how long it was used for, and what subtasks were performed. The CPA study was approved by the Virginia Tech Institutional Review Board, and all drivers provided informed consent.

A key step of risk assessment is to estimate the total duration that drivers were engaged in VM tasks. This was done by combining information from both cell phone records and video reduction from a sub-samples cellphone use. The cellphone records provided the total number of cellphone uses, and the video reduction from the samples provided the durations of VM tasks for the sample. The estimation of total VM duration for a driver *i* is illustrated in Figure 4. In this example, E_i is total duration of VM for driver *i*; n_i is the number of VM tasks for driver *i*, and L_{ik} is duration of each VM activity (see Figure 4). Because it is impractical to measure durations of all the VM tasks, only a subset of cellphone use activities was sampled. This estimation process will inherently contain uncertainty.

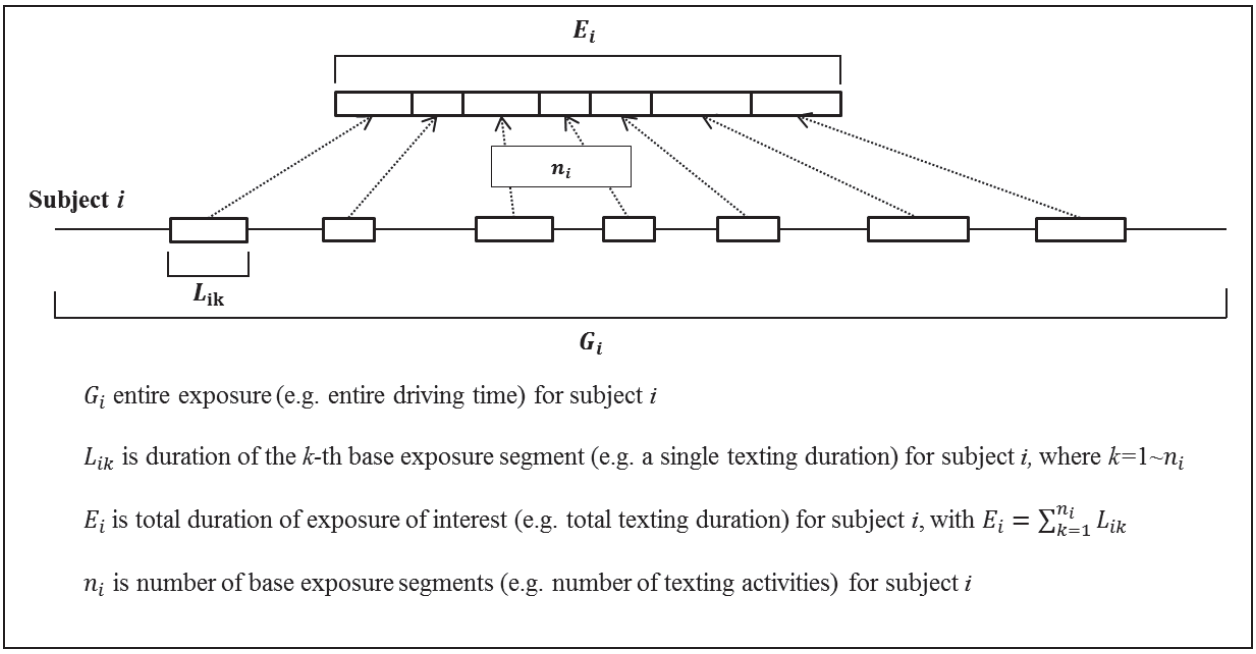

Figure 4 Example of Random Exposure Scenario

3.3 Poisson Regression Models with Random Exposure

3.3.1 Theoretical properties of MLE when exposure is random

In the case of a traditional Poisson model without covariates, we proved that the Maximum Likelihood Estimate (MLE) for the event rate parameter λ has an unbiased mean and a larger variance when the exposure is random, as compared with when exposure is considered as fixed. Consider a Poisson model without covariates as

$$Y_i \sim Poisson(E_i\lambda)$$

where:

Y_i is the number of events for subject i, with i =1~I;

E_i is the total length of exposure for subject i;

λ is the expected event rate.

Regardless of whether E_i is fixed or random, the MLE for λ conditional on E_i's is as follows:

$$L(\lambda|E_i) = \prod_{i=1}^{I} \frac{e^{-E_i\lambda} \times (E_i\lambda)^{y_i}}{y_i!},$$

$$l = lnL = -\lambda\sum E_i + \sum y_i \ln(E_i) + ln(\lambda)\sum y_i - \sum ln(y_i!),$$

$$\frac{\partial l}{\partial \lambda} = -E\sum n_i + \frac{1}{\lambda}\sum y_i \overset{set}{=} 0,$$

$$\Rightarrow \hat{\lambda}_{mle} \mid E_i = \frac{\sum y_i}{\sum E_i}$$

When exposure is fixed

When the total exposure E_i's are known and fixed, we define $E_i = \mu_i$, the MLE for $\lambda \mid E_i$ is giving by $\hat{\lambda}_{fixed\_E} = \frac{\sum y_i}{\sum \mu_i}$.

Based on the properties of unconditional mean and variance, the unconditional mean of the estimator $\hat{\lambda}_{fixed\_E}$ is as follows. It is an unbiased estimator for λ .

$$E\left(\hat{\lambda}_{fixed\_E}\right) = E_Y\left[\frac{\sum y_i}{\sum E_i}\right] = \frac{\sum E_i \lambda}{\sum E_i} = \lambda .$$

The unconditional variance of the estimator $\hat{\lambda}_{fixed\_E}$ is as follows. The variance converges to zero with the increase in $\sum \mu_i$.

$$Var\left(\hat{\lambda}_{fixed\_E}\right) = Var_Y\left[\frac{\sum y_i}{\sum E_i}\right] = \frac{\sum E_i \lambda}{\left(\sum E_i\right)^2} = \frac{\lambda}{\sum E_i} = \frac{\lambda}{\sum \mu_i}$$

When exposure is random

When E_i is a random variable that follows a distribution with mean μ_i and variance Σ_i , the MLE for $\lambda \mid E_i$ is $\hat{\lambda}_{random\_E} = \frac{\sum y_i}{\sum E_i}$.

Corollary 1: *The MLE for the Poisson rate λ is an unbiased estimator for λ when the exposure E_i is random; the variance of the MLE for the Poisson rate λ is larger when the exposure is random than when the exposure is considered as fixed.*

This can be shown easily as follows:

$$E\left(\hat{\lambda}_{random\_E}\right) = E_{E_i}\left[E_{Y|E_i}\left[\frac{\sum y_i}{\sum E_i}\right]\right] = E_{E_i}\left[\frac{\sum E_i \lambda}{\sum E_i}\right] = \lambda .$$

$$Var\left(\hat{\lambda}_{random\_E}\right) = E_{E_i}\left[Var_{Y|E_i}\left[\frac{\sum y_i}{\sum E_i}\right]\right] + Var_{E_i}\left[E_{Y|E_i}\left[\frac{\sum y_i}{\sum E_i}\right]\right]$$

$$= E_{E_i}\left[\frac{\sum E_i\lambda}{\left(\sum E_i\right)^2}\right] + Var_{E_i}\left[\lambda\right] = E_{E_i}\left[\frac{\lambda}{\sum E_i}\right] + 0 \geq \frac{\lambda}{E_{E_i}\left[\sum E_i\right]} = \frac{\lambda}{\sum \mu_i}.$$

The inequality is a direct result of Jensen's Inequality $\mathrm{E}\left(\frac{1}{\mathrm{X}}\right) \geq \frac{1}{\mathrm{E}(\mathrm{X})}$ since the exposure is always positive.

The results can be extended to Poisson regression model with covariates. Poisson regression models with log link function do not have closed-form analytical solution for MLE estimators of regression parameters. Here, we considered a simplified situation - the Poisson regression with identity link, with only one covariate and with no intercept, and assume that $\sum E_i x_i > 0$. This situation has a closed form analytical solution for MLE estimator.

Consider a Poisson regression model with one covariate and identity link function as

$$Y_i \sim Poisson\left(E_i\lambda_i\right)$$

$$\text{with} \quad g\left(\lambda_i\right) = \lambda_i = x_i\beta_1$$

where

Y_i is the number of events for subject *i*, with *i* =*1*~*I*;

E_i is the exposure for subject *i*;

λ_i is the expected event rate for subject *i*;

β_1 is the regression parameter;

x_i is the covariate for subject *i*;

g() is the identity link function for Poisson model.

And we have the assumption that $\sum E_i x_i > 0$.

The MLE estimator for $\beta_1 \mid E_i$ is derived as

$$L = \prod_{i=1}^{K} \frac{e^{-E_i\lambda_i} \times \left(E_i\lambda_i\right)^{y_i}}{y_i!},$$

$$l = \ln L = -\sum E_i \lambda_i + \sum y_i \ln\left(E_i \lambda_i\right) - \sum \ln\left(y_i !\right)$$
$$= -\beta_1 \sum E_i x_i + \sum y_i \ln\left(E_i x_i\right) + \ln \beta_1 \sum y_i - \sum \ln\left(y_i !\right)$$

$$\frac{\partial l}{\partial \beta_1} = -\sum E_i x_i + \frac{1}{\beta_1} \sum y_i \overset{set}{=} 0 \ ,$$

$$\Rightarrow \hat{\beta}_{1,mle} \mid E_i = \frac{\sum y_i}{\sum E_i x_i}$$

When exposure is fixed

When E_i 's are fixed and known, we define $E[E_i] = \mu_i$, the MLE for $\beta_1 \mid E_i$ is

$\hat{\beta}_{1,fixed\_E} = \frac{\sum y_i}{\sum E_i x_i}$. Based on the properties of unconditional mean and variance, the

unconditional mean of the estimator $\hat{\beta}_{1,fixed\_E}$ is

$E\left(\hat{\beta}_{1,fixed\_E}\right) = E_Y\left[\frac{\sum y_i}{\sum E_i x_i}\right] = E_Y\left[\frac{\sum E_i \beta_1 x_i}{\sum E_i x_i}\right] = \beta_1$. The unconditional variance of the estimator

$\hat{\beta}_{1,fixed\_E}$ is as follows. It converges to zero as $\sum E_i x_i$ increases.

$$Var\left(\hat{\beta}_{1,fixed\_E}\right) = Var_Y\left[\frac{\sum y_i}{\sum E_i x_i}\right]$$
$$= Var_Y\left[\frac{\sum E_i x_i \beta_1}{\left(\sum E_i x_i\right)^2}\right] = \frac{\beta_1}{\sum E_i x_i} = \frac{\beta_1}{\sum \mu_i x_i}$$

When exposure is random

When E_i is a random variable that follows a distribution with mean μ_i and variance Σ_i , the

MLE for $\beta_1 \mid E_i$ is $\hat{\beta}_{random\_E} = \frac{\sum y_i}{\sum E_i x_i}$.

Corollary 2: *For Poisson regression models with identity-link function, one covariate and no intercept, and assume that* $\sum E_i x_i > 0$ *, the MLE for the regression coefficient is unbiased when exposure* E_i *is random. The variance of the MLE is larger when exposure is random than the when exposure is considered as fixed.*

This can be shown as follows:

$$E\left(\hat{\beta}_{random\_E}\right) = E_{E_i}\left[E_{Y|E_i}\left[\frac{\sum y_i}{\sum E_i x_i}\right]\right] = E_{E_i}\left[\frac{\sum E_i \beta_1 x_i}{\sum E_i x_i}\right] = \beta_1 .$$

$$Var\left(\hat{\beta}_{random\_E}\right) = E_{E_i}\left[Var_{Y\mid E_i}\left[\frac{\sum y_i}{\sum E_i x_i}\right]\right] + Var_{E_i}\left[E_{Y\mid E_i}\left[\frac{\sum y_i}{\sum E_i x_i}\right]\right]$$

$$= E_{E_i}\left[\frac{\sum E_i \beta_1 x_i}{\left(\sum E_i x_i\right)^2}\right] + Var_{E_i}\left[\beta_1\right] = E_{E_i}\left[\frac{\beta_1}{\sum E_i x_i}\right] + 0 \geq \frac{\beta_1}{E_{E_i}\left[\sum E_i x_i\right]} = \frac{\beta_1}{\sum \mu_i x_i}$$

The inequality is the result of Jensen's Inequality $\mathrm{E}\left(\frac{1}{\mathrm{X}}\right) \geq \frac{1}{\mathrm{E}(\mathrm{X})}$ when we assume $\sum E_i x_i > 0$.

3.3.2 Bayesian random exposure Poisson models

3.3.2.1 Three prior setups for Bayesian Poisson regression models

As discussed above, treating exposure as fixed will lead to underestimation for the variance and potentially invalid inference. We proposed random exposure Bayesian models to incorporate the uncertainty associated with exposures. The information on exposure was incorporated into the model via informative priors.

Consider a Bayesian Poisson model as follows,

$$Y_i \sim Poisson\left(\Lambda_i\right),$$

$$\log \Lambda_i = \log\left(E_i\right) + \beta_0 + \beta_1 \times x_i ,$$

Vague priors can be used on regression coefficients, i.e. $\beta_0, \beta_1 \sim N\left(0, precision = 10^{-6}\right)$, this way the data will dominate posterior distribution. The Rate Ratio (RR) equals to $\exp(\beta_1)$.

The prior for exposure E_i reflects the *a priori* information on exposure. The form of the prior distribution is determined by the stochastic process from which the exposure was generated and the randomness in the estimation procedure. The following three setups were used for the prior structure.

- Model 1 (Fixed): fixed exposure.
- Model 2 (Random 1): Exposure is generated from an exponential distribution; estimated via sample mean

- Model 3 (Random 2): Exposure is generated from an exponential distribution; estimated via estimated distribution.

Model 1 is the classical scenario where the exposure can be fully observed. There is no uncertainty with the exposure information. Model 2 and Model 3 fits the scenarios where the exposure is from a random distribution and has to be estimated and thus there is inherited uncertainty.

3.3.2.2 Prior elicitation for Random 1 model

The prior for Model 2 is motivated by the CPA study. In this study, there are 116 drivers with log record of text messaging activities. 12 drivers of them do not have any text messaging at all. To maintain the consistency of the simulation study and the real data analysis, we excluded these 12 drivers, retaining 104 drivers with text messaging activities. By comparing the NDS data and texting log, there are 5819 texting activities occurred during driving from the 104 drivers. Notation-wise, assume the number of texting while driving for driver *i* is n_i ; the total number of texting event is $N = \sum_{i=1}^{I} n_i$. A subsample of *M* texting activities were randomly sampled for video reduction and used for estimating total texting-related VM task duration for each driver. The estimation procedure is illustrated in Figure 5.

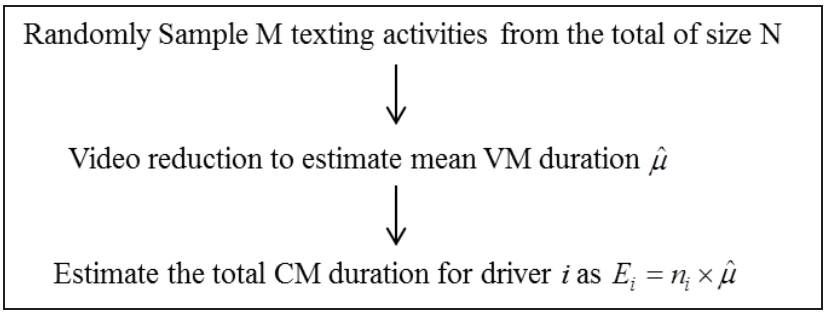

Figure 5 Estimation procedure for Model 2

Under this scenario, the primary source of uncertainty comes from the random sampling of texting activities and estimation procedure. To elicit prior from this procedure, following procedure was used.

I proposed a Gamma-distributed informative prior for the random exposure by considering the precision of the sample. I consider that the base exposure duration L_{ik} (i.e. the *k*-th texting activity for driver *i*) follows an Exponential distribution with mean μ , i.e.

$L_{ik} \sim Exponential(mean = \mu)$. This consideration is supported by CPA data. Table 8 and Figure 6 show that, the texting durations in CPA texting samples follow an exponential distribution. The mean parameter μ can be estimated using an independent and identically distributed (i.i.d.) random sample of size M from the pool all L_{ik}'s. The advantage of sampling is that we do not need to make video reduction on all activities, which could be cost prohibitive. I denote the sample by L_j's (neglecting the driver index i), with $j=1...M$. The estimation for μ is $\hat{\mu} = \frac{1}{M}\sum_{j=1}^{M} L_j$. By not including a driver-level index, I imply that the samples from all drivers are treated equally.

Table 8 Goodness-of-fit tests of exponential distribution on CPA texting samples

Test	Statistic		p-Value	
Kolmogorov-Smirnov	D	0.033	Pr > D	>0.250
Cramer-von Mises	W-Sq	0.099	Pr > W-Sq	>0.250
Anderson-Darling	A-Sq	0.579	Pr > A-Sq	>0.250

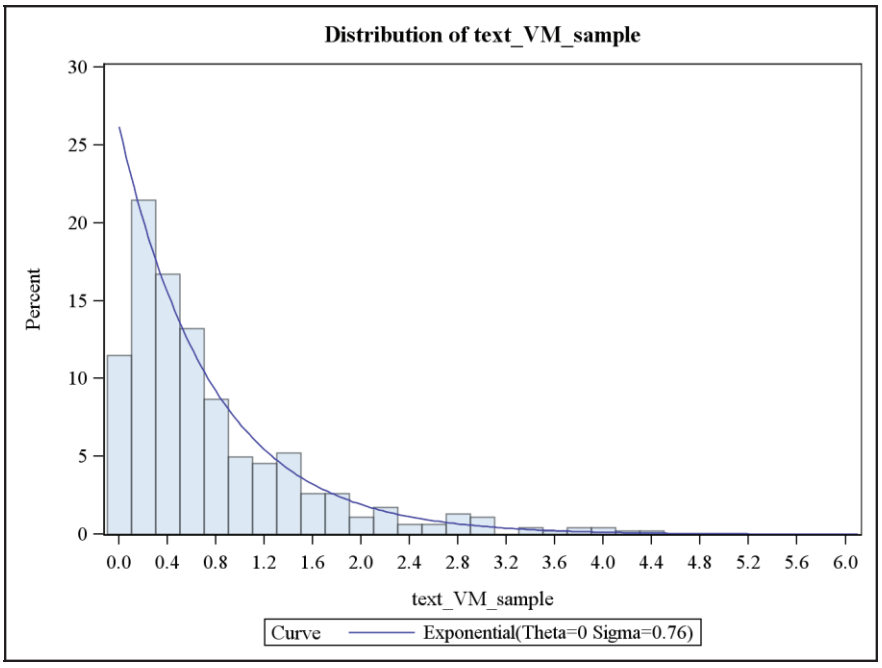

Figure 6 Distribution of texting durations in CPA texting samples

Because the summation of the i.i.d. exponential random variables is equivalent to a Gamma random variable, it follows that

$$\sum_{j=1}^{M} L_j \sim Gamma\left(shape = M, rate = \frac{1}{\mu} \right),$$

$$\text{with } E\left(\sum_j L_j \right) = M\mu \text{, and } Var\left(\sum_j L_j \right) = M\mu^2 .$$

Then based on the properties of Gamma random variable, it follows that

$$\bar{L} = \frac{1}{M} \sum_{j=1}^{M} L_j \sim Gamma\left(shape = M, rate = \frac{M}{\mu} \right),$$

$$\text{With } E\left(\bar{L}\right) = \mu \text{, and } Var\left(\bar{L}\right) = \frac{\mu^2}{M} .$$

The exposure E_i can be estimated as $\hat{E}_i = n_i \times \bar{L}$, with n_i the number of texting activities for subject *i* (known from the phone log). It follows that

$$\hat{E}_i = n_i \times \bar{L} = n_i \times \frac{1}{M} \sum_{j=1}^{M} L_j \sim Gamma\left(shape = M, rate = \frac{M}{n_i \mu} \right),$$

$$\text{with } E\left(\hat{E}_i\right) = n_i \mu \text{, and } Var\left(\hat{E}_i\right) = \frac{n_i^2 \mu^2}{M} .$$

By replacing μ with $\hat{\mu} = \frac{1}{M} \sum_{j=1}^{M} L_j$, I propose the first Gamma informative prior for model 2 as

$$E_i \mid M, n_i, \hat{\mu} \sim Gamma\left(shape = M, rate = \frac{M}{n_i \hat{\mu}} \right).$$

When M is large and n_i is small, the variance of E_i becomes small, and this prior can converge to a point mass with the value equivalent to the fixed-exposure prior in Model 1 (i.e. $E_i = n_i \hat{\mu}$) .

3.3.2.3 Prior elicitation for Random 2 model

In model 2, the informative prior when we estimate E_i as $E_i = n_i \times \hat{\mu}$ only take the sampling error into account. It is based on an important assumption: the pool of *N* texting activities is fixed but unknown. In other words, if we increase the sample size to *M=N*, then it is for certain that $\hat{\mu} \equiv \mu$, and $E_i = n_i \times \hat{\mu} = n_i \times \mu$ will be without little randomness.

However, if we consider the pool of *N* texting activities are from an infinitely large population of texting activities (beyond the scope of CPA), then the pool is random, and we should incorporate this portion of uncertainty. The estimation procedure is illustrated in Figure 7.

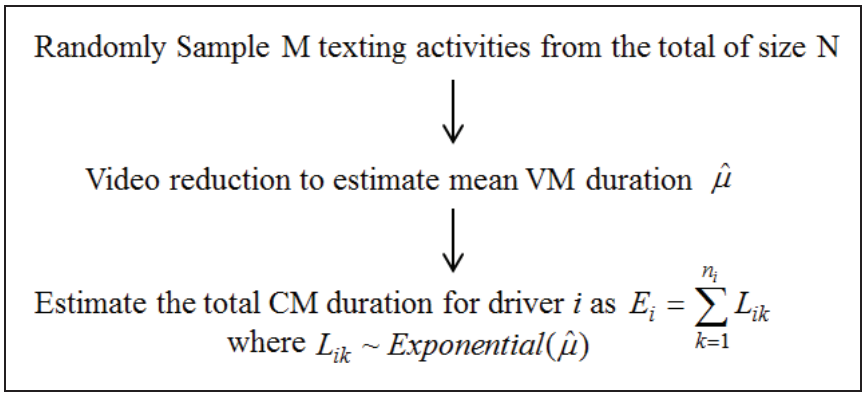

Figure 7 Estimation procedure for Model 3

Under this scenario, the source of uncertainty is the combination of random sampling of *M* samples from the pool of *N* texting activities and the randomness of *N* testing activities from the entire population of texting activities. To elicit prior from this procedure, following procedure was used.

In Model 3, I proposed a more general Gamma-distributed informative prior by considering $E_i = \sum_{k=1}^{n_i} L_{ik}$. As we assume that $L_{ik} \sim Exponential(mean = \alpha)$, based on the distributional properties, it follows that

$$E_i = \sum_{k=1}^{n_i} L_{ik} \sim Gamma\left(shape = n_i, rate = \frac{1}{\mu}\right),$$

$$\text{with } E(E_i) = n_i\mu, \text{ and } Var(E_i) = n_i\mu^2 .$$

By replacing μ with $\hat{\mu} = \frac{1}{M}\sum_{j=1}^{M} L_j$, I propose the Gamma informative prior for model 3 as

$$E_i \mid n_i, \hat{\mu} \sim Gamma\left(shape = n_i, rate = \frac{1}{\hat{\mu}}\right).$$

The variance of E_i becomes large as n_i increases, but it is not affected by the value of *M*.

3.3.3 Bayesian model formulation

3.3.3.1 Models for data setting 1 - one type of exposure

In the first data setting, there is only one type of exposure. So each subject has one row of data.

With $\hat{\mu} = \frac{1}{M}\sum_{j=1}^{M} L_j$, the Poisson regression model is as follows.

$$Y_i \sim Poisson\left(\Lambda_i\right)$$

$$\log \Lambda_i = \log\left(E_i\right) + \beta_0 + \beta_1 \times x_i$$

$$\text{Rate Ratio (RR)} = \exp\left(\beta_1\right)$$

$$\beta_0, \beta_1 \sim N\left(0, precision = 10^{-6}\right)$$

Model 1 (Fixed): $E_i = n_i \hat{\mu}$

Model 2 (Random 1): $E_i \mid M, n_i, \hat{\mu} \sim Gamma\left(shape = M, rate = \frac{M}{n_i \hat{\mu}}\right)$

Model 3 (Random 2): $E_i \mid n_i, \hat{\mu} \sim Gamma\left(shape = n_i, rate = \frac{1}{\hat{\mu}}\right)$

In each model, two MCMC chains, each with 30,000 iterations, were performed, with first 20,000 as burn-in period, and the last 10,000 iterations pooled together for posterior. Model convergence was checked and assured using Gelman-Rubin statistic (Gelman and Rubin 1992) and trace plots. For demonstration purpose, two examples of trace plot from Section 3.4.3.2 (data setting 2, $\beta_0 = 2$, $\beta_1 = 4$) are shown in Figure 8. After several attempts, the selected numbers of iterations ensure the model convergence.

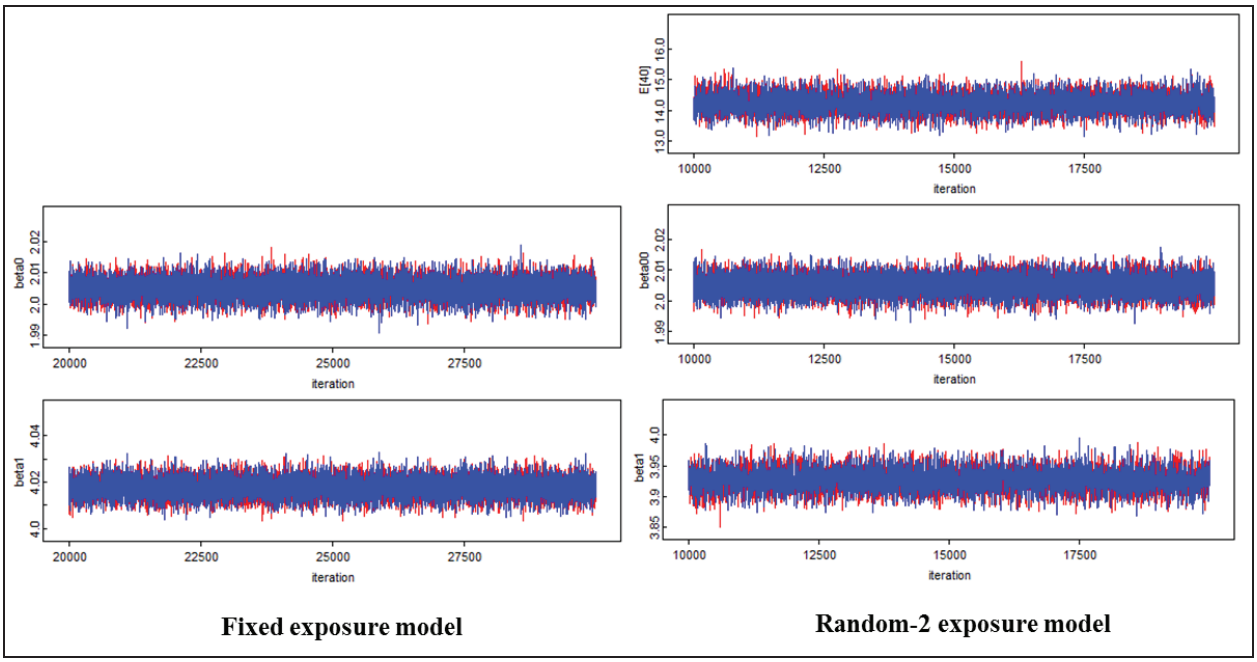

Figure 8 Example of trace plot

3.3.3.2 Models for data setting 2 - two types of exposures

In this data setting, there are two types of exposures, and the covariate is an indicator variable. Under CPA context, the covariate represents the driving with texting activity (1) or just driving (0), and the entire exposure, i.e. the total driving time G_i (the combination of driving with texting and without texting), is fixed and known. In this situation, each driver has 2 rows of data, indexed by *z=1* (driving with texting) and *z=2* (driving without texting). Thus, the covariate is defined as $x_{i,z=1}=1$, and $x_{i,z=2}=0$ .The Poisson regression model is as follows.

$$Y_{iz} \sim Poisson\left(\Lambda_{iz}\right)$$

$$\log \Lambda_{iz} = \log\left(E_{iz}\right) + \beta_0 + \beta_1 \times x_{iz}$$

$$\text{Rate Ratio (RR)} = \exp\left(\beta_1\right)$$

$$\beta_0, \beta_1 \sim N\left(0, precision = 10^{-6}\right)$$

Model 1 (Fixed): $E_{i,z=1} = n_i\hat{\mu}$, $E_{i,z=2} = G_i - n_i\hat{\mu}$

Model 2 (Random 1): $E_{i,z=1} \mid M, n_i, \hat{\mu} \sim Gamma\left(shape = M, rate = \frac{M}{n_i\hat{\mu}}\right)$

$$E_{i,z=2} \mid E_{i,z=1} = G_i - E_{i,z=1}$$

Model 3 (Random 2): $E_{i,z=1} \mid n_i, \hat{\mu} \sim Gamma\left(shape = n_i, rate = \frac{1}{\hat{\mu}}\right)$

$$E_{i,z=2} \mid E_{i,z=1} = G_i - E_{i,z=1}$$

For Model 1, two MCMC chains, each with 30,000 iterations, are performed, with first 20,000 as burn-in period, and the last 10,000 iterations from each chain pooled together for posterior. For Model 2 and Model 3, because there was severe autocorrelation in the MCMC samples of β_0 and β_1, thinning technique is used as remedy. The MCMC samples are sequentially sample in a distance of 20 iterations. The thinned samples are less auto-correlated. This way, two MCMC chains (each with 400,000 iterations) are performed, with first 200,000 iterations as burn-in period. The 10,000 samples from each chain of last 200,000 iterations were pooled together for posterior distribution.

3.4 Simulation Studies

Section 3.4.1 introduced the data generating mechanism based on the estimation procedure of Model 3 (see Figure 7 for details), which is the primary focus of this chapter. In Section 3.4.2 and 3.4.3, I conducted extensive simulation studies to investigate the model fitting performances and parameter coverage probabilities of the proposed Bayesian Poisson regression models.

3.4.1 Data generating mechanism

3.4.1.1 Data setting 1 - one type of exposure

In this data setting, each subject has one row of data (y_i , x_i , n_i , E_i), generated as follows.

- Define the number of subject $n_{subject}$; sample size M; mean parameter μ ; lower and upper bonds of discrete uniform distribution n_{lower}, n_{upper}; and regression parameters β_0 and β_1
- For $i = 1 \ldots I$, sample n_i as $n_i \mid \mathrm{n_{lower}}, \mathrm{n_{upper}} \sim DiscreteUniform(\min = n_{lower}, \max = n_{upper})$
- Generate the exposure data $E_{i,true}$'s and compute $\hat{\mu}$ by
    - Sample all N base exposure segments into a vector $L_{all,\mathrm{vector}}$

      i.e. $L_{all,vector} \mid N, \mu \sim Exponential(nsample = N = \sum n_i, scale = \mu)$

- Randomly sample M base exposure segments from $L_{all,vector}$, and create a vector $L_{j,vector}$ (each element is denoted as L_j), and calculate $\hat{\mu}$ as $\hat{\mu} = \frac{1}{M}\sum_{j=1}^{M} L_j$
- For each subject i, randomly sample (without replacement) n_i base exposure segments from the pool $L_{all,\text{vector}}$ to form a vector $L_{i,\text{vector}}$ (each element is denoted as L_{ik}, with $k=1 \ldots n_i$). Then calculate the total exposure $E_{i,true}$ as $E_{i|true} = \sum_{k=1}^{n_i} L_{ik}$

- For each subject i, generate a covariate x_i as $x_i \sim Uniform(-1,1)$
- For each subject i, generate the events y_i as $y_i \mid \beta_0, \beta_1 \sim Poisson\left(E_{i|true} \times \exp\left(\beta_0 + \beta_1 x_i\right)\right)$
- For each subject i, calculate the fixed-exposure estimate as $E_{i|fixed} = n_i \times \hat{\alpha}$

3.4.1.2 Data setting 2 - two types of exposures

In this data setting, each subject has two rows of data: ($y_{i,z=1}$, n_i, $x_{i,z=1}=1$, $E_{i,z=1}$) and ($y_{i,z=2}$, $x_{i,z=2}=0$, $E_{i,z=2}$, G_i). Data are generated as follows.

- Define $n_{subject}$, M, μ, n_{lower}, n_{upper}, β_0, β_1, G_{lower}, G_{upper}.
- For each subject i, sample G_i as

  $G_i \mid G_{lower}, G_{upper} \sim DiscreteUnif(\min = G_{lower}, \max = G_{upper})$
- For each subject i, sample n_i as

  $n_i \mid \mathrm{n}_{\text{lower}}, \mathrm{n}_{\text{upper}} \sim DiscreteUniform(\min = n_{lower}, \max = n_{upper})$
- Generate the exposure data $E_{i,\mathrm{z}=1|true}$'s, $E_{i,\mathrm{z}=2|true}$'s and calculate $\hat{\mu}$ by
  - Sample all N texting-related base exposure segments into a vector $L_{all,\text{vector}}$, i.e.

    $L_{all,vector} \mid N, \mu \sim Exponential(nsample = N = \sum n_i, scale = \mu)$
  - Randomly sample M base exposure segments from $L_{all,vector}$ to form a vector

    $L_{j,vector}$ (each element is denoted as L_j), and calculate $\hat{\mu}$ as $\hat{\mu} = \frac{1}{M}\sum_{j=1}^{M} L_j$

- For each subject i, sample (without replacement) n_i base exposure segments from $L_{all,\text{vector}}$ to form a vector $L_{i,\text{vector}}$ (each element is denoted as L_{ik}, with $k=1...n_i$). Then calculate the total texting-related exposure $E_{i,z=1|true}$ as $E_{i,z=1|true} = \sum_{k=1}^{n_i} L_{ik}$.
- Calculate $E_{i,z=2|true}$ as $E_{i,z=2|true} = G_i - E_{i,z=1|true}$.

- For each subject i, generate the count data $y_{i,z=1}$ and $y_{i,z=2}$ as

  $y_{i,z=1} \mid \beta_0, \beta_1 \sim Poisson\left(E_{i,z=1|true} \exp\left(\beta_0 + \beta_1 \times 0\right)\right)$,

  $y_{i,z=2} \mid \beta_0, \beta_1 \sim Poisson\left(E_{i,z=2|true} \exp\left(\beta_0 + \beta_1 \times 1\right)\right)$.

- For each subject i, calculate the fixed-exposure estimate as $E_{i,z=1|fixed} = n_i \times \hat{\mu}$, and $E_{i,z=2|fixed} = G_i - E_{i,z=1|fixed}$.

3.4.2 Simulation studies for model fitting performances

3.4.2.1 Model fitting performance for data setting 1

In Section 3.4.2, I generated data for Poisson regression models with different parameter settings and data structures, and evaluated the model fitting performances using Deviance Information Criterion (DIC) measure (Spiegelhalter *et al.* 2002, Gelman *et al.* 2004) and the standard deviations of the Poisson rate ratio estimates.

The parameter settings for data setting 1 are listed in Table 9. A total 16 parameter settings cover the high (H) and low (L) levels of: number of subjects (I); number of base exposure segments for each subject (n_{lower} and n_{upper}); sample size for estimating $\hat{\mu}$ (M); magnitude of regression parameters (β_0 and β_1). Because the magnitude of x_i's is not of the primary interest, I set lower and upper bonds of x_i's as -1 and 1 respectively. Moreover, I set $\mu = 2$. For each parameter setting, 3 Bayesian models were implemented: Fixed model (Model 1), Random 1 model (Model 2) and Random 2 model (Model 3).

Table 9 Parameter settings for data setting 1 (for DIC)

setting	I	n_{lower}	n_{upper}	M	β_0	β_1
1	100	10	20	400	1	1
2	100	10	20	400	2	4
3	100	10	20	40	1	1
4	100	10	20	40	2	4
5	40	10	20	400	1	1
6	40	10	20	400	2	4
7	40	10	20	40	1	1
8	40	10	20	40	2	4
9	100	2	9	100	1	1
10	100	2	9	100	2	4
11	100	2	9	40	1	1
12	100	2	9	40	2	4
13	40	2	9	100	1	1
14	40	2	9	100	2	4
15	40	2	9	40	1	1
16	40	2	9	40	2	4

The model fitting performances (measured in DICs) of three Bayesian models are shown in Figure 9 and Figure 10. In all cases, Fixed model yielded the largest DIC. Random 2 model yielded the lowest DIC values, indicating that Random 2 model has the best model fitting. Moreover, when β_0 and β_1 are large, Random 1 and Random 2 models significantly outperform the Fixed model. While when β_0 and β_1 are small, the DICs of three models become similar.

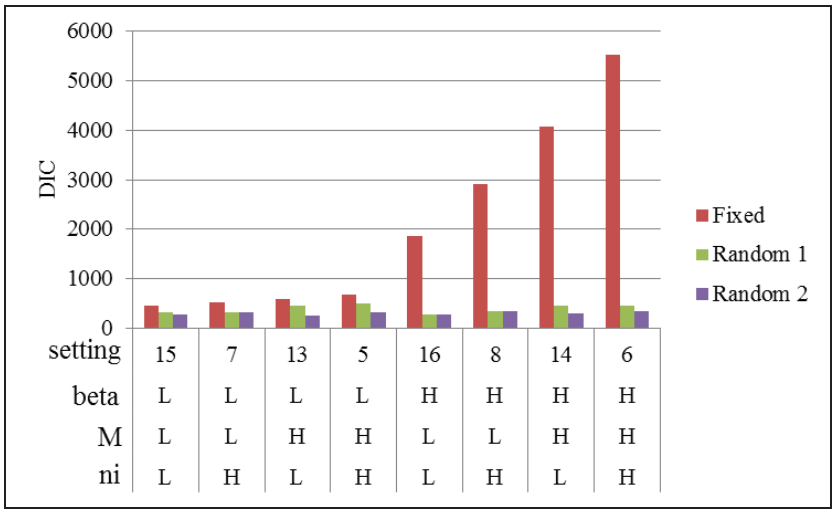

Figure 9 Model fitting DICs for data setting 1 (I = 40)

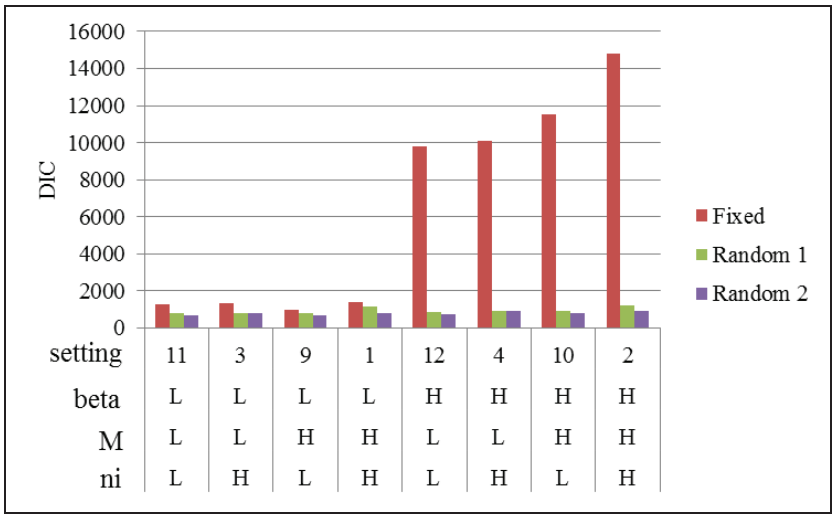

Figure 10 Model fitting DICs for data setting 1 (I=100)

The standard deviations of Poisson rate ratios in three Bayesian models are shown in Figure 11 and Figure 12. In all cases, Fixed model yielded the smallest standard deviation. Random 2 model yielded the largest standard deviation. The results completely are agreed with the theoretical proof (in Section 3.3.1) of a larger variance when exposure is random than when exposure is treated as fixed.

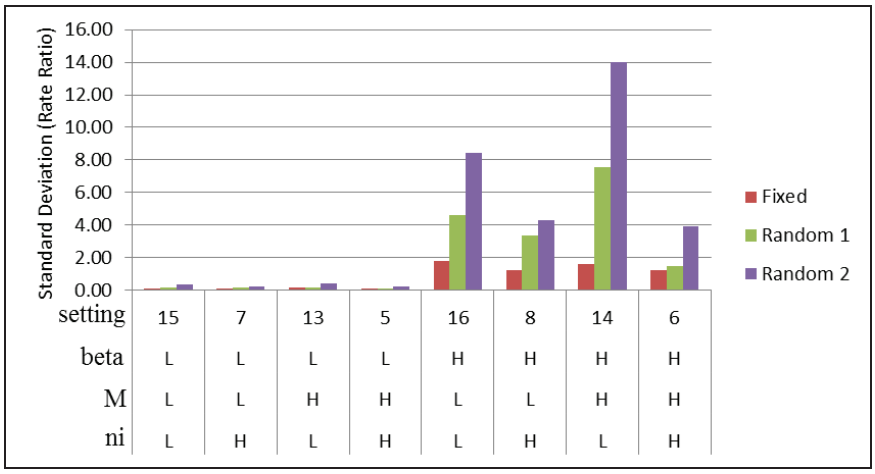

Figure 11 Standard deviations of Poisson rate ratios for data setting 1 (I=40)

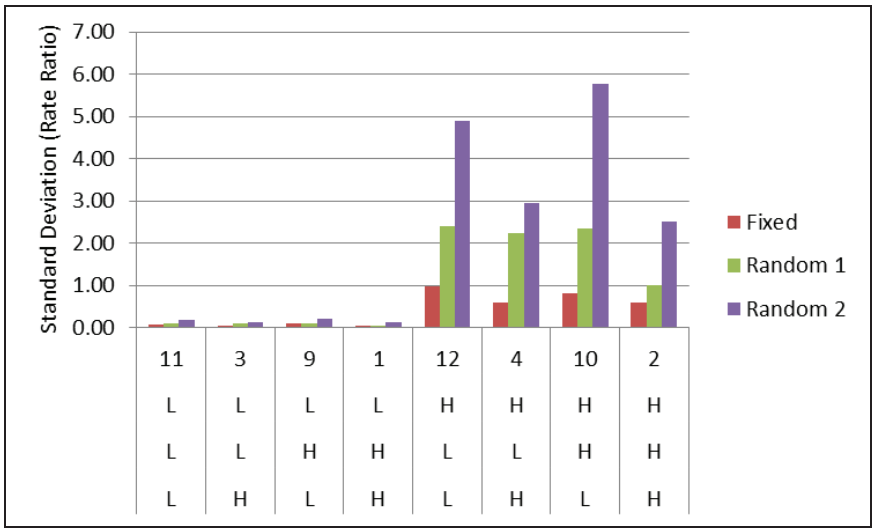

Figure 12 Standard deviations of Poisson rate ratios for data setting 1 (I=100)

3.4.2.2 Model fitting performances for data setting 2

The parameter settings for data setting 2 are listed in Table 10. A total of 12 settings are implemented to cover 4 levels of regression coefficients (HH=very high, H=high, L=low, LL=very low); High (H) and Low (L) number of base exposure segments for each subject; High (H) and Low (L) values of known entire exposure (G_i's). Moreover, we set I=100, M=400 and $\mu = 0.7$.

Table 10 Parameter settings for data setting 2 (for DIC)

settting	n_{lower}	n_{upper}	G_{upper}	G_{upper}	β_0	β_1
1	10	20	300	400	1	3
2	10	20	300	400	1	1
3	10	20	300	400	-1	0.6
4	10	20	300	400	-1	-0.6
5	2	10	300	400	1	3
6	2	10	300	400	1	1
7	2	10	300	400	-1	0.6
8	2	10	300	400	-1	-0.6
9	2	10	700	1000	1	3
10	2	10	700	1000	1	1
11	2	10	700	1000	-1	0.6
12	2	10	700	1000	-1	-0.6

The model fitting performances (measured DICs) of three Bayesian models are shown in Figure 13. In all cases, Fixed model yielded the largest DIC. Random 2 model yielded the lowest DICs, indicating that Random 2 model has the best model fitting. When β_0 and β_1 are large, Random 1 and Random 2 models significantly outperform the Fixed exposure model; in other cases, the DICs of three models are similar.

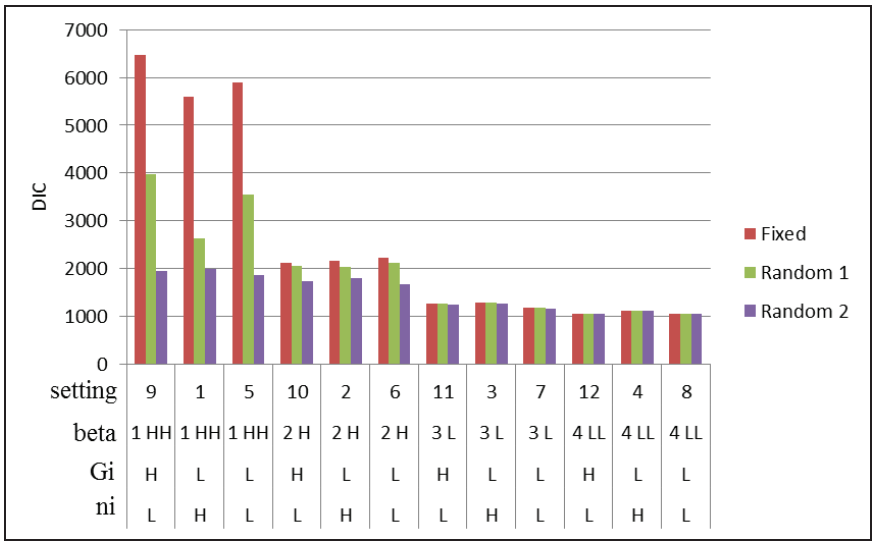

Figure 13 Model fitting DICs for data setting 2

The standard deviations of Poisson rate ratios of three Bayesian models are shown in Figure 14. In all cases, Fixed model yielded the smallest standard deviation. Random 2 model

yielded the largest standard deviation. The results are agreed with the theoretical proof of a larger variance when exposure is random than when exposure is treated as fixed.

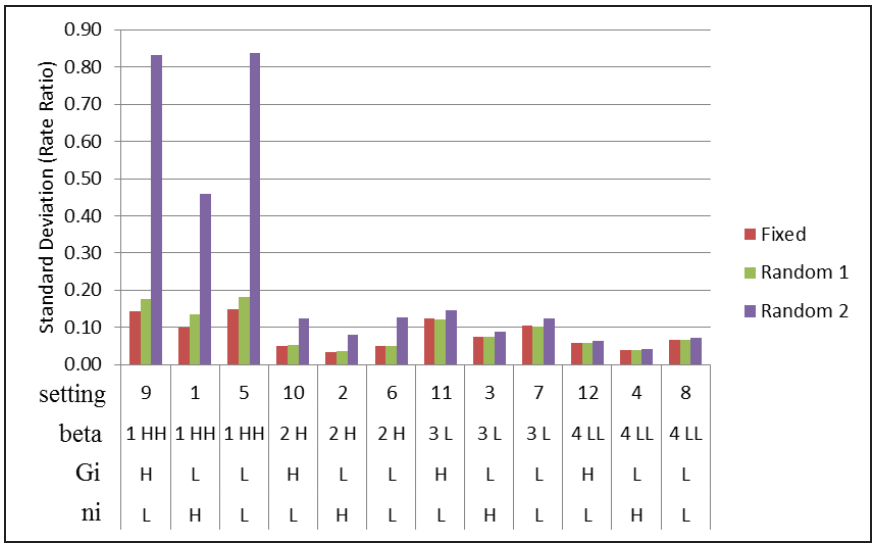

Figure 14 Standard deviations of Poisson rate ratios for data setting 2

3.4.3 Simulation studies for parameter coverage probabilities

In Section 3.4.3, I performed extensive simulation studies to investigate the parameter coverage probabilities of three Bayesian Models. For each parameter setting, T=1,000 simulation runs were performed. The empirical coverage probabilities were calculated based on the percentage of times that the 95% credible interval covers the pre-defined values of β_1 in the parameter setting. Because β_1 is associated with our primary research interest - the Poisson rate ratio, I primarily focused on the coverage probabilities of β_1. The simulation results for β_0 also indicated (detailed results not shown), although the coverage probabilities of β_0 are not consistently high, Random 2 model still outperformed the Fixed model and Random 1 model.

3.4.3.1 Coverage probabilities under data setting 1

A total of 72 parameter settings were studied in order to cover a broad range of parameter space as follows.

- Three levels of (β_0, β_1): High=(2,4), Median=(1,1), Low=(-1,-0.6)
- Three levels of μ: High=4, Median=2, Low=0.7
- Two levels of M: High=100, Low=40
- Two levels of (n_{lower}, n_{upper}): High= (10,20), Low=(2,9)

- Two levels of I: High=100, Low=40

The coverage probabilities of three Bayesian models are shown in Figure 15. Vertical reference lines are added to indicate different levels of β and μ. In all parameter settings, Fixed model has the lowest coverage probability. Random 2 model yielded has the highest coverage probabilities, and consistently keep 95% nominal level. This indicates that Random 2 model has best coverage probability performance for β_1. The coverage probability of Random 1 model is lower than nominal level, but is still higher than Fixed model. In contrast, the parameter coverage probability in Fixed model significantly decreases as the values of β_0 and β_1 increase. Moreover, parameter coverage probability of Random 1 model is affected by M and n_i. Random 1 model has higher coverage probabilities when M and range of n_i are relatively large.

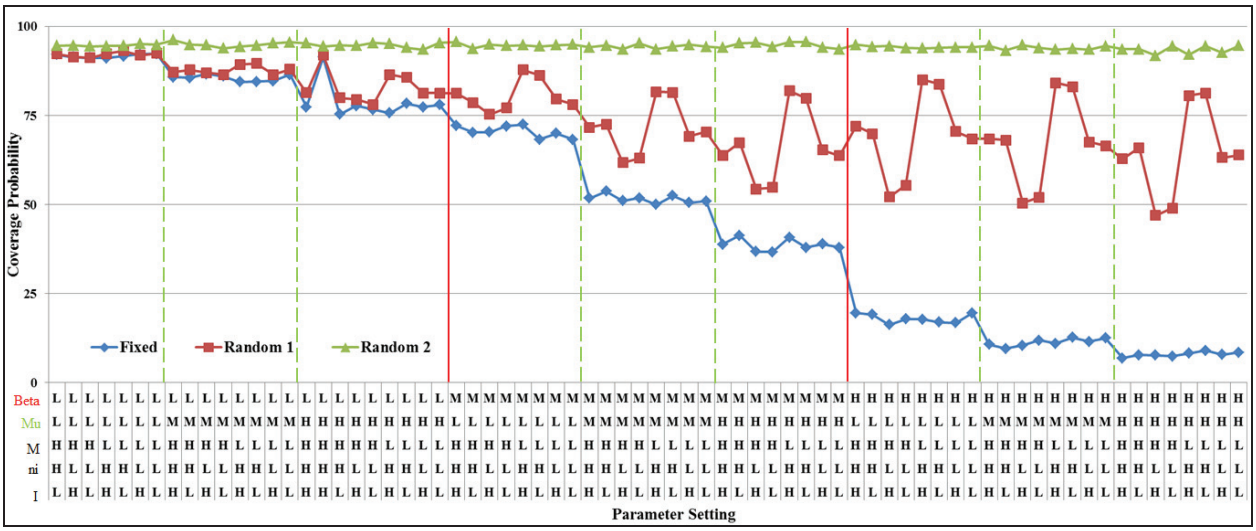

Figure 15 Coverage probabilities for regression parameter beta1 (data setting 1)

3.4.3.2 Coverage probabilities under data setting 2.

A total of 18 parameter settings were studied. In each setting I defined (I, M, G_{lower}, G_{upper}) = (40, 100, 300, 400). Other tuning parameters are set as follows.

- Three levels of (β_0, β_1): High=(2,4), Median=(1,1), Low=(-1,-0.6)
- Three levels of μ: High=4, Median=2, Low=0.7
- Two levels of (n_{lower}, n_{upper}): High= (10,20), Low=(2,9)

The coverage probabilities of three Bayesian models are shown in Figure 16. Vertical reference lines are added to indicate different levels of β and n_i. In all cases, Fixed model has

the lowest coverage probabilities. Random 2 model yielded has the highest coverage probability for β_1 . Moreover, the coverage probability of Fixed model and Random 1 model significantly decrease as the values β_0 and β_1 are increase. Moreover, the coverage probability of Random 2 model is affected by the range of n_i . Random 2 model has higher coverage probabilities when range of n_i are relatively low.

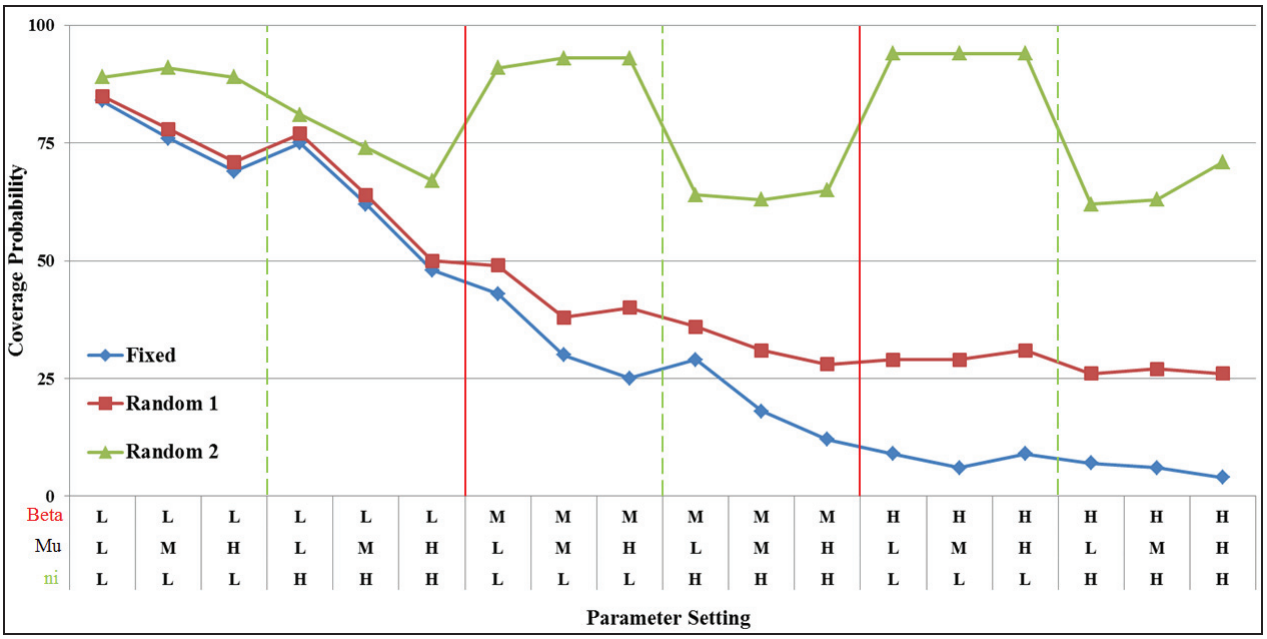

Figure 16 Coverage probabilities for regression parameter beta1 (data setting 2)

3.5 Implementation of Random Exposure Method on CPA Data

3.5.1 Introduction of texting-related VM subtask data in CPA study

In Section 3.5, I implemented the random exposure models to the data from the CPA study (Fitch *et al.* 2013). The primary goals of CPA study were to evaluate whether cell phone use activities during driving will increase driving risk. The study concluded that: (1) cell phone talking in any cell phone types does not significantly increase driving risk; (2) call-related Visual-Manual (VM) distraction, text-related VM distraction, and the overall VM distraction (combining call and text-related VM distractions), are associated with an increased driving risk; (3) overall cell phone use with hand-held cell phone type is associated with an increased driving risk, but the portable-hands-free and integrated-hands-free devices types do not significantly increase driving risk.

Different types of cell phone use subtasks were investigated in CPA study. Some examples are overall cell phone use, cell phone talking, and Visual-Manual (VM) cell phone use (i.e. non-talking cell phone use). All these cell phone activities are involved with multiple cellphone subtasks. Moreover, the directions (incoming, outgoing, or unknown direction) of call or text-messaging are considered. CPA study used stratified samples to estimate the durations of base exposure segments of each subtask, and applied weighted average to compute the total durations of the combined exposures. In addition, the entire exposure, i.e. entire driving time, is known for each driver. Therefore, the data structures in CPA study are similar to data setting 2 discussed in Section 3.3.3.2, but can have more than two types of exposures of interest.

In Section 3.5, I implemented the random exposure method for a less complicated exposure: text-related VM cell phone use. It fits the context of data setting 2. Preliminary analysis showed that the durations of the base exposure segment for incoming/outgoing/ unknown directions of text-related VM subtasks are similar. So cell phone use direction was not considered as a confounding factor in this section.

In CPA study, there are 116 drivers with texting activities recorded in the cellphone use log. Among them, 12 drivers do not have any texting at all. It is not difficult to modify the models for data setting 2 to incorporate these 12 drivers without E_i exposure (only with G_i exposure). But, in order to maintain the consistency of the simulation study and the real data analysis, I excluded these 12 drivers, retaining 104 drivers with non-zero text messaging activities (with E_i and $G_i - E_i$ exposures).

The summary statistics for the data are shown in Table 11. The text-related VM, number of texting, number of SCEs, and SCE rate are all right skewed, suggesting that some drivers have large numbers of text messaging, SCE events, and large SCE rate. The standard deviation of SCE rate is 4 times of the mean SCE rate, indicating a potential variance over-dispersion problem in Poisson regression model. Rather using Negative Binomial regression, I used Poisson regression with subject random effect to account for extra variation.

Table 11 Summary for text-related visual-manual tasks

Variable	N	Sum	Mean	Std. Dev.	Median	Min.	Max.
Total driving (hr)	104	184620.6	1775.2	809.1	1725.9	79.0	5417.2
Text-related VM (hr)	104	4435.9	42.7	108.8	13.3	0.8	993.3
Just driving (hr)	104	180184.7	1732.5	757.8	1702.7	76.7	4423.9
Number of texting (count)	104	5819	56	142.8	17.5	1	1303
Number of SCEs (count)	208	192	0.9	1.8	0	0	11
SCE rate (per 1000 hrs)	208	264.2	1.3	4.7	0	0.0	54.7

The text-related VM subtask duration (i.e. base exposure segment) is estimated using video reduction samples. After deleting outliers using the +/- 3 standard deviation rule, there are M=461 individual text-related VM duration (Figure 6). The goodness-of-fit test (Table 8) failed to reject the null hypothesis that the sample data follow an Exponential distribution. Using video reduction samples, we obtained $\hat{\mu} = 0.76$.

3.5.2 Model implementation results

We implemented 6 Bayesian models to the data.

- (1-3) Fixed model, Random 1 model and Random 2 model as discussed in Section 3.3.3.
- (4-6) Based on (1-3), add a driver-level random intercept b_i, with prior as

  $b_i \mid \tau_b \sim N(0, precision = \tau_b)$ and $\tau_b \sim Gamma(shape = 0.01, rate = 0.01)$.

Model fitting and estimation results are shown in Table 12. The model fitting (DIC values), parameter estimates and 95% credible intervals are similar in Model 1 to Model 3 and in Model 4 to Model 6. The similarity is not surprising. Because in simulation studies for data setting 2 we already observed that, when the regression parameters are small, Model 1 to Model 3 have similar results. Therefore, in the situations with small regression coefficients, Fixed model is an alternative to Random 1 and Random 2 models.

Moreover, the models with driver random effects have significantly lower DICs, indicating that the driver-level random effects successfully account for extra data variation. Fixed model and Random 1 model have 95% credible interval above 1. Random 2 model has 95% credible interval slightly covered the value of 1. On the other hand, all 3 counterparts with random intercept showed that text-related VM tasks are associated with increased driving risk.

Table 12 Model estimation results of text-related VM data

Model	DIC	Parameter	Mean	Std. dev.	95% LCL	95% UCL
Fixed	470.4	β_0	-6.89	0.07	-7.04	-6.75
		β_1	0.63	0.36	-0.13	1.28
		Risk RR	2.00	0.69	0.88	3.59
Random 1	470.3	β_0	-6.90	0.07	-7.04	-6.76
		β_1	0.64	0.35	-0.11	1.27
		Risk RR	2.01	0.68	0.90	3.56
Random 2	470.3	β_0	-6.89	0.07	-7.04	-6.75
		β_1	0.64	0.35	-0.08	1.28
		Risk RR	2.02	0.69	0.92	3.59
Fixed + Random intercept	388.9	β_0	-7.24	0.14	-7.54	-6.98
		β_1	0.75	0.36	0.01	1.40
		Risk RR	2.24	0.79	1.01	4.05
Random 1 + Random intercept	388.9	β_0	-7.25	0.14	-7.55	-6.99
		β_1	0.75	0.36	0.01	1.41
		Risk RR	2.25	0.79	1.01	4.08
Random 2 + Random intercept	389.1	β_0	-7.25	0.14	-7.55	-6.98
		β_1	0.74	0.37	-0.03	1.40
		Risk RR	2.24	0.80	0.97	4.07

3.6 Conclusion and Discussion

In this chapter, I proposed a random exposure method for Poisson regression model to accommodate the situations where the exposure needs to be estimated. Section 3.3.1 proved that the MLEs are unbiased estimator in Poisson model and Poisson regression model, while the variance of MLEs are larger when exposure is random than when exposure is treated as fixed. Section 3.3.2 investigated the situations where base exposure segments follow an Exponential distribution. In Section 3.3.3, two Gamma-distributed informative priors were elicited under Bayesian framework, and the models for 2 different data settings were proposed.

In Section 3.4, we conducted simulation studies to investigate the model fitting performance and parameter coverage probabilities proposed models. The results indicate that Random 2 model outperform Fixed exposure model and Random 1 model in all cases, especially when regression coefficients are large.

In Section 3.5, we implemented the random exposure method to the text-related VM subtasks data from CPA study. The Models with random intercept have lowest DICs, and yield significant 95% credible intervals not covering the null value of 1, suggesting that text-related VM subtasks are associated with increased driving risk. On the other hand, the model fitting and estimation results are similar among Fixed, Random 1 and Random 2 models.

As shown in simulation studies and CPA real data study, when Poisson regression coefficients are small, the model fitting performance and parameter coverage probability of Fixed exposure model is similar to Random 1 and Random 2 models. In these cases, fixed exposure model can serve as an alternative to random exposure model. This provided a support for the validity of the results in CPA report (Fitch *et al.* 2013), as the regression coefficients in CPA are relatively small. However, the random exposure method (especially Random 2 model) is a more robust method on broader application area, and shall be used when the exposure information is unknown and needs to be estimated.

Chapter 4 Performance Analysis on Meta-analysis Models

4.1 Introduction

Meta-analysis makes statistical inference on a summary effect size by summarizing comparable results of multiple independent studies. By including multiple studies, meta-analysis tends to achieve a higher statistical power than any study included (Hedges and Pigott 2001, Borenstein *et al.* 2009).

To date, meta-analysis has primarily focused on the parameter estimation and hypothesis test of the summary effect size and population heterogeneity (Laird and Mosteller 1990, Normand 1999, Sutton and Higgins 2008). The statistical power of hypothesis test on summary effect size has not been thoroughly investigated. Theoretical derivation and simulation-based empirical approach are two major approaches to calculate power. There has been limited number of studies on the theoretical power of hypothesis test on summary effect size (Hedges and Pigott 2001, 2004, Valentine *et al.* 2010, Thorlund and Mills 2012). Hedges and Pigott (2001) provided a method to calculate the theoretical power of hypothesis test on summary effect size in meta-analysis. Hedges and Pigott (2004) extended this method to meta-analysis with moderators (i.e. meta-regression). Valentine *et al.* (2010) based on the same method and outlined the frameworks for prospective and retrospective power analyses. Thorlund and Mills (2012) proposed a method to calculate theoretical power in network meta-analysis. In these studies, the theoretical power were derived based on normal distribution theory, and usually the original scales of effect size measures (e.g. odds ratio, correlation) need to be transformed into some measures that have approximately normal distribution. Theoretical methods are fast in computation and have a consistent performance when study size and sample size are sufficient.

Monte-Carlo simulation study is another widely used method that can calculate statistical power. In meta-analysis research area, simulation studies have been widely used to study the performances (e.g. bias, variance, MSE, coverage probability, relative efficiency) of different estimators of the population heterogeneity parameter (Sidik and Jonkman 2005, Viechtbauer 2005, Sidik and Jonkman 2007), as well as the empirical powers and type-I errors of hypothesis tests for population heterogeneity (Jackson 2006, Viechtbauer 2007). Simulation-based empirical approach does not depend on distributional assumption when calculating power and type-I error. Meanwhile, they can be directly applied to meta-analysis models that work on original scale of

effect sizes (e.g. Mantel-Haenszel and Peto methods) without a need to transform to normal. Moreover, empirical approach can handle more complicated meta-analysis models (e.g. meta-analysis on survival data) where theoretical power derivation may not be easily achieved. Only a limited number of studies have used simulation to study the power, type-I error and coverage probability on different estimators of summary effect size (Hartung and Knapp 2001b, Knapp and Hartung 2003, Bradburn *et al.* 2007).

Fixed-effects and random-effects models are two major statistical inference approaches in meta-analysis (Hedges and Vevea 1998, Borenstein *et al.* 2009). Fixed-effects models treat effect size parameters as fixed and unknown constants to be estimated, and typically have a homogeneity assumption that all individual effect sizes are equal. The statistical inference in fixed-effects models is conditional on studies that are observed; the inference results cannot be generalized to unobserved studies (Hedges and Vevea 1998). In contrast, random-effects models treat effect size parameters as a random sample from a population of parameters with a certain distribution. The mean and variance of the distribution are of interest. The inference results of random-effects models can be generalized to the population of studies with similar characteristics (Hedges and Vevea 1998). Because of the fundamental difference, the statistical powers of fixed- and random-effects models are not directly comparable.

In this chapter, we used simulation studies to investigate the statistical powers, type-I errors and coverage probabilities on summary effect size estimators in meta-analysis. Three fixed-effects models and 14 (7 non-KH-adjusted models and 7 KH-adjusted versions) random-effects models were investigated. We measured the model performances as a function of study size, sample size, effect size and population heterogeneity. Empirical powers from simulation studies are compared with theoretical powers, for which I modified the calculation method to increase precision.

The chapter is structured as follows. Section 4.2 introduced the modern meta-analysis models and the simulation-based performance analysis method. Section 4.3 discussed the simulation procedures and results. Section 4.4 summarize the findings and conclude this chapter.

4.2 Methods

This chapter focused on binary data and the log-odds ratio effect size. Binary data type is popular in medicine, epidemiology, and social science fields. The results of study i can be described in a

2×2 contingency table as shown in Table 13. Section 4.2.1 and 4.2.2 introduced modern classical (frequentist's) approach fixed- and random-effects meta-analysis models. Section 4.2.3 and 4.2.4 outlined the simulation-based performance analysis method for summary effect size. Section 4.2.5 introduced the adjustments for zero-cell count issue in binary data as well as the Knapp-Hartung adjustment for uncertainty of heterogeneity parameter in random-effects models.

Table 13 Contingency table for binary data

	Event	non-event	row total
Treatment group	a_i	b_i	n_{1i}
Control group	c_i	d_i	n_{2i}
column total	m_{1i}	m_{2i}	N_i

4.2.1 Fixed-effects models

Suppose $Y_1,\ldots,Y_k$ are the effect size variables from *k* independent studies. A fixed-effects meta-analysis model is given by $Y_i=\theta_i+\varepsilon_i$ (i =1…k), where θ_i is the effect size parameter for *i*-th study; ε_i is the within-study variance, with $\varepsilon_i \sim N(0,\sigma_i^2)$. In convention, σ_i^2 's are estimated by sample variance (i.e. $\hat{\sigma}_i^2=s_i^2$). Fixed-effects models treat θ_i 's as fixed and unknown parameters, and usually have a underlying assumption that all θ_i 's are equal. Thus, a fixed-effects model usually can be written as $Y_i=\theta+\varepsilon_i$, where θ is the summary effect size parameter to be estimated.

Three widely-use classical fixed-effects models were studied in this chapter: inverse-variance weighted least squares estimator (FE) (Hedges and Olkin 1985, Hedges and Vevea 1998), Mantel-Haenszel (MH) estimator (Mantel and Haenszel 1959) and Peto (PT) estimator (Yusuf *et al.* 1985) (Table 14). Based on normal assumption, the α -level confidence interval for θ is given by $\hat{\theta}\pm z_{1-\alpha/2}\sqrt{Var\left(\hat{\theta}\right)}$.

Table 14 Classical fixed-effects models

$\hat{\theta}$	Formulation	Notation
FE	$\hat{\theta}_{FE} = \sum_{i=1}^{k} \hat{w}_i y_i \Big/ \sum_{i=1}^{k} \hat{w}_i$ $Var\left(\hat{\theta}_{FE}\right) = 1 \Big/ \sum_{i=1}^{k} \hat{w}_i$	$y_i = \log(a_i d_i / b_i c_i)$, $\hat{w}_i = 1/\hat{\sigma}_i^2$, $\hat{\sigma}_i^2 = s_i^2 = \left(1/a_i + 1/b_i + 1/c_i + 1/d_i\right)$
MH	$\hat{\theta}_{MH} = \log\left(\sum_{i=1}^{k} \hat{w}_i y_i \Big/ \sum_{i=1}^{k} \hat{w}_i\right)$ $Var\left(\hat{\theta}_{MH}\right) = \frac{1}{2}\left(\frac{E}{R^2} + \frac{F+G}{RS} + \frac{H}{S^2}\right)$	$\hat{w}_i = b_i c_i / N_i$, $y_i = a_i d_i / b_i c_i$, $E = \sum_i \left(a_i + d_i\right) a_i d_i \Big/ N_i^2$ $F = \sum_i \left(a_i + d_i\right) b_i c_i \Big/ N_i^2$, $G = \sum_i \left(b_i + c_i\right) a_i d_i \Big/ N_i^2$ $H = \sum_i \left(b_i + c_i\right) b_i c_i \Big/ N_i^2$, $R = \sum_i a_i d_i / N_i$ $S = \sum_i b_i c_i / N_i$
PT	$\hat{\theta}_{PT} = \log\left(\sum_{i=1}^{k} \left(O_i - E_i\right) \Big/ \sum_{i=1}^{k} I_i\right)$ $Var\left(\hat{\theta}_{PT}\right) = 1 \Big/ \sum_{i=1}^{k} I_i$	$O_i = a_i$, $E_i = (a_i + b_i)(a_i + c_i)/N_i$ $I_i = \left(a_i + b_i\right)\left(c_i + d_i\right)\left(a_i + c_i\right)\left(b_i + d_i\right) \Big/ \left[N_i^2 \left(N_i - 1\right)\right]$

4.2.2 Random-effects models

A random-effects meta-analysis model can be written as $Y_i = \theta_i + \varepsilon_i = \left(\theta + \delta_i\right) + \varepsilon_i$, with $\delta_i \sim N\left(0, \tau^2\right)$, and $\varepsilon_i \sim N\left(0, \sigma_i^2\right)$. Random-effects models consider individual effect size parameters as a random sample from a population of parameters with typically a Normal distribution, i.e. $\theta_i \sim \text{Normal}(\theta, \tau^2)$. The mean ($\theta$) and variance ($\tau^2$) of the distribution are of interest. The variance component τ^2 is also called population heterogeneity parameter, representing the degree of variation among individual studies.

So far, most classical random-effects models are based on inverse-variance weighted least squares approach (Hedges and Vevea 1998). The estimator for θ can be derived as $\hat{\theta}_{random} = \sum_{i=1}^{k} \hat{w}_i^* y_i \Big/ \sum_{i=1}^{k} \hat{w}_i^*$, with $\hat{w}_i^* = 1\Big/\left(\hat{\sigma}_i^2 + \hat{\tau}^2\right)$. Under normal assumption, the variance is given by $Var\left(\hat{\theta}_{random}\right) = 1 \Big/ \sum_{i=1}^{k} \hat{w}_i^*$, and the α-level confidence interval is $\hat{\theta}_{random} \pm z_{1-\alpha/2} \sqrt{Var\left(\hat{\theta}_{random}\right)}$.

The distinctions among random-effects models depend on different estimation methods for τ^2. Seven widely-used estimators for τ^2 were investigated in this chapter (Table 15). They are Hedges (HE) estimator (Hedges 1983), DerSimonian-Laird (DL) estimator (DerSimonian

and Laird 1986), Hunter-Schmidt (HS) estimator (Hunter and Schmidt 1990, Viechtbauer 2005), Sidik-Jonkman (SJ) estimator (Sidik and Jonkman 2005, Sidik and Jonkman 2007), maximum-likelihood (ML) estimator (DerSimonian and Laird 1986, Hardy and Thompson 1996, Thompson and Sharp 1999), restricted maximum-likelihood (REML) estimator (Harville 1977, Raudenbush and Bryk 1985, DerSimonian and Laird 1986), and empirical Bayes (EB) estimator (Morris 1983, Raudenbush and Bryk 1985, DerSimonian and Laird 1986, Thompson and Sharp 1999, Knapp and Hartung 2003). Among them, HE, DL, HS and SJ estimators have closed-form analytic solutions, while ML, REML, and EB estimators need iterative numerical solutions.

Table 15 Estimators for heterogeneity parameter in classical random-effects models

$\hat{\tau}^2$	Formulation	Notation
HE	$Max\left(0, \frac{1}{k-1}\sum_{i=1}^{k}\left(y_i-\bar{y}\right)^2-\frac{1}{k}\sum_{i=1}^{k}\hat{\sigma}_i^2\right)$	$y_i=\log\left(a_i d_i/b_i c_i\right),\ \bar{y}=\sum_{i=1}^{k} y_i/k$ $\hat{\sigma}_i^2=\left(1/a_i+1/b_i+1/c_i+1/d_i\right)$ (the same as follows)
DL	$Max\left(0, \frac{Q_{\hat{w}}^2-(k-1)}{\sum_{i=1}^{k}\hat{w}_i-\left(\sum_{i=1}^{k}\hat{w}_i^2\right)/\left(\sum_{i=1}^{k}\hat{w}_i\right)}\right)$	$Q_{\hat{w}}^2=\sum_{i=1}^{k}\hat{w}_i\left(y_i-\bar{y}_{\hat{w}}\right)^2,\ \hat{w}_i=1/\hat{\sigma}_i^2$ $\bar{y}_{\hat{w}}=\sum_{i=1}^{k}\hat{w}_i y_i/\sum_{i=1}^{k}\hat{w}_i$
HS	$Max\left(0, \frac{\sum_{i=1}^{k}\hat{w}_i\left(y_i-\bar{y}_{\hat{w}}\right)^2}{\sum_{i=1}^{k}\hat{w}_i}-\frac{\sum_{i=1}^{k}\hat{w}_i\hat{\sigma}_i^2}{\sum_{i=1}^{k}\hat{w}_i}\right)$	$\bar{y}_{\hat{w}}=\sum_{i=1}^{k}\hat{w}_i y_i/\sum_{i=1}^{k}\hat{w}_i$, $\hat{w}_i=1/\hat{\sigma}_i^2$
SJ	$\frac{1}{k-1}\sum_{i=1}^{k}\hat{v}_i^{-1}\left(y_i-\bar{y}_{\hat{v}}\right)^2$	$\hat{v}_i=\hat{r}_i+1\ ,\ \hat{r}_i=k\hat{\sigma}_i^2/\sum_{i=1}^{k}\left(y_i-\bar{y}\right)^2$ $\bar{y}_{\hat{v}}=\left(\sum_{i=1}^{k}\hat{v}_i^{-1}y_i\right)/\left(\sum_{i=1}^{k}\hat{v}_i^{-1}\right)$
ML	$Max\left(0, \frac{\sum_{i=1}^{k}\hat{w}_i^2\left[\left(y_i-\bar{y}_{\hat{w}}\right)^2-\hat{\sigma}_i^2\right]}{\sum_{i=1}^{k}\hat{w}_i^2}\right)$	$\bar{y}_{\hat{w}}=\frac{\sum_{i=1}^{k}\hat{w}_i y_i}{\sum_{i=1}^{k}\hat{w}_i},\ \hat{w}_i=1/\left(\hat{\sigma}_i^2+\hat{\tau}_{ML}^2\right)$
REML	$Max\left(0, \frac{\sum_{i=1}^{k}\hat{w}_i^2\left[\left(y_i-\bar{y}_{\hat{w}}\right)^2+1/\sum_{i=1}^{k}\hat{w}_i^2-\hat{\sigma}_i^2\right]}{\sum_{i=1}^{k}\hat{w}_i^2}\right)$	$\bar{y}_{\hat{w}}=\frac{\sum_{i=1}^{k}\hat{w}_i y_i}{\sum_{i=1}^{k}\hat{w}_i}$, $\hat{w}_i=1/\left(\hat{\sigma}_i^2+\hat{\tau}_{REML}^2\right)$
EB	$Max\left(0, \frac{\sum_{i=1}^{k}\hat{w}_i\left[k\left(y_i-\bar{y}_{\hat{w}}\right)^2/\left(k-1\right)-\hat{\sigma}_i^2\right]}{\sum_{i=1}^{k}\hat{w}_i}\right)$	$\bar{y}_{\hat{w}}=\frac{\sum_{i=1}^{k}\hat{w}_i y_i}{\sum_{i=1}^{k}\hat{w}_i},\ \hat{w}_i=1/\left(\hat{\sigma}_i^2+\hat{\tau}_{EB}^2\right)$

4.2.3 Simulation-based performance analysis method

In this chapter, we used Monte-Carlo simulation method to study the power, type-I error and coverage probability of meta-analysis models. The core idea of Monte-Carlo simulation is to use a large number of random samples of parameters or inputs to numerically evaluate the behavior of a complex system or process (Robert and Casella 2005). In the context of this chapter, the theoretical computations of power, type-I error and coverage probability involve the integration of complex probability process over a large set of parameters, where an analytical solution may not be easily derived. Monte-Carlo simulation replaces the analytical integration by making a large number of independent random samples from the complex system, and evaluating the numerical results from the samples. With a large number of simulation runs, the empirical outcome from Monte Carlo simulation will be consistent with theoretical outcome.

In our simulation-based method, seven parameters can be controlled and need pre-defined values. They are summary log-odds ratio (θ_{sim}), squire root of population heterogeneity parameter (τ_{sim}), study size (K), treatment-group and control-group sample sizes (Nt_i's and Nc_i's), and lower and upper bounds of control-group event probabilities (Pc_{lower} and Pc_{upper}). A simulation run consists of two steps:

(1) Data generation: for $i = 1...K$, sample $\theta_i \sim \text{Normal}\left(\theta_{sim}, \tau_{sim}^2\right)$, $Pc_i \sim Uniform(Pc_{lower}, Pc_{upper})$, and calculate $Pt_i = \dfrac{\exp(\theta_i) \times Pc_i}{1 + \left(\exp(\theta_i) - 1\right) \times Pc_i}$; then generate the binary data by sampling $a_i \sim \text{Binomial}\left(Nt_i, Pt_i\right)$, $c_i \sim \text{Binomial}\left(Nc_i, Pc_i\right)$, and calculated $b_i = Nt_i - a_i$, $d_i = Nc_i - c_i$;

(2) Model implementation: apply different meta-analysis models to the data, and record the model inference results.

The simulation run is repeated for T times, then the empirical power, type-I error, and coverage probability can be calculated as follows in Section 4.2.4.

4.2.4 Calculation of power, type-I error and coverage probability

4.2.4.1 Hypothesis tests in fixed- and random-effects models

The hypothesis tests in fixed- and random-effects models are different. In fixed-effect models, a two-sided hypothesis test for a significant (non-zero) θ can be written as $H_0 : \theta_{fixed} = 0$ vs. $H_a : \theta_{fixed} \neq 0$, where θ_{fixed} is the summary effect size parameter shared by all observed studies. With the homogeneity assumption and the treatment of θ_i's and θ_{fixed} as fixed and unknown parameters, the hypothesis test in a fixed-effects model is essentially testing whether $\theta_1 = ... = \theta_k = \theta_{fixed} = 0$ or $\theta_1 = ... = \theta_k = \theta_{fixed} \neq 0$. While in random-effects models, a hypothesis test of significant θ can be written as $H_0 : \theta_{random} = 0$ vs. $H_a : \theta_{random} \neq 0$, where the θ_{random} is mean of population of the parameters that includes the θ_i's. Therefore, the hypothesis test in a random-effects model is testing whether the mean of the distribution is equal to zero. Because of the conceptual difference, the inference results between the fixed-effects models and random-effects models are not directly comparable. Hence, in this chapter we presented and compared the performances of different models within each category respectively.

4.2.4.2 Calculation of power

Empirical power

With a significant (non-zero) true summary effect size $\theta_{\text{sim}} \neq 0$, the empirical statistical power for detecting the effect $\delta = \theta_{sim}$ can be calculated as

$$\text{Power}^{\text{empirical}} = \sum_{t=1}^{T} I\left(UCL < 0 \mid \text{LCL} > 0\right) \Big/ T ,$$

where LCL and UCL are lower and upper bounds of confidence interval for θ in a meta-analysis model, and I(.) is the indicator function.

Modified theoretical power

The methods for computing theoretical powers of classical meta-analysis models have been proposed in previous studies (Hedges and Pigott 2001, 2004). The theoretical power for a fixed-effects model is given by

$$\text{Power}_{\text{fixed}}^{\text{theoretical}} = 1 - \Phi\left(z_{\alpha/2} - \theta_{sim} \Big/ \sqrt{\bar{v}}\right) + \Phi\left(-z_{\alpha/2} - \theta_{sim} \Big/ \sqrt{\bar{v}}\right),$$

with $\bar{v} = 1 \Big/ \left(\sum_{i=1}^{k} \hat{w}_i\right)$ and $\hat{w}_i = 1 / s_i^2$. And the theoretical power for a random-effects model is

$$\text{Power}_{\text{random}}^{\text{theoretical}} = 1 - \Phi\left(z_{\alpha/2} - \theta_{sim} \Big/ \sqrt{\bar{v}^*}\right) + \Phi\left(-z_{\alpha/2} - \theta_{sim} \Big/ \sqrt{\bar{v}^*}\right),$$

with $\bar{v}^* = 1\Big/\left(\sum_{i=1}^{k} \hat{w}_i^*\right)$ and $\hat{w}_i^* = 1/\left(s_i^2 + \hat{\tau}^2\right)$.

Note that in original Hedges and Pigott (2001) paper, $\bar{v}$, $\bar{v}^*$ and τ^2 are arbitrarily defined with some subjective judgement. Although such user-defined values are easy to implement, they do not incorporate the information of specific data. In this chapter, I proposed a modified theoretical power calculation method by determining the values of $\bar{v}$, $\bar{v}^*$ and τ^2 using simulations.

When we need to perform a prospective performance analysis on meta-analysis models, we assigned $\hat{\tau}^2 = \hat{\tau}_{sim}^2$ to rule out of noise in the determination of $\hat{\tau}^2$. On the other hand, when we need to perform retrospective performance analysis based on existing meta-analysis results, we assign $\hat{\tau}^2$ as the DerSimonian-Laird estimate, which is consistent with the treatment in Hedges and Pigott (2001). Moreover, to provide good estimates of $\bar{v}$ and $\bar{v}^*$, we simulated 10,000 collections of meta-analysis data (each with *K* studies) with some specific specifications (more details in Section 4.3.1). Then we can calculate the value of $\bar{v}$ and $\bar{v}^*$ for each collection of data, and use the average of $\bar{v}$'s and $\bar{v}^*$'s as inputs to compute theoretical powers.

4.2.4.3 Calculation of empirical type-I error and coverage probability

When setting $\theta_{\text{sim}} = 0$, the empirical type-I error of a model can be calculated as

$$\text{Type-I Error}^{\text{empirical}} = \sum_{t=1}^{T} \text{I}\left(\text{UCL}<0 \mid \text{LCL}>0\right)\Big/T.$$

Moreover, the empirical coverage probability for θ_{sim} is given by

$$\text{Coverage Probability} = \sum_{t=1}^{T} \text{I}\left(\text{LCL} \le \theta_{\text{sim}} \le \text{UCL}\right)\Big/T.$$

4.2.5 Other adjustments

4.2.5.1 Adjustment for zero cell count

To deal with zero cell count problem in contingency tables, I add a constant of 0.5 to all the cells in the studies that have zero-cell counts. This is a widely used strategy for binary data (Agresti 2002, Bradburn *et al.* 2007, Borenstein *et al.* 2009, Viechtbauer 2010). By default, Mantel-Haenszel and Peto fixed-effects models do not require zero-count adjustment (except for some

extreme situations, see Viechtbauer (2014)). However, in order to be consistent with other estimators, we also applied zero cell count adjustment for MH and PT estimators. Moreover, in most cases, only one event cell count (a_i or c_i) may be zero in a study. If both cells are zero, the study is considered non-informative and is dropped off the meta-analysis as recommended by many studies (Bradburn *et al.* 2007, Higgins and Green 2008, Viechtbauer 2014).

4.2.5.2 Knapp-Hartung adjustment for uncertainty of heterogeneity parameter

In classical random-effects models, τ^2 is usually estimated first and then plugged in as a constant to calculate θ. This way the inference does not consider uncertainty in the estimation of τ^2, which may result in underestimation of variance and a narrower confidence interval for θ. To account for the uncertainty associated with estimating τ^2, Hartung and Knapp (Hartung and Knapp 2001b, a, Knapp and Hartung 2003) proposed a method to adjust confidence interval. The Knapp-Hartung (KH) adjusted confidence interval is given by

$$\hat{\theta}_{random} \pm t_{k-1,1-\alpha/2}\sqrt{\hat{q}},$$

where $\hat{q} = \sum_{i=1}^{k} \hat{w}_i^* \left(y_i - \hat{\theta}_{random}\right)^2 \Big/ \left[(k-1)\sum_{i=1}^{k} \hat{w}_i^*\right]$, $\hat{w}_i^* = 1\Big/\left(\hat{\tau}^2 + s_i^2\right)$, and $\hat{\theta}_{random}$ and $\hat{\tau}^2$ the point estimates from a specific random-effects model. KH-adjusted intervals are better in keeping nominal significance level (Hartung and Knapp 2001a, b). This chapter investigated both KH-adjusted and un-adjusted random-effect models.

4.3 Simulation Studies

4.3.1 Study procedures

The simulation procedures I used are similar to previous studies that investigated the properties of population heterogeneity parameter (e.g. (Sidik and Jonkman 2005, Bradburn *et al.* 2007, Sidik and Jonkman 2007)). In this chapter, I controlled some key parameters in simulations to cover the most influential (rather than maximum) parameter space. Specifically, I chose θ_{sim} = (0, 0.2, 0.4, 0.6, 0.8, 1.0), which are equivalent to odds ratios = (1.0, 1.2, 1.5, 1.8, 2.2, 2.7); square root of population heterogeneity parameter τ_{sim} = (0.001, 0.05, 0.1, 0.2, 0.3); study size *K* = (4, 8, 12, 16, 20, 24); group sample size *N* = (25, 50, 100, 150, 200, 250, 300), where we considered

equal-arm study and equal sample sizes in all studies (i.e. $Nt_i = Nc_i = N$); the lower and upper bounds of control-group event probabilities $\left(Pc_{lower}, Pc_{upper}\right) = \left(0.1, 0.2\right)$.

For each parameter setting, I conducted a simulation study with T=10,000 simulation runs. Three fixed-effects models (FE, MH, PT) and 14 random-effects models (HE, DL, HS, SJ, ML, REML, EB, and their KH-adjusted versions) were investigated. I implemented classical models in R software and used the Metafor R package (Viechtbauer 2010, 2014).

I calculated empirical powers, type-I errors, and coverage probabilities of summary effect size based on simulation results. I also derived theoretical powers based on Hedges and Pigott (2001), which required three pre-determined inputs: θ_{sim}, $\bar{v}$ (or $\bar{v}^*$), and $\hat{\tau}^2$. To rule out noise in the determination of $\hat{\tau}^2$, we assigned the optimal value for $\hat{\tau}^2$ as $\hat{\tau}^2 = \hat{\tau}^2_{sim}$, which could lead to an optimal theoretical power. To provide good estimates of $\bar{v}$ and $\bar{v}^*$, we simulated 10,000 collections of meta-analysis data (each with K studies) using the same method as the data generation step discussed above. Then we calculated the $\bar{v}$ and $\bar{v}^*$ for each collection of data, and used the average of $\bar{v}$'s and $\bar{v}^*$'s samples as inputs to compute theoretical powers.

4.3.2 Study results

4.3.2.1 Type-I error

The empirical type-I errors from simulation studies are shown in Figure 17 and Figure 18. A reference line of 0.05 was provided to indicate nominal level.

Figure 17 showed that, type-I errors will be inflated in the presence of study heterogeneity. A primary reason is that fixed-effects models assume no study heterogeneity and thus are not equipped to account for it. So when non-trivial between-study variation exists, the total variance (the combination of within-study variance and between-study variance) will be underestimated, which will lead to a tighter confidence interval and higher rejection rate. In addition, the inflation of Type-I error is positively associated with sample size. A possible reason is that, the increase in sample size reduces the within-study variance and thus the proportion of within-study variance in the total variance. Because fixed-effects models only consider within-study variance (rather than total variance) to make model inference, the variance underestimation will be more severe as the sample size increases.

Figure 18 shows that, KH-adjusted random-effects models perfectly keep nominal level of type-I error. On the other hand, the unadjusted models still have increased type-I errors, but the magnitude is much smaller than those happened in fixed-effects models. Moreover, type-I error is also positively associated with sample size, but is negatively associated with study size. The possible reason is that, the increased study size reduces the between-study variance and thus the proportion of between-study variance in the total variance. Therefore, the total variance underestimation is alleviated as the study size increases.

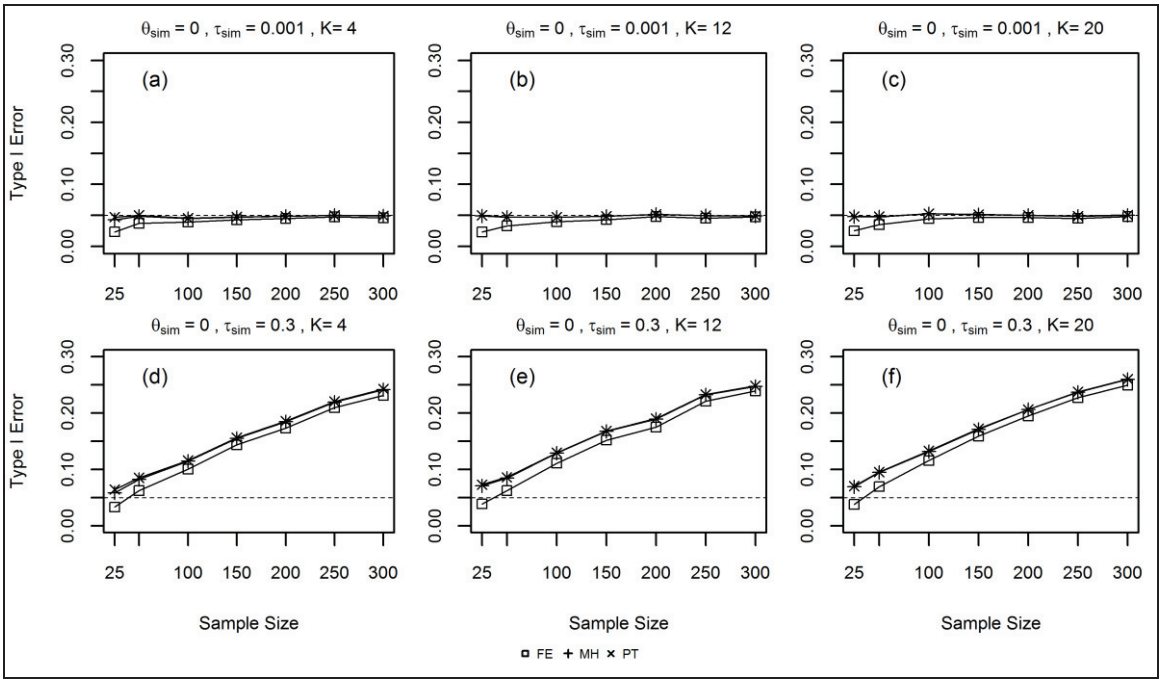

Figure 17 Type-I error vs. sample size in fixed-effects models

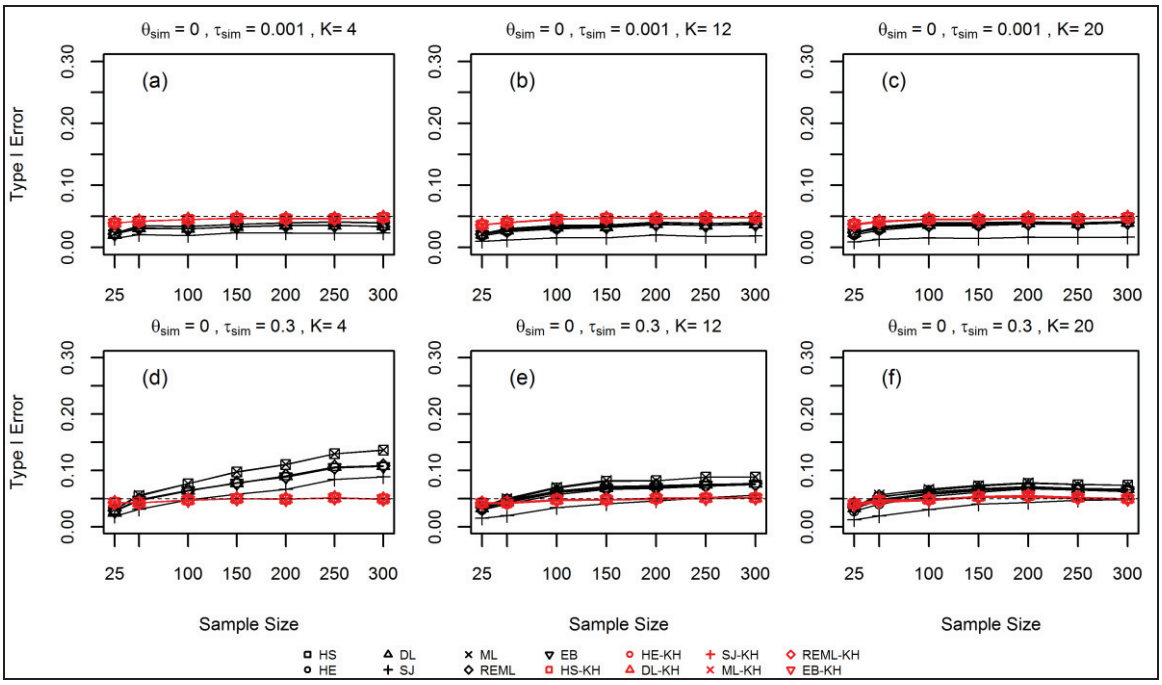

Figure 18 Type-I error vs. sample size in random-effects models

4.3.2.2 Statistical power

The empirical and theoretical powers are plotted against study size, sample size, and population heterogeneity and effect size to be detected. A reference line of 0.8 indicates a conventional threshold for high power. Because τ^2 is not a significant impact factor (shown later in this section) for statistical power, I only used one setting of τ^2 when plotting power against study size, sample size and effect size to be detected.

Power vs. study size

Figure 19 and Figure 20 showed that, study size is a significant impact factor for the power. Power increases as the study size increases. This impact is interacted with sample size and the effect size to be detected.

In fixed-effect models (Figure 19), when the effect to be detected is small ($\delta = \theta_{sim} = 0.2$), it is difficult to obtain a high power even with a large study size. When the effect to be detected is larger (0.4), it is easier to reach a high power with moderate to large study size. Moreover, when sample size is small, the MH and PT models have similar powers, and both are larger than the FE model. With moderate to large sample size, all fixed-effects models have similar powers. Similar relationships are observed in random-effects models. But in random-

effects models, a larger study size is needed to reach a high statistical power. Moreover, KH-adjusted models have lower powers than unadjusted versions, especially when study size is small.

In addition, theoretical powers are similar to empirical powers in fixed-effects models and unadjusted random-effects models, except when sample size is relatively small. However, the theoretical power and empirical powers of KH-adjusted models are different. A possible reason is that theoretical power calculation is based on normal distribution and z-statistic, while KH-adjusted models use t-statistic with study size as degree of freedom. When study size (and thus degree of freedom) is small, t-statistic deviates largely from z-statistic.

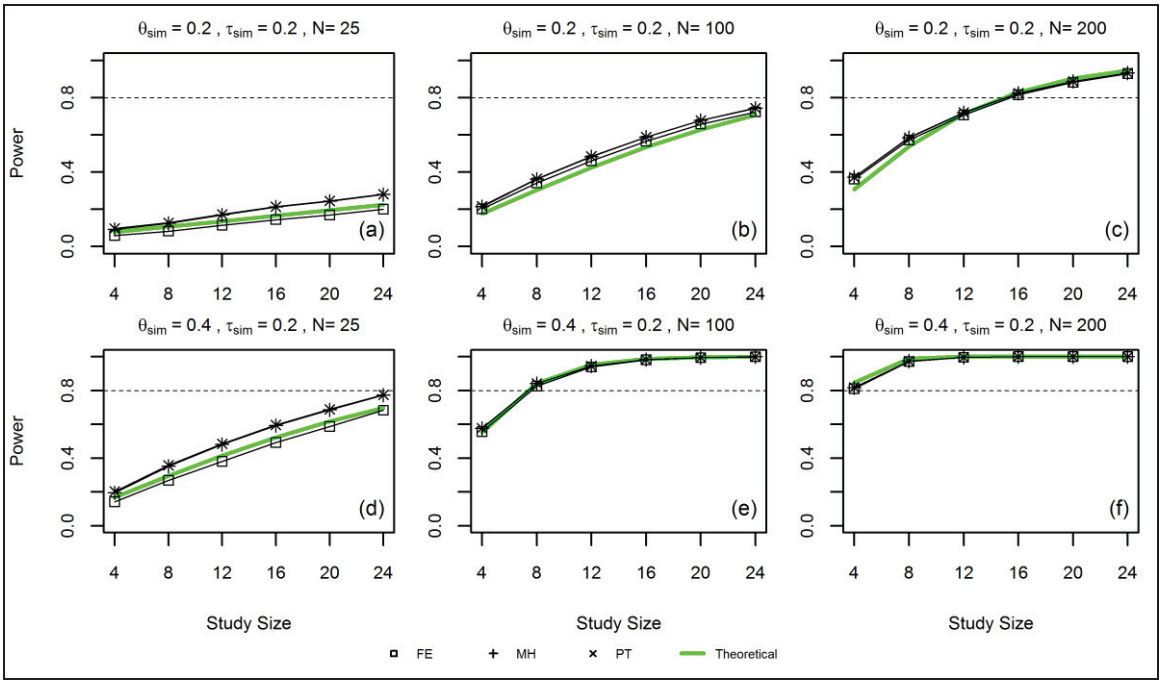

Figure 19 Power vs. study size in fixed-effects models

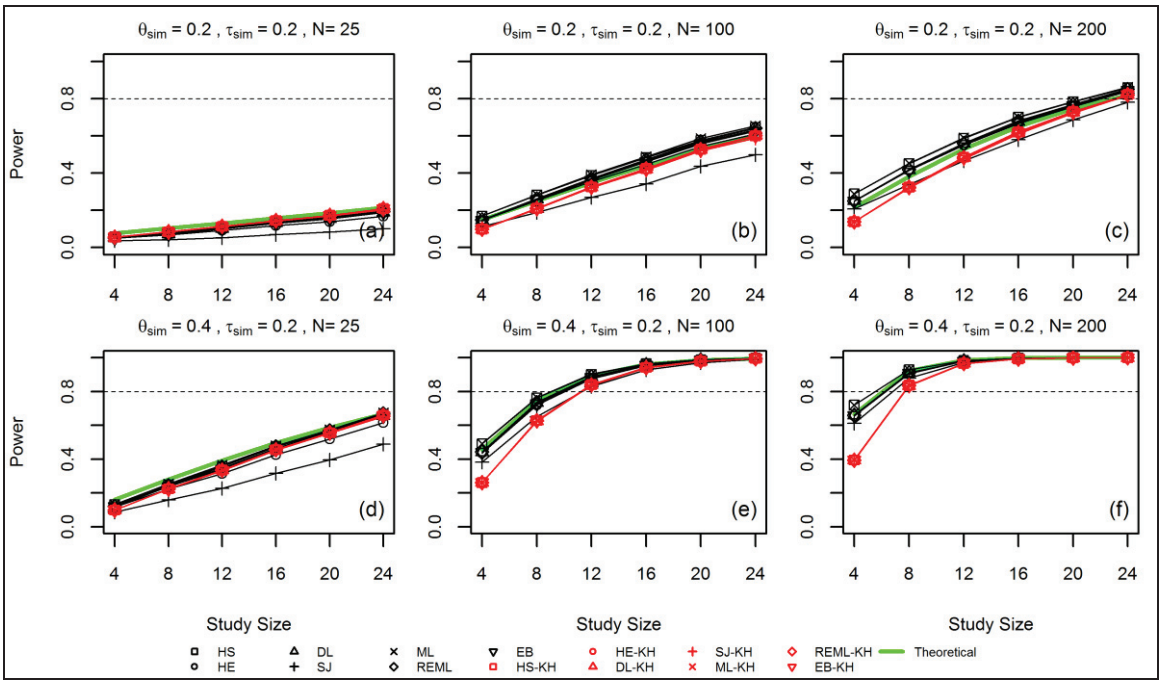

Figure 20 Power vs. study size in random-effects models

Power vs. sample size

Figure 21 and Figure 22 show that sample size is another significant impact factor for statistical power. The Power increases with increased sample size. This impact is interacted with study size and the effect to be detected. With small study size (K=4), a very large sample size is needed to reach a high power, especially when the effect to be detected is small. While for moderate (K=12) to large (K=20) study size, a sample size of 50 to 250 is sufficient to reach a high power.

Moreover, Figure 22 shows that the KH-adjusted random-effects models have significantly lower powers than the unadjusted versions when study size is small. But when we take into account fact that KH-adjusted models have much better Type-I error performance than unadjusted versions in small study size (e.g. Figure 18.d), KH-adjusted models are actually a better choice.

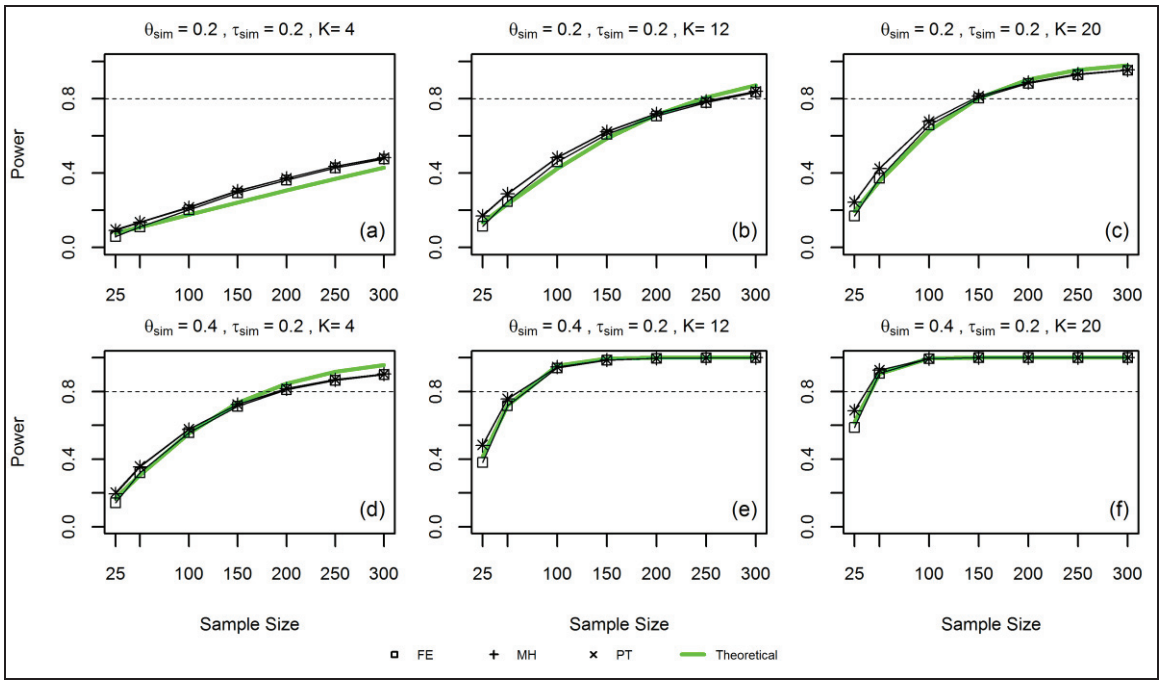

Figure 21 Power vs. sample size in fixed-effects models.

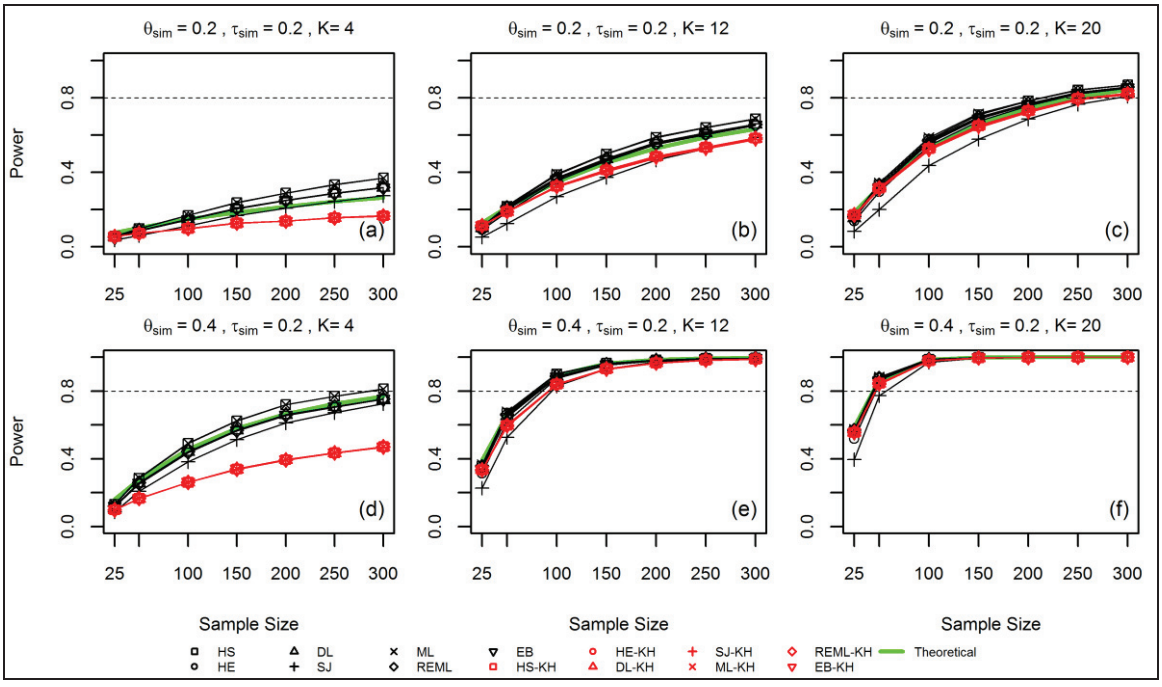

Figure 22 Power vs. sample size in random-effects models

Power vs. population heterogeneity

Figure 23 and Figure 24 show that τ^2 (and τ) is generally not a significant impact factor for statistical power. With range of τ from 0.001 to 0.3, there is less than 15% change of powers in most situations.

In fixed-effects models, the empirical powers have a small (but not significant) increasing trend as the degree of population heterogeneity increases. It is interesting to see that the theoretical powers remain constant in all ranges of heterogeneity. The reason for this is that the calculation of theoretical power in fixed-effects models does not incorporate input of τ^2. While in random-effects models, both empirical powers and theoretical powers have small (but not significant) decreasing trend as the population heterogeneity increases.

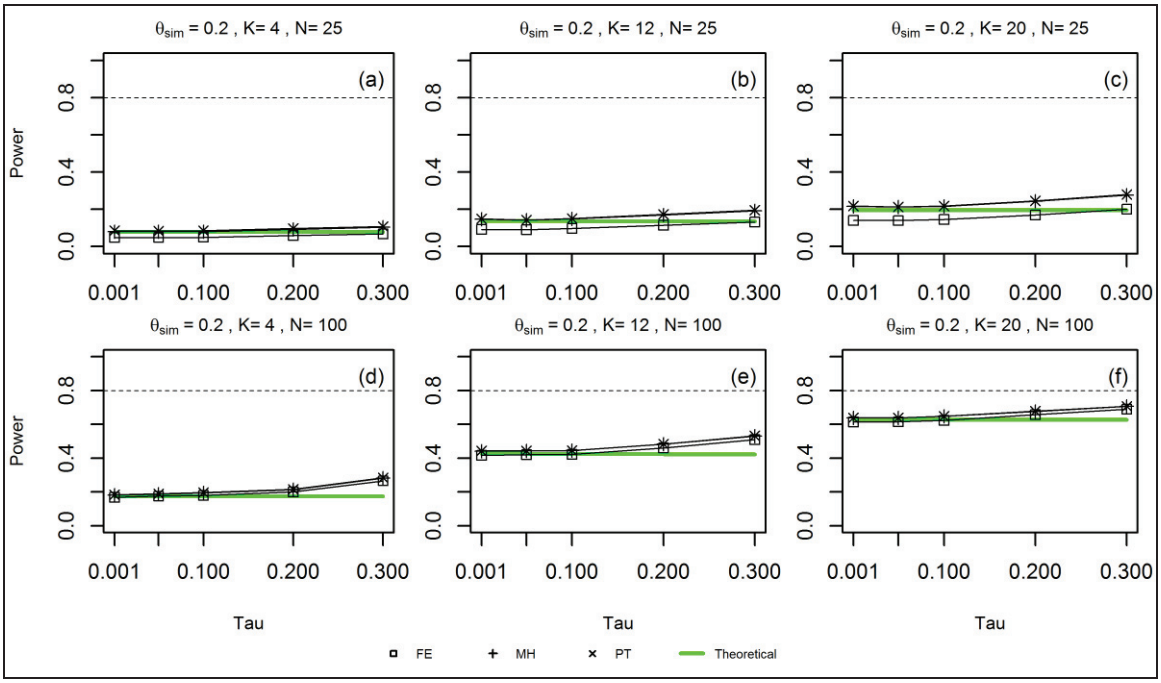

Figure 23 Power vs. population heterogeneity in fixed-effects models

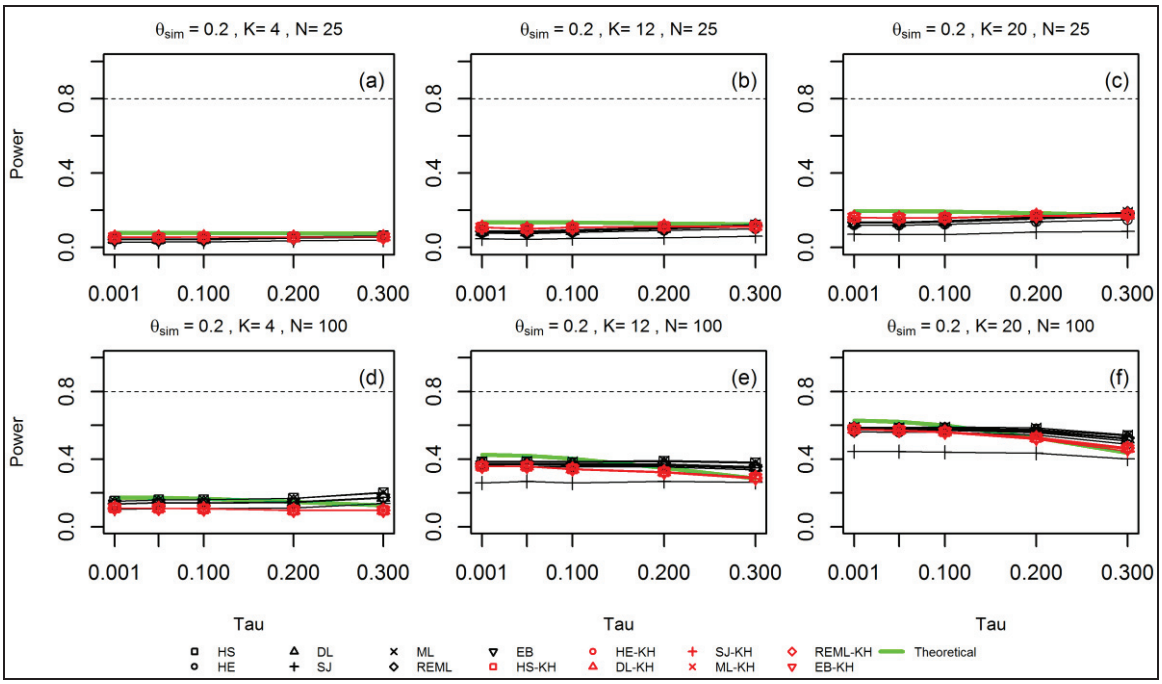

Figure 24 Power vs. population heterogeneity in random-effects models

Power vs. effect size to be detected

Figure 25 and Figure 26 show that, the statistical power increases as the summary effect to be detected increases. This relationship is positively associated with increased study size and sample size. When both study size and sample size are small, it is difficult to achieve a high power even if effect size to be detected is large.

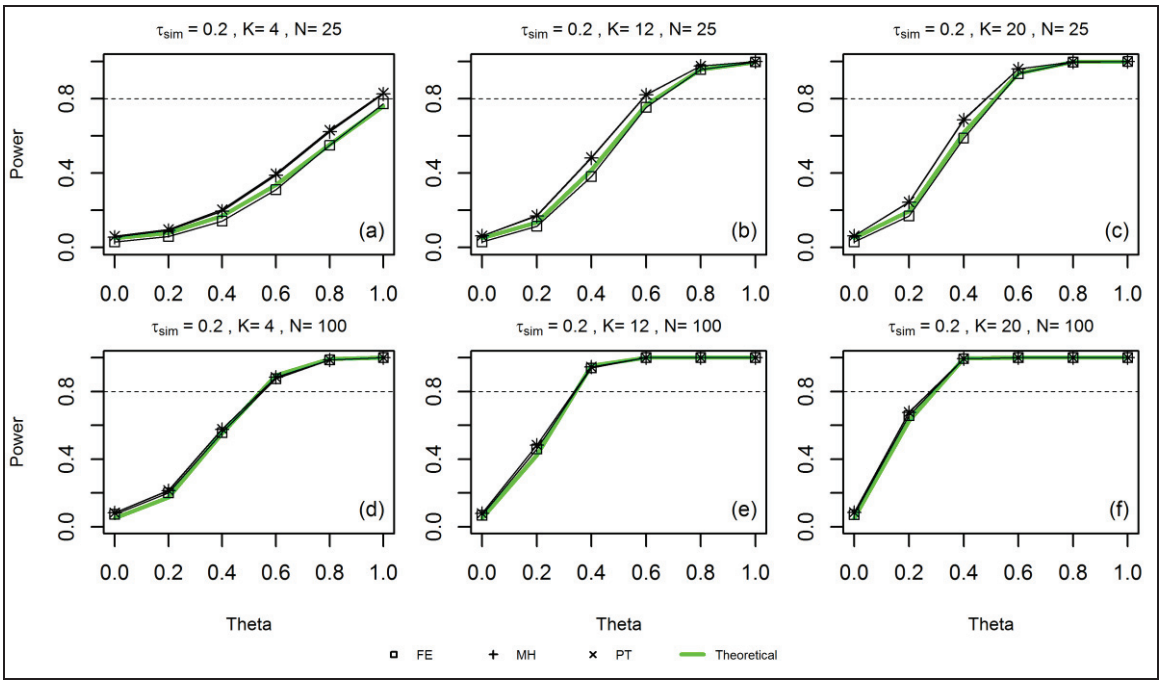

Figure 25 Power vs. effect size to be detected in fixed-effects models

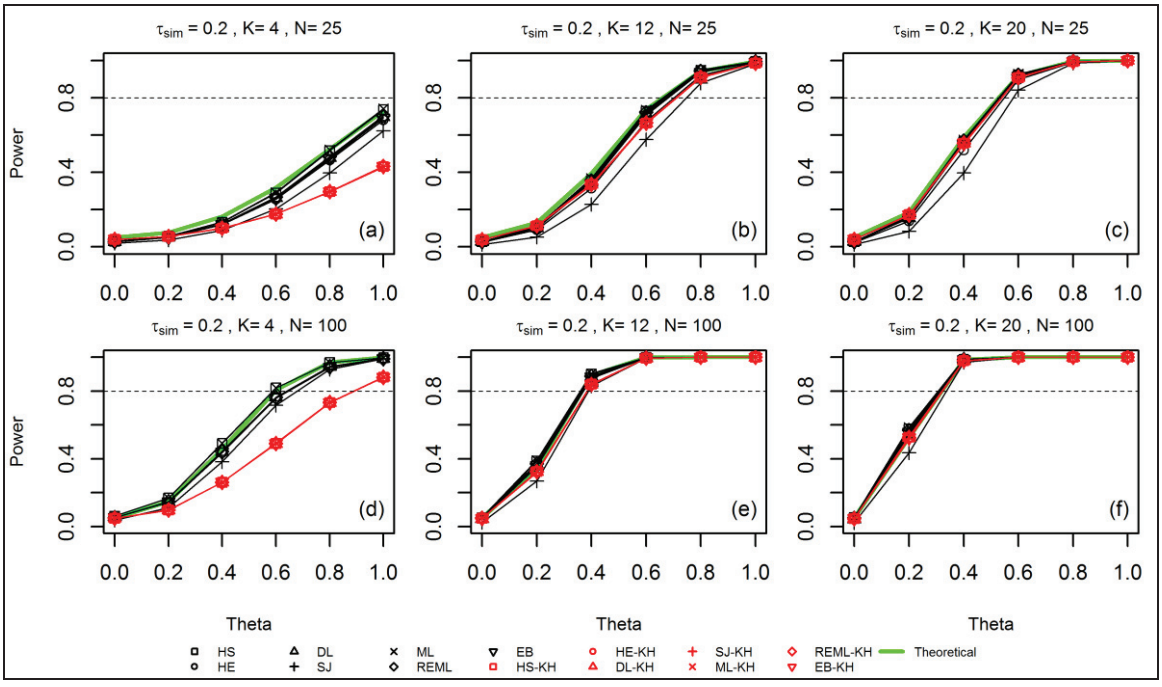

Figure 26 Power vs. effect size to be detected in random-effects models

4.3.2.3 Coverage probability

Figure 27 and Figure 28 show that, fixed- and random-effects models have dramatically different performances in keeping nominal 95% parameter coverage probabilities.

In fixed-effects model, the empirical coverage probabilities fall below nominal level as the degree of population heterogeneity increases. This decrease is more significant with increased sample size, but is not related to study size.

In random-effects models, KH-adjusted models maintain nominal coverage level very well. Unadjusted random-effects models have slightly reduced coverage as degree of heterogeneity increases. This reduction is more significant with increase sample size but with decreased study size.

The reasons for such performances are similar to the reasons discussed for Type-I error (Section 4.3.2.1). In fact, type-I error is equivalent to (1 – coverage probability) when the parameter setting θ_{sim} is set as zero.

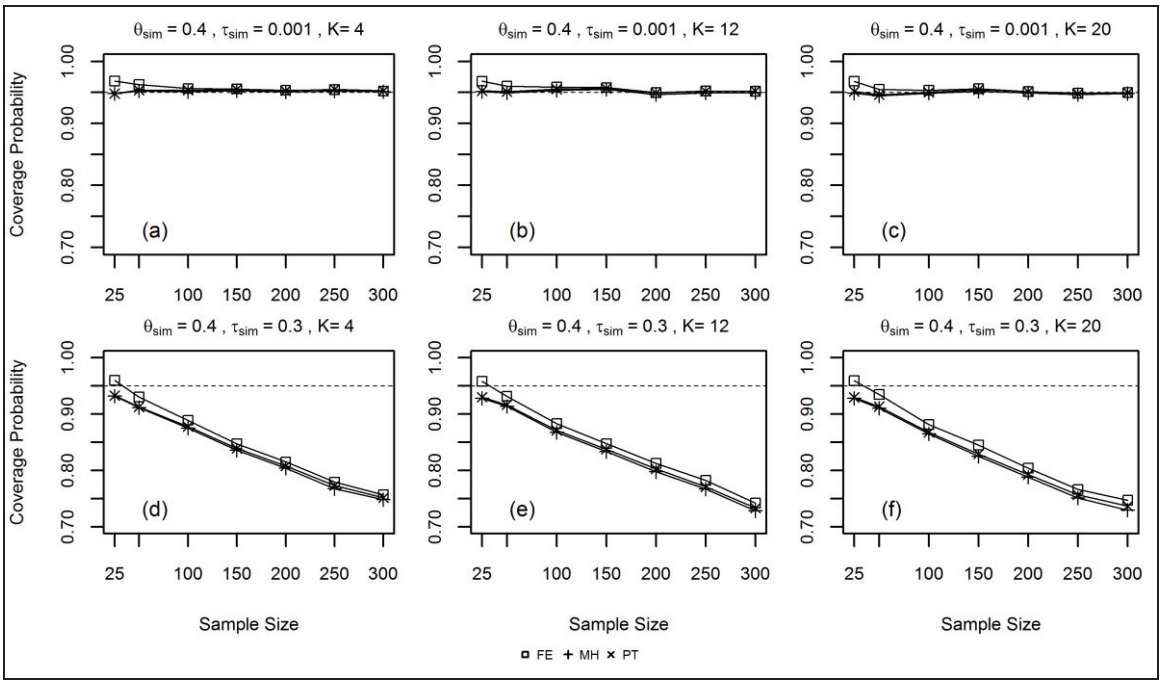

Figure 27 Coverage probability vs. sample size fixed-effects models

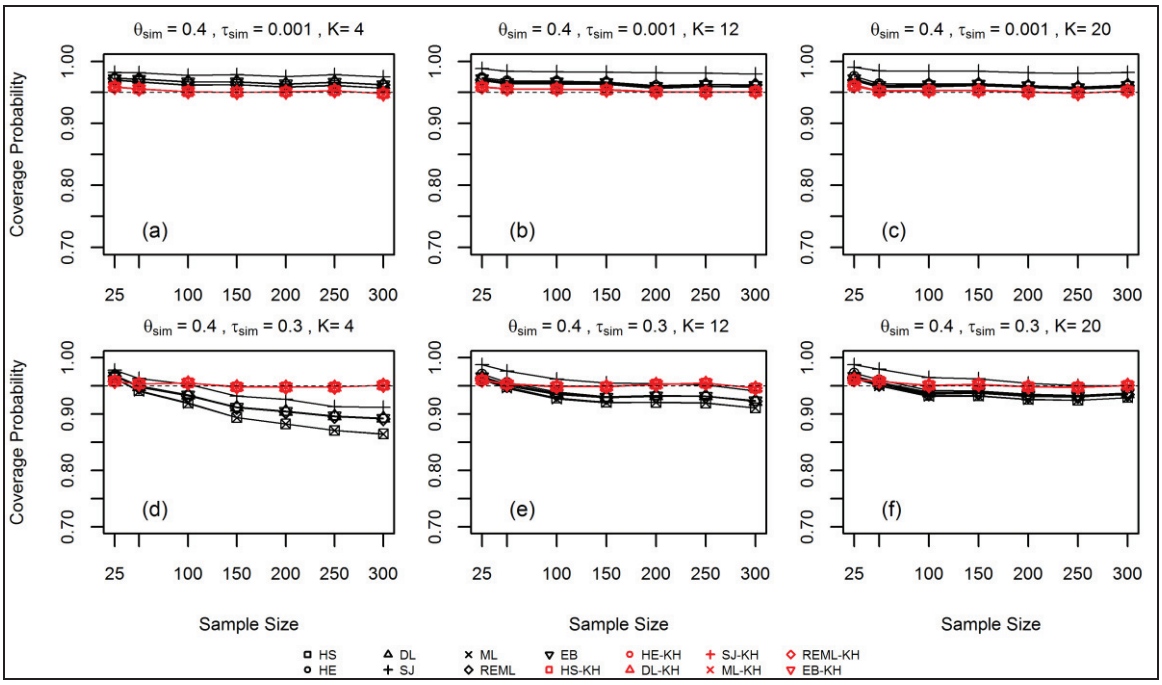

Figure 28 Coverage probability vs. sample size random-effects models

4.4 Conclusion and Discussion

In this chapter, I performed simulation studies to investigate the statistical powers and type-I errors of hypothesis tests on summary effect sizes, as well as the confidence interval coverage probabilities of classical meta-analysis models. Three fixed-effects models and 14 random-effects models were studied. I evaluated the impacts of study size, sample size, effect size to be detected, and population heterogeneity on the performances of meta-analysis models. Empirical powers from simulation studies and modified version of theoretical power were compared.

There are five major findings. (1) The study size and sample size significantly impact statistical power. But population heterogeneity parameter does not significantly affect the power. (2) All fixed-effects models have similar powers. Among non-KH-adjusted random-effects models, HE and ML models have highest powers, and SJ model has lowest power. On the other hand, all KH-adjusted random-effects models have similar powers. (3) All fixed-effects models have inflated type-I errors and reduced coverage probabilities in the presence of population heterogeneity. KH-adjusted random-effects models keep nominal level of Type-I error and coverage probability very well. Non-KH-adjusted random-effects models have slightly increased type-I errors and reduced coverage probabilities when population heterogeneity is present. (4) Theoretical powers and empirical powers are generally consistent within the classes of fixed-

effects models and non-KH-adjusted random-effects models. But the empirical powers of KH-adjusted random-effects models are lower than theoretical powers.

This chapter shed some new light on power consideration on summary effect size in the planning, conducting and reviewing meta-analysis. The results of this chapter indicate that power is not always as high as we desire in meta-analysis. If study size, sample size, and the effect size to be detected are not in some appropriate ranges, the statistical power of meta-analysis can be unacceptably low. Therefore, power analysis is highly recommended in both the planning stage of the new meta-analysis and the review stage of the existing meta-analysis.

Moreover, fixed-effects models have dramatically inflated Type-I errors and reduced coverage probabilities in the presence of population heterogeneity. Hence, the use of fixed-effects models in such situations requires extra caution. As for random-effects models, the KH-adjusted random-effects models have outstanding performances in keeping nominal level of Type-I error and coverage probability. Although the KH-adjusted models have slightly compromised power performances as compared with non-KH adjusted models, I still highly prefer KH-adjusted random-effects models because of their outstanding type-I error performances.

In addition, the modified version of theoretical power is generally consistent with empirical power using simulation approach, except in some cases when both study size and sample size are very small. But it is worth mentioning that, this consistency partially relies on the new modification proposed in this chapter. If $\bar{v}$, $\bar{v}^*$ and τ^2 are arbitrarily defined like in Hedges and Pigott (2001), rather than being estimated using simulations in this chapter, the calculated theoretical power will differ more from empirical power. Therefore, the modified version of theoretical power calculation is a good alternative when sample size and study size are not very small. While simulation-based power analysis still has a big advantage. It can be applied to essentially any kind of meta-analysis models as long as the data can be simulated. Some meta-analysis models, such as meta-analysis on survival data, are more complicated and current theoretical power calculation method may not be easily derived. In these situations, simulation-based power analysis is definitely a robust choice.

Chapter 5 Multi-Group Hierarchical Models for 100-Car data

5.1 Introduction

The impact of risk factors on the individual driving risk usually varies among different drivers. Driver demographic characteristics, such as age and gender, may play a significant role in differentiating drivers into different groups. The driving risk within a driver group may be similar, while the risks among different driver groups may vary significantly. In many situations, the researchers are interested in modeling not only the driving risk in each driver group, but also a summary driving risk across all driver groups. In this chapter, I proposed and implemented Bayesian approach multi-group hierarchical models to address this need.

In hierarchical models, the group driving risks and summary driving risk are modeled simultaneously in a hierarchical structure. Hierarchical models have several advantages over non-hierarchical models on separate group data or over aggregated models on combined data. First, hierarchical models reflect the researcher's belief on the relationship between group driving risks and summary driving risk. The derived summary driving risk will be more representative for overall risk. Second, there is a "borrow-strength" effect in hierarchical models. In Bayesian hierarchical model context, lower-level parameters share information with upper-level parameters, and upper-level parameters serve as a hub for lower level parameters to communicate and borrow strength from each other. Borrow-strength effect is more apparent when the sample sizes of the some groups are small. Third, multi-group hierarchical models can use stratified data from each group, eliminating potential confounding factors which can be a major issue in aggregated models.

Multi-group hierarchical models are closely related to meta-analysis. In mathematical perspective, the formulation of meta-analysis models and two-level hierarchical models are almost identical. The individual effect sizes and summary effect size in meta-analysis models are corresponding to the individual group effects and summary effect in multi-group hierarchical models. A major difference between meta-analysis and multi-group hierarchical model is that, in meta-analysis, the effect sizes are from multiple independent studies; while in multi-group hierarchical models, group effects are from multiple groups in a single study. While since the mathematical formulations are very similar in both contexts, I adopted the classical- and

Bayesian-approach fixed- and random-effects meta-analysis models and implement them under multi-group hierarchical model context in this chapter.

In this chapter, I am interested in modeling the impact of complex secondary tasks (a category of driving distraction) on driving risk using 100-Car NDS data. This impact may vary with different demographic features, such as age and gender. The objectives of this chapter are two folds: (1) use multi-group hierarchical models to investigate the group driving risks and overall driving risk; (2) perform simulation-based retrospective performance analysis on multi-group hierarchical models that are developed.

5.2 Data and Methods

5.2.1 100-Car complex secondary task data

The 100-Car NDS data have been introduced in Chapter 2. For the complex secondary task data, a total of 104 drivers and 8974 cases were involved.

The data collection follows a case-cohort study design. In a case-cohort design, two exposure cohorts (with secondary task vs. without secondary task) were defined at the beginning of the study, but the exposure information was not extracted in the NDS data collection process (cannot be extracted until video data reduction later). After data collection was complete, all 492 crashes and near-crashes (CNCs) and their exposure conditions were identified by kinematic triggers and video data reduction. On the other hand, 8482 the non-CNC baseline samples were randomly sampled from non-CNC driving data, and their exposure conditions were also examined by video data reduction.

Drivers are divided into 6 driver groups based on age (young: <25, mid-age: 25-55, elder >55) and gender (male, female). Using the same notations as shown in Table 13, the 100-Car complex secondary task data are shown in Table 16. The events/non-events are corresponding to CNCs/non-CNCs, and the treatment/control groups are corresponding to with secondary task/without secondary task.

Table 16 100-Car complex secondary task data

Group	St_i	Nt_i	Sc_i	Nc_i
young male	4	78	100	913
Young female	15	148	118	1473
mid-age male	15	211	116	3823
mid-age female	16	115	64	1197
elder male	3	39	37	738
elder female	0	22	4	217
(total)	53	613	439	8361

5.2.2 Multi-group hierarchical models

Both fixed- and random-effects models can be used. Fixed-effects models are corresponding to the conditional inference, meaning that the inference results are only based on existing groups of data. Random-effects models are related to unconditional inference, indicating that the inference results can be extended to a larger universe of groups that include the observed groups. Fixed-effects models are actually a better fit in terms of data context, because the defined driver groups basically cover entire range of age and gender. Nevertheless, random-effects models can be very useful in situations where the defined driver groups are only a subset of the universe of driver groups. Om this chapter, for methodology comparison purpose, I performed both fixed- and random-effects models on the data.

5.2.2.1 Classical-approach multi-group hierarchical models

Classical fixed- and random-effects meta-analysis models were extensively discussed in Chapter 4. I adopted these models and implemented them interchangeably as multi-group hierarchical models in this chapter.

5.2.2.2 Bayesian-approach multi-group hierarchical models

Bayesian fixed-effects models

Bayesian fixed-effects models are effective counterparts to classical fixed-effects models (Sutton and Abrams 2001). Two Bayesian fixed-effects models were studied in this chapter: Bayesian fixed-effects model with binomial likelihood function (BFB) and Bayesian fixed-effects model with normal likelihood function (BFN). The model formulations are shown in Table 17, with *Pt* and *Pc* being the event probabilities for treatment and control groups. The formation of BFB model is inspired by Smith *et al.* (1995) paper; while the BFN model directly model log-odds

ratios using normal likelihood functions, and can be considered as a Bayesian counterpart to classical FE model. As to the prior specification, we adopted the informative priors proposed in Smith *et al.* (1995), where the priors were carefully elicited based on the numerical properties of log-odds ratio measure.

Table 17 Bayesian Fixed-Effects Models

Model	Likelihood	Priors
BFB	$a_i \mid Pt \sim Binomial(Pt, n_{1i})$, $c_i \mid Pc \sim Binomial(Pc, n_{2i})$ $\mu = \text{logit}(Pc)$, $\theta = \text{logit}(Pt) - \text{logit}(Pc)$	$\mu \sim Normal(0,12.4)$ $\theta \sim Normal(0,10)$
BFN	$y_i = \log(a_i d_i / b_i c_i)$, $s_i = \sqrt{a_i^{-1} + b_i^{-1} + c_i^{-1} + d_i^{-1}}$ $y_i \mid \theta \sim Normal(\theta, s_i^2)$	$\theta \sim Normal(0,10)$

Bayesian random-effects models

Bayesian random-effects meta-analysis models have been studied in many literatures (e.g. (Smith *et al.* 1995, Thompson *et al.* 1997, Normand 1999, Sutton and Abrams 2001, Warn *et al.* 2002)). The flexibility in modeling hierarchical structure of parameters makes them a good counterpart to classical models. Three Bayesian random-effects models were studied in this chapter (shown in Table 18): Bayesian random-effects binomial-1 (BRB1) and -2 (BRB2) models, and Bayesian random-effects normal (BRN) model. The models and prior specifications were based on Smith *et al.* (1995) with some modifications.

Table 18 Bayesian Random-Effects Models

Model	Likelihood	Priors
BRB1	$a_i \mid Pt \sim Binomial(Pt, n_{1i})$, $c_i \mid Pc \sim Binomial(Pc, n_{2i})$, $\mu_i = \text{logit}(Pc_i)$, $\theta_i = \text{logit}(Pt_i) - \text{logit}(Pc_i)$, $\theta_i \mid \theta, \tau^2 \sim Normal(\theta, \tau^2)$	$\mu_i \sim Normal(0, 12.4)$ $\theta \sim Normal(0, 10)$ $(\tau^2)^{-1} \sim Gamma(shape = 3, rate = 1)$
BRB2	$a_i \mid Pt \sim Binomial(Pt, n_{1i})$, $c_i \mid Pc \sim Binomial(Pc, n_{2i})$, $\theta_i = \text{logit}(Pt_i) - \text{logit}(Pc_i)$, $\theta_i \mid \theta, \tau^2 \sim Normal(\theta, \tau^2)$	$Pc_i \sim Uniform(0.0001, 0.9999)$ $\theta \sim Normal(0, 10)$ $(\tau^2)^{-1} \sim Gamma(shape = 3, rate = 1)$
BRN	$y_i = \log(a_i d_i / b_i c_i)$, $s_i = \sqrt{a_i^{-1} + b_i^{-1} + c_i^{-1} + d_i^{-1}}$ $y_i \sim Normal(\theta_i, s_i^2)$, $\theta_i \mid \theta, \tau^2 \sim Normal(\theta, \tau^2)$	$\theta \sim Normal(0, 10)$ $(\tau^2)^{-1} \sim Gamma(shape = 3, rate = 1)$

5.3 Results

5.3.1 Model fitting results

Five fixed-effects models and 22 random-effects models were implemented. The model fitting results of the summary effect (summary log-Odds Ratio) are shown in Section 5.3.1.

5.3.1.1 Fixed-effects models

The estimated summary log-odds ratios using fixed-effects models are shown in Table 19 and Figure 29. They show that all fixed-effects models, including both classical- and Bayesian-approach models, yielded similar estimates for summary effect. In fixed-effects model context, complex secondary tasks are significantly associated with increased driving risk.

Table 19 Estimated summary effect using fixed-effects models

Approach	Model	Summary log-Odds Ratio		
		Mean	95% LCL	95% UCL
classical	FE	0.57	0.27	0.87
	MH	0.44	0.14	0.74
	PT	0.52	0.17	0.87
Bayesian	BRB	0.53	0.22	0.82
	BRN	0.57	0.27	0.88

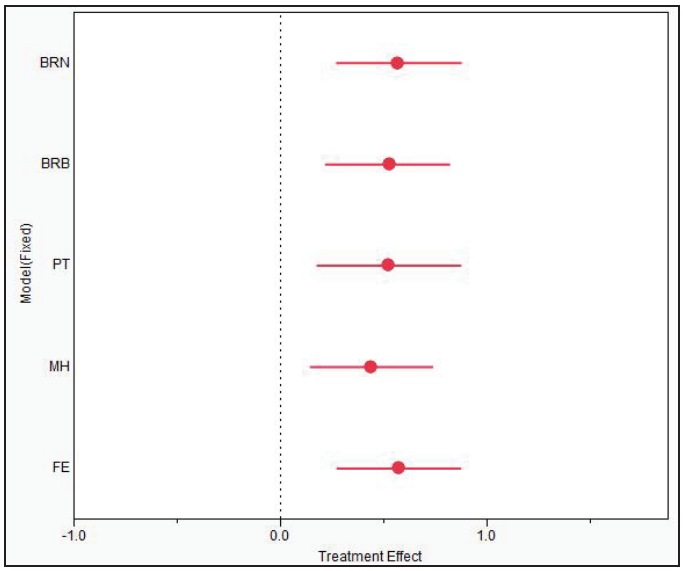

Figure 29 Summary effect using fixed-effects models

5.3.1.2 Random-effects models

The estimated summary log-odds ratios using random-effects models are shown in Table 20 and Figure 30. In random-effects model context, most models (except for non-KH-adjusted HS and HE models) do not show a significant increased driving risk due to complex secondary tasks. Moreover, Bayesian approach random-effects models have larger estimates of $\hat{\tau}^2$, and estimates of θ are towards zero. This is because of the incorporation of prior information on θ and τ^2. In classical random-effects models, KH-adjusted versions have wider confidence intervals than non-KH-adjusted versions, even both use the same estimates of $\hat{\tau}^2$. This is because KH-adjusted versions use t-statistic, rather than z-statistic. Moreover, it is interesting that HS and HE models yield much smaller estimates of $\hat{\tau}^2$ than other models. HE model yielded an estimate of 0 for τ^2, which is theoretically valid (see Table 15) because the lower bound of the τ^2 estimate is truncated to zero.

Table 20 Estimated summary effect using random-effects models

Approach	Model	Summary log-Odds Ratio			$\hat{\tau}^2$
		Mean	95% LCL	95% UCL	
Classical	HS	0.48	0.01	0.95	0.15
	HE	0.57	0.27	0.87	0
	DL	0.45	-0.08	0.98	0.23
	SJ	0.44	-0.13	1.01	0.29
	ML	0.46	-0.04	0.97	0.19
	REML	0.44	-0.14	1.01	0.29
	EB	0.45	-0.09	0.99	0.24
	HS-KH	0.48	-0.22	1.17	0.15
	HE-KH	0.57	-0.05	1.19	0
	DL-KH	0.45	-0.26	1.16	0.23
	SJ-KH	0.44	-0.28	1.15	0.29
	ML-KH	0.46	-0.24	1.17	0.19
	REML-KH	0.44	-0.28	1.15	0.29
	EB-KH	0.45	-0.26	1.16	0.24
Bayesian	BRB1	0.32	-0.38	0.97	0.33
	BRB2	0.30	-0.40	0.94	0.31
	BRN	0.42	-0.25	1.05	0.33

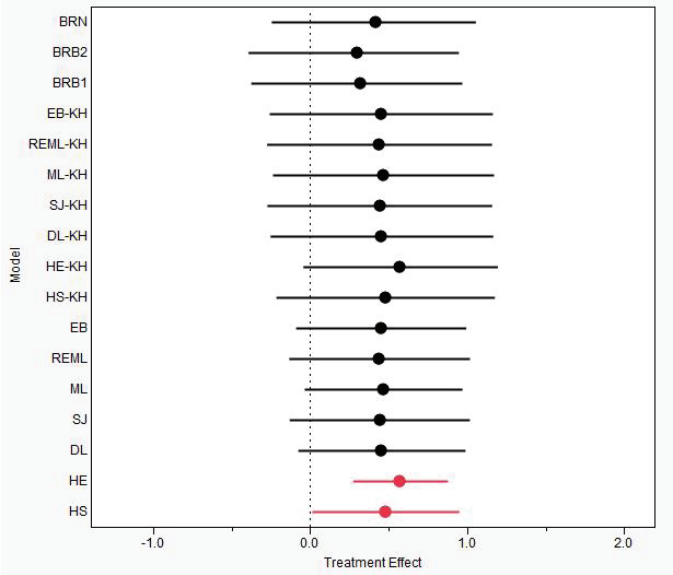

Figure 30 Summary effect using random-effects models

5.3.2 Retrospective performance analysis

In Section 5.3.2, we used simulation-based performance analysis method proposed in Chapter 4 to investigate the empirical power, type-I error and coverage probability of the multi-group hierarchical models based on 100-Car complex secondary task data.

5.3.2.1 Methods of retrospective performance analysis

The data generation and simulation methods in Section 5.3.2 are very similar to the methods proposed in Chapter 4. The major difference is that, here I performed a retrospective performance analysis based on existing data and model, while in Chapter 4 I performed prospective performance analysis based on arbitrarily defined parameter settings.

To perform retrospective performance analysis, first I need to extract information from existing data. Based on the data in Table 16 and Table 20, I set K=6, $\tau_{sim} = 0.4$, θ_{sim} = (0 to 0.7 by 0.05), Nc = (913, 1473, 3823, 1197, 738, 217), Nt = (78, 148, 211, 115, 39, 22), and $Pc_i \sim Uniform(0.02, 0.11)$. Then simulation procedures proposed in Chapter 4 (Section 4.2.3) were implemented with T=10,000 simulation runs.

Table 16 shows that CNCs are rare events, and thus the CNC event probabilities in both exposed and unexposed groups are low. This imposed a zero cell count problem in the simulation studies. In preliminary studies I have observed that, a large portion of simulation runs include more than one zero cell within in a group. The small cell counts significantly affected the model implementation. The most notable outcome is the convergence issue in classical models that require iterative numerical solution (i.e. ML, REML, EB). Therefore, in this section, I only investigated classical models with closed form solutions, i.e. FE, MH, PT fixed-effects models and HS, HE, DL, SJ random-effects models.

Furthermore, because the concepts of hypothesis test and power are under classical context, but not in Bayesian context, I did not include Bayesian hierarchical models in Section 5.3.2.

5.3.2.2 Power and type-I error

The empirical powers and type-I errors of classical fixed- and random-effects models are shown in Figure 31 and Figure 32. The red vertical dashed line indicates the type-I error, and blue vertical dashed line indicates the power based on the actual data that are observed. The results show that, the powers of fixed-effects models for real data are above 0.8, but type-I errors are as

high as about 0.2. On the other hand, random-effects models without KH-adjustment have powers between 0.5 and 0.75, and also have type-I errors between 0.1 and 0.2. In contrast, KH-adjusted models only have powers between 0.4 and 0.5, but with a low type-I error between 0.05 and 0.1.

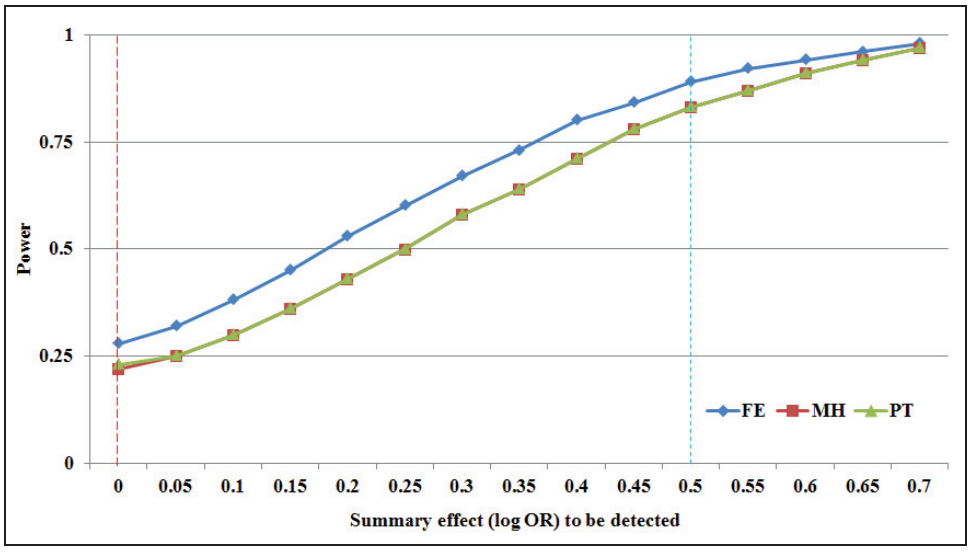

Figure 31 Powers and type-I errors in fixed-effects models

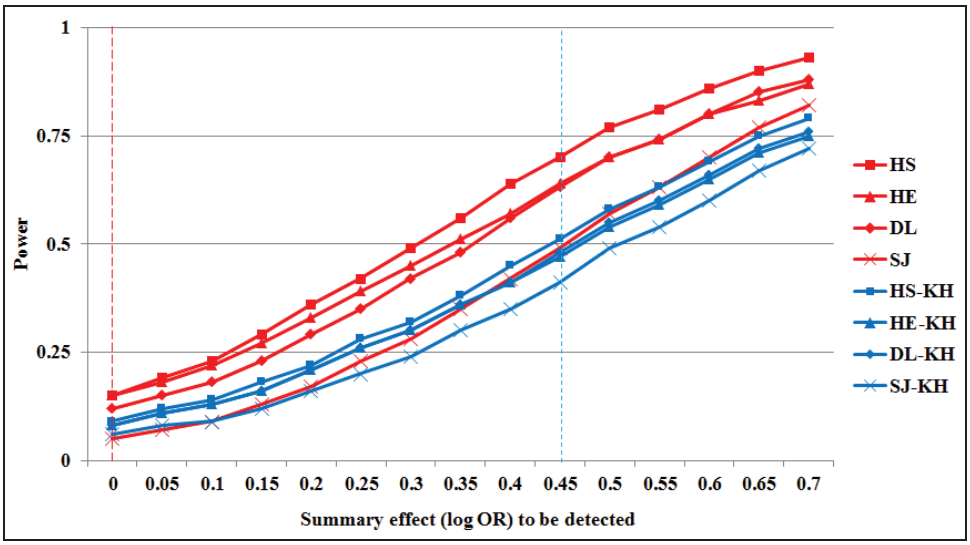

Figure 32 Powers and type-I errors in random-effects models

5.3.2.3 Coverage probability

The coverage probabilities of summary effect (summary log-odds ratio) in classical fixed- and random-effects models are shown in Figure 33 and Figure 34. The blue vertical dashed line

indicates the coverage probabilities based on real data that are observed. The results show that, the coverage probabilities in fixed-effects models based on the real data are below 0.75. On the other hand, non-KH-adjusted random-effects models (except SJ) have coverage probabilities between 0.85 and 0.9, while KH-adjusted random-effects models and non-KH-adjusted SJ models have coverage probabilities as high as between 0.9 and 0.95.

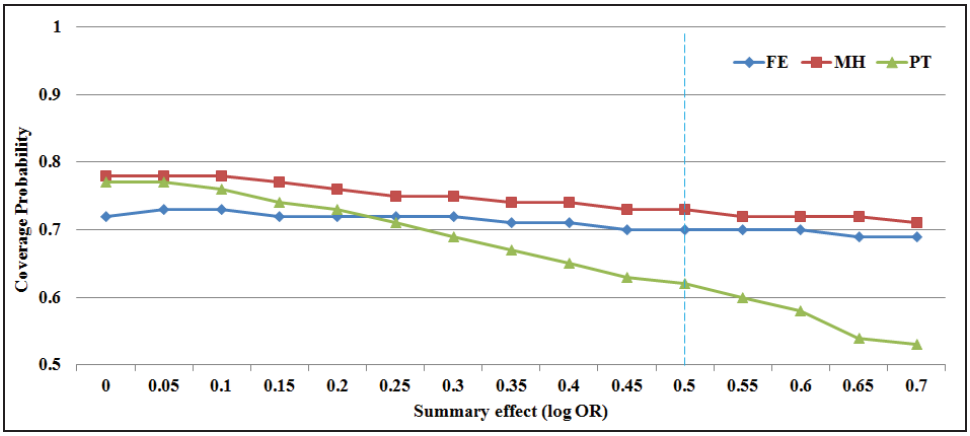

Figure 33 Coverage probabilities in fixed-effects models

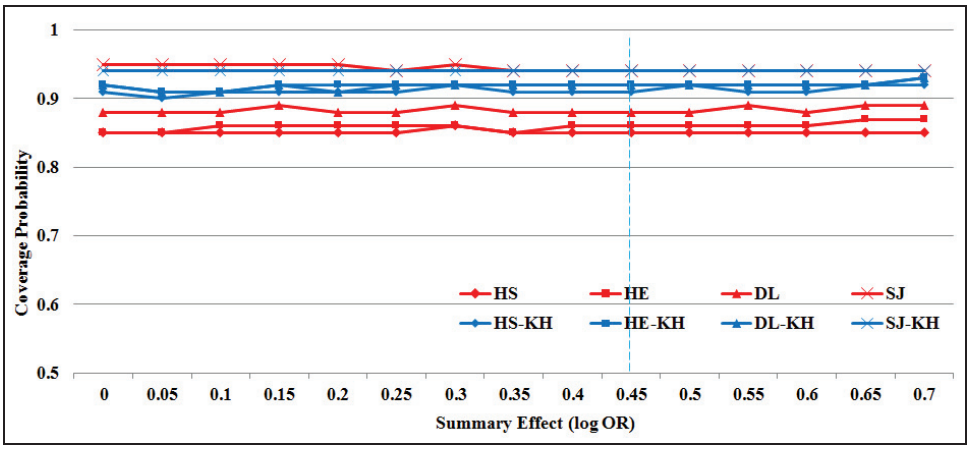

Figure 34 Coverage probabilities in random-effects models

5.3.3 Additional analysis on coverage probabilities in classical and Bayesian models

Independent of Section 5.3.1 and 5.3.2, in this Section 5.3.3 I performed a simulation-based coverage probability analysis using procedures similar to Chapter 4 (Section 4.3.2.3). In this section, I considered both classical- and Bayesian-approach models.

The results in Figure 35 and Figure 36 show that, the performances of Bayesian approach and classical approach fixed-effects models are similar. The coverage probability decrease as the

population heterogeneity increases. This relationship is affected by group sample size, but is not affected by group size (number of groups). On the other hand, for random-effects models, the coverage probabilities in Bayesian random-effects models are consistently higher than nominal level, in the levels as high as 0.97 to 0.99. These results imply that the Bayesian inference (based on existing model specification and prior settings) are very conservative.

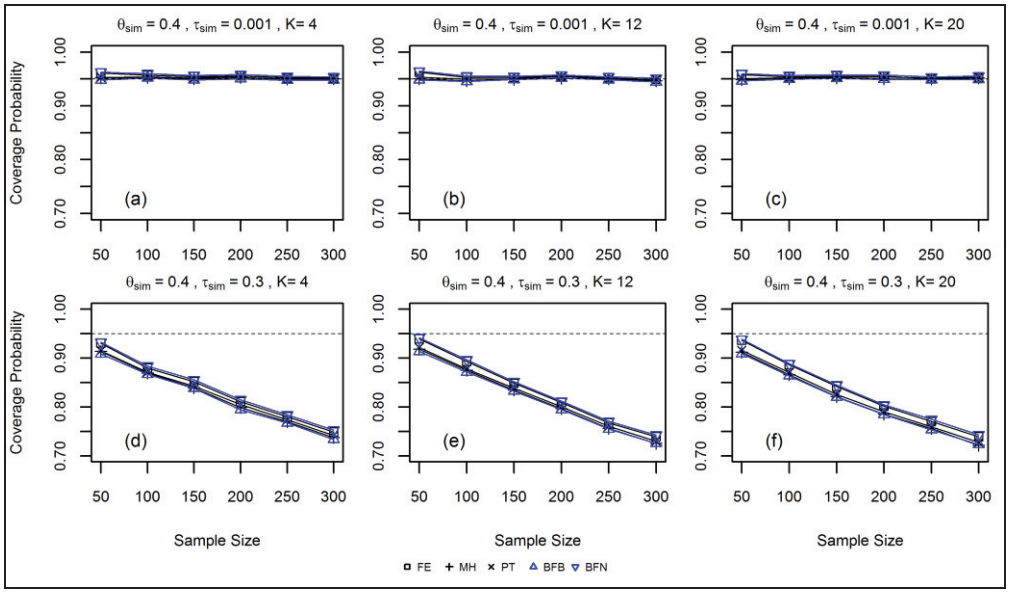

Figure 35 Coverage probabilities in fixed-effects models

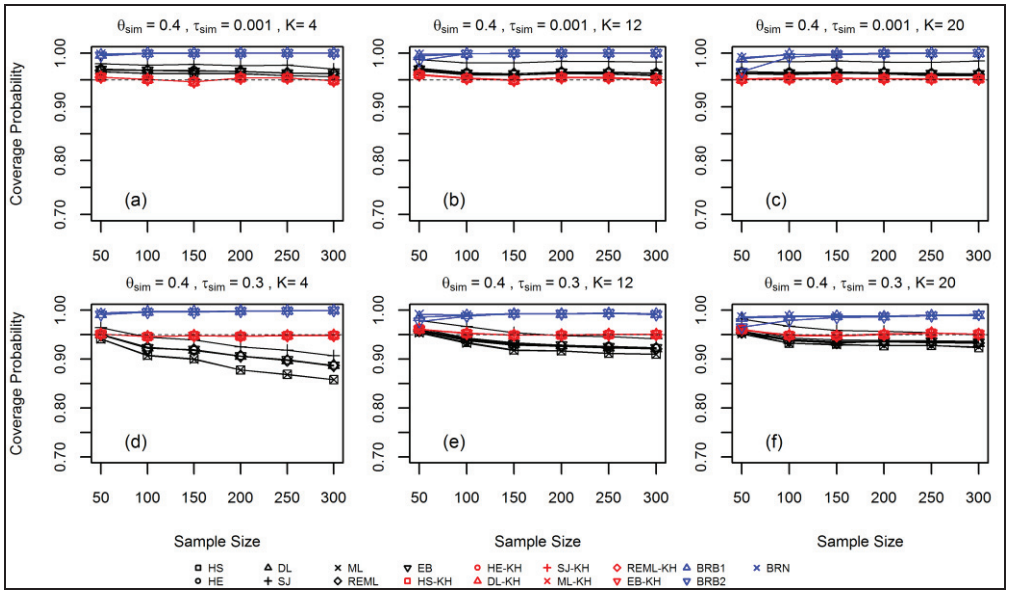

Figure 36 Coverage probabilities in random-effects models

5.4 Conclusion and Discussion

In this chapter, I implemented multi-group hierarchical models on 100-Car complex secondary task data. The hierarchical model can simultaneously model group-level and overall-level driving risks.

The model fitting results on 100-Car data showed that, under fixed-effects model context, complex secondary tasks are associated with higher driving risk. These inference results have powers over 0.8, but with type-I errors as high as over 0.2. In contrast, under random-effects model context, complex secondary tasks are not significantly associated with higher driving risk. These inference results have powers over 0.4 to 0.7, and with type-I errors over 0.05 to 0.2. Because fixed- and random-effects models are conceptually different (see Chapter 4 for detailed discussion), researchers need to use appropriate model contexts based on specific research goals, rather than simply seeking for a high power or low type-I error.

Simulation-based retrospective performance analysis showed that, the coverage probabilities in classical and Bayesian fixed-effects models are lower than nominal level in the presence of population heterogeneity. For random-effects models, classical random-effects models (especially the KH-adjusted versions) maintain nominal level of coverage probability relatively well. In contrast, Bayesian random-effects models have much higher coverage probabilities than nominal level, implying that the inferences based on Bayesian random-effects models are conservative.

Reference

Agresti, A., 2002. Categorical data analysis, 2nd ed. Wiley-Interscience, New York.

Antin, J.F., Lockhart, T.E., Stanley, L.M., Guo, F., 2012. Comparing the impairment profiles of older drivers and non-drivers: Toward the development of a fitness-to-drive model. Safety Science 50, 333-341.

Arthur, W., Graziano, W.G., 1996. The five-factor model, conscientiousness, and driving accident involvement. Journal of Personality 64 (3), 593-618.

Benmimoun, A., Benmimoun, M., Van Noort, M., Wilmink, I., 2009. Eurofot: Large scale field operational test - impact assessment. 16th ITS World Congress, Stockholm.

Borenstein, M., Hedges, L.V., Higgins, J.P.T., Rothstein, H.R., 2009. Introudction to meta-analysis John Wiley & Sons Ltd, United Kindom.

Bradburn, M.J., Deeks, J.J., Berlin, J.A., Localio, R., 2007. Much ado about nothing: A comparison of performance of meta-analytical methods with rare events. Statistics in Medicine 26, 53-77.

Cdc, 2010. Wisqars (web-based injury statistics query and reporting system). US Department of Health and Human Services, CDC, Atlanta, GA.

Committee for the Strategic Highway Research Program 2: Implementatation, 2009. Implementing the results of the second strategic highway research program:Saving lives, reducing congestion, improving quality of life - special report 296 The National Academies Press.

Costa, P.T., Mccrea, R.R., 1992. Revised neo personality inventory (neo pi-r) and neo five-factor inventory Psychological Assessment Resources, Odessa, Fl.

Council, N.S., 2010. Understanding the distracted brain: Why driving while using hands-free cell phones is risky behavior.

Dahlen, E.R., White, R.P., 2006. The big five factors, sensation seeking, and driving anger in the prediction of unsafe driving. Personality and Individual Differences 41 (5), 903-915.

Deery, H.A., Fildes, B.N., 1999. Young novice driver subtypes: Relationship to high-risk behavior, traffic accident record, and simulator driving performance. Human Factors: The Journal of the Human Factors and Ergonomics Society 41 (4), 628-643.

Dersimonian, R., Laird, N., 1986. Meta-analysis in clinical trials. Controlled Clinical Trials 7, 177-188.

Dingus, T.A., Klauer, S.G., Neale, V.L., Petersen, A., Lee, S.E., Sudweeks, J., Perez, M.A., Hankey, J., Ramsey, D., Gupta, S., Bucher, C., Doerzaph, Z.R., Jermeland, J., Knipling, R.R., 2006. The100-car naturalistic driving study: Phase ii – results of the 100-car field experiment.

Donmez, B., Boyle, L.N., Lee, J.D., 2010. Differences in off-road glances: Effects on young drivers' performance. Journal of Transportation Engineering-Asce 136 (5), 403-409.

Fitch, G.M., Hanowski, R.J., Year. The risk of a safety-critical event associated with mobile device use as a function of driving task demands. In: Proceedings of the 2nd International Conference on Driver Distraction and Inattention, Gothenburg, Sweden.

Fitch, G.M., Soccolich, S.A., Guo, F., Mcclafferty, J., Fang, Y., Olson, R.L., Perez, M.A., Hanowski, R.J., Hankey, J.M., Dingus, T.A., 2013. The impact of hand-held and hands-free cell phone use on driving performance and safety-critical event risk.

Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 2004. Bayesian data analysis, 2nd ed. CRC press.

Gelman, A., Rubin, D.B., 1992. Inference from iterative simulation using multiple sequences. Statistical Science 7 (4), 457-511.

Guo, F., Fang, Y., 2013. Individual driver risk assessment using naturalistic driving data. Accident Analysis and Prevention. pp. 3-9.

Guo, F., Hankey, J.M., 2009. Modeling 100-car safety events: A case-based approach for analyzing naturalistic driving data. the National Surface Transportation Safety Center for Excellence.

Guo, F., Klauer, S.G., Hankey, J.M., Dingus, T.A., 2010a. Near crashes as crash surrogate for naturalistic driving studies. Transportation Research Record: Journal of the Transportation Research Board 2147 (1), 66-74.

Guo, F., Klauer, S.G., Hankey, J.M., Dingus, T.A., 2010b. Using near-crashes as a crash surrogate for naturalistic driving studies. Transportation Research Record: Journal of the Transportation Research Board 2147 (66-74).

Hardy, R.J., Thompson, S.G., 1996. A likelihood approach to meta-analysis with random effects. Statistics in Medicine 15, 619-629.

Hartung, J., Knapp, G., 2001a. On tests of the overall treatment effect in meta-analysis with normally distributed responses. Statistics in Medicine 20, 1771-1782.

Hartung, J., Knapp, G., 2001b. A refined method for the meta-analysis of controlled clinical trials with binary outcome. Statistics in Medicine 20, 3875-3889.
Harville, D.A., 1977. Maximum likelihood approaches to variance component estimation and to related problems. Journal of American Statistical Association 72, 320-339.
Hauer, E., Garder, P., 1986. Research into the validity of the traffic conflicts technique. Accident Analysis & Prevention 18 (6), 471-481.
Hauer, E., Ng, J., Lovell, J., 1988. Estimation of safety at signalized intersections. Transportation Research Record: Journal of the Transportation Research Board, 48-61.
Hedges, L.V., 1983. A random effects model for effect sizes. Psychological Bulletin 93, 388-395.
Hedges, L.V., Olkin, I., 1985. Statistical methods for meta-analysis, 1st ed. Academic Press.
Hedges, L.V., Pigott, T.D., 2001. The power of statistical tests in meta-analysis. Psychological Methods 6 (3), 203-217.
Hedges, L.V., Pigott, T.D., 2004. The power of statistical tests for moderators in meta-analysis. Psychological Methods 9 (4), 426-445.
Hedges, L.V., Vevea, J.L., 1998. Fixed- and random- effects models in meta-analysis. Psychological Methods 3, 486-504.
Hickman, J.S., Hanowski, R.J., Bocanegra, J., 2010. Distraction in commercial trucks and buses: Assessing prevalence and risk in conjunction with crashes and near-crashes. Report No. FMCSA-RRR-10-049, Washington, DC: Federal Motor Carrier Safety Administration. .
Higgins, J.P., Green, S., 2008. Cochrane handbook for systematic reviews of interventions Wiley, Chichester, England.
Hunter, J.E., Schmidt, F.L., 1990. Methods of meta-analysis: Correcting error and bias in research findings Sage, Newbury Park, CA.
Ishigami, Y., Klein, R.M., 2009. Is a hands-free phone safer than a handheld phone? Journal of Safety Research 40 (2), 157-164.
Jackson, D., 2006. The power of the standard test for the presence of heterogeneity in meta-analysis. Statistics in Medicine 25, 2688-2699.
Jolliffe, I.T., 2002. Principal component analysis Springer.
Jonah, B.A., 1997. Sensation seeking and risky driving: A review and synthesis of the literature. Accident Analysis & Prevention 29 (5), 651-665.
Jonah, B.A., Thiessen, R., Au-Yeung, E., 2001. Sensation seeking, risky driving and behavioral adaptation. Accident Analysis & Prevention 33 (5), 679-684.
Klauer, S.G., Dingus, T.A., Neale, V.L., Sudweeks, J.D., Ramsey, D.J., 2006. The impact of driver inattention on near-crash/crash risk: An analysis using the 100-car naturalistic driving study data. (Report No. DOT HS 810 594). Washington, DC: National Highway Traffic Safety Administration.
Klauer, S.G., Guo, F., Simons-Morton, B.G., Ouimet, M.C., Lee, S.E., Dingus, T.A., 2014. Distracted driving and risk of road crashes among novice and experienced drivers. New England Journal of Medicine (370), 54-59.
Klauer, S.G., Guo, F., Sudweeks, J.D., Dingus, T.A., 2010. An analysis of driver inattention using a case-crossover approach on 100-car data.
Knapp, G., Hartung, J., 2003. Improved tests for a random effects meta-regression with a single covariate. Statistics in Medicine 22, 2693-2710.
Laird, N.M., Mosteller, F., 1990. Some statistical methods for combining experimental results. International Journal of Technology Assessment in Health Care 6, 5-30.
Loo, R., 1979. Role of primary personality factors in the perception of traffic signs and driver violations and accidents. Accident Analysis & Prevention 11 (2), 125-127.
Lord, D., Mannering, F., 2010. The statistical analysis of crash-frequency data: A review and assessment of methodological alternatives. Transportation Research Part A 44 (5), 291-305.
Machin, M.A., Sankey, K.S., 2008. Relationships between young drivers' personality characteristics, risk perceptions, and driving behaviour. Accident Analysis & Prevention 40 (2), 541-547.
Mantel, N., Haenszel, W., 1959. Statistical aspects of the analysis of data from retrospective studies of disease. Journal of the National Cancer Institute 22, 719-748.
Maze, T.H., Agarwai, M., Burchett, G., 2006. Whether weather matters to traffic demand, traffic safety, and traffic operations and flow. Transportation Research Record: Journal of the Transportation Research Board 1948, 170-176.

Mcevoy, S.P., Stevenson, M.R., Mccartt, A.T., Woodward, M., Haworth, C., Palamara, P., Cercarelli, R., 2005. Role of mobile phones in motor vehicle crashes resulting in hospital attendance: A case-crossover study. BMJ 331 (7514), 428.
Morris, C.N., 1983. Parametric empirical bayes inference: Theory and applications. Journal of American Statistical Association 78, 47-55.
National Highway Traffic Safety Administration, 2008. Traffic safety facts 2006: A compliation of motor vehicle crash data from the fatality analysis reporting system and the general estimates system.
Normand, S.-L.T., 1999. Tutorial in biostatistics meta-analysis: Formulating, evaluating, combing, and reporting. Statistics in Medicine 18, 321-359.
Nsc, 2009. National safety council attributable risk estimate model. December 2009. In: Council, N.S. ed.
Olson, R.L., Hanowski, R.J., Hickman, J.S., Bocanegra, J., 2009. Driver distraction in commercial vehicle operations: Final report. Contract dtmc75-07-d-00006, task order 3. Federal Motor Carrier Safety Administration, Washington, D.C.
Ouimet, M.C., Brown, T.G., Guo, F., Klauer, S.G., Simons-Morton, B.G., Fang, Y., Lee, S.E., Gianoulakis, C., Dingus, T.A., 2014. Higher crash and near-crash rates in teenage drivers with lower cortisol reactivity: An 18-month longitudinal, naturalistic study. JAMA Pediatrics 168 (6), 517-522.
Poch, M., Mannering, F., 1996. Negative binomial analysis of intersection-accident frequencies. Journal of Transportation Engineering 122, 105-113.
Raudenbush, S.W., Bryk, A.S., 1985. Empirical bayes meta-analysis. Journal of Educational Statistics 10, 75-98.
Robert, C.P., Casella, G., 2005. Monte carlo statistical methods, 2nd ed. Springer, New York, NY, USA.
Segovia-Gonzalez, M.M., Guerrero, F.M., Herranz, P., 2009. Explaining functional principal component analysis to actuarial science with an example on vehicle insurance. Insurance: Mathematics and Economics 45, 278-285.
Shaw, L., Sichel, H.S., 1971. Accident proneness: Research in the occurrence, causation, and prevention of road accidents Oxford, England: Pergamon.
Sidik, K., Jonkman, J.N., 2005. Simple heterogeneity variance estimation for meta-analysis. Journal of the Royal Statistical Society, Series C 54, 367-384.
Sidik, K., Jonkman, J.N., 2007. A comparison of heterogeneity variance estimators in combining results of studies. Statistics in Medicine 26, 1964-1981.
Smith, T.C., Spiegelhalter, D.J., Thomas, A., 1995. Bayesian approaches to random-effects meta-analysis: A comparative study. Statistics in Medicine 14, 2685-2699.
Spiegelhalter, D.J., Best, N.G., Carlin, B.P., Linde, A.V.D., 2002. Bayesian measures of model complexity and fit. Journal of the Royal Statistical Society: Series B 64 (4), 583-639.
Stutts, J.C., Reinfurt, D.W., Staplin, L., Rodgman, E.A., 2001. The role of driver distraction in traffic crashes. AAA Foundation for Traffic Safety.
Sutton, A.J., Abrams, K.R., 2001. Bayesian methods in meta-analysis and evidence synthesis. Statistical Methods in Medical Research 10, 277-303.
Sutton, A.J., Higgins, J.P.T., 2008. Recent developments in meta-analysis. Statistics in Medicine 27 (625-650).
Tan, P.N., Steinbach, M., Kumar, V., 2005. Introduction to data mining Addison-Wesley.
Tarko, A., Davis, G., Saunier, N., Sayed, T., Washington, S., 2009. Surrogate measure of safety: White paper. Transportation Research Board ANB20(3) Subcommittee on Surrogate Measures of Safety.
Thompson, S.G., Sharp, S.J., 1999. Explaining heterogeneity in meta-analysis: A comparison of methods. Statistics in Medicine 18, 2693-2708.
Thompson, S.G., Smith, T.C., Sharp, S.J., 1997. Investigating underlying risk as a source of heterogeneity in meta-analysis. Statistics in Medicine 16, 2741-2758.
Thorlund, K., Mills, E.J., 2012. Sample size and power considerations in network meta-analysis. Systematic Reviews 1.
Tiwari, G., Mohan, D., Fazio, J., 1998. Conflict analysis for prediction of fatal crash locations in mixed traffic streams. Accident Analysis & Prevention 30 (2), 207-215.
U.S. Census Bureau, 2012. Statistical abstract of the united states: 2012, table 1103. In: Bureau, U.S.C. ed.
Ulleberg, P., 2001. Personality subtypes of young drivers. Relationship to risk-taking preferences, accident involvement, and response to a traffic safety campaign. Transportation Research Part F: Traffic Psychology and Behaviour 4 (4), 279-297.
Ulleberg, P., Rundmo, T., 2003. Personality, attitudes and risk perception as predictors of risky driving behaviour among young drivers. Safety Science 41 (5), 427-443.

University of Michigan Transportation Research Institute, 2005. Automotive collision avoidance system field operational test methodology and results appendices. National Highway Traffic Safety Administration.
Valentine, J.C., Pigott, T.D., Rothstein, H.R., 2010. How many studies do you need? : A primer on statistical power for meta-analysis. Journal of Educational and Behavioral Statistics 35 (2), 215-247.
Viechtbauer, W., 2005. Bias and efficiency of meta-analytic variance estimators in the random effects model. Journal of Educational and Behavioral Statistics 30, 261-293.
Viechtbauer, W., 2007. Hypothesis tests for population heterogeneity in meta-analysis. British Journal of Mathematical and Statistical Psychology 60, 29-60.
Viechtbauer, W., 2010. Conducting meta-analyses in r with the metafor package. Journal of Statistical Software 36 (3), 1-48.
Viechtbauer, W., 2014. Package 'metafor'. Meta-Analysis Package for R, http ://cran.r-project.org/web/packages/metafor/metafor.pdf (2011, accessed 11 June 2011).
Walters, M.A., 1981. Risk classification standards. Proceedings of the Casualty Actuarial Society 68, 1-18.
Warn, D.E., Thompson, S.G., Spiegelhalter, D.J., 2002. Bayesian random effects meta-analysis of trials with binary outcomes: Methods for the absolute risk difference and relative risk scales. Statistics in Medicine 21, 1601-1623.
Williams, M.J., 1981. Validity of the traffic conflicts technique. Accident Analysis & Prevention 13 (2), 133-145.
Yusuf, S., Peto, R., Lewis, J., Collins, R., Sleight, P., 1985. Beta blockade during and after myocardial infarction: An overview of the randomized trials. Progress in Cardiovascular Diseases 25, 335-371.