Accident Analysis and Prevention 48 (2012) 401–415

Contents lists available at SciVerse ScienceDirect

Accident Analysis and Prevention

journal homepage: www.elsevier.com/locate/aap

Nonresponse analysis and adjustment in a mail survey on car accidents

Emma Tivesten[a,b,*], Sofia Jonsson[a,c], Lotta Jakobsson[a,b], Hans Norin[b]

[a] Dept 91410, Volvo Cars Safety Centre, Volvo Car Corporation, Göteborg, Sweden
[b] Vehicle Safety Division, Department of Applied Mechanics, Chalmers University of Technology, Göteborg, Sweden
[c] Volvia/If, Göteborg, Sweden

ARTICLE INFO

Article history:
Received 8 August 2011
Received in revised form 20 January 2012
Accepted 16 February 2012

Keywords:
Road safety
Nonresponse bias
Statistical accident data
Post survey adjustment
Driver distraction
Driver vigilance

ABSTRACT

Statistical accident data plays an important role for traffic safety development involving the road system, vehicle design, and driver education. Vehicle manufacturers use data from accident mail surveys as an integral part of the product development process. Low response rates has, however, lead to concerns on whether estimates from a mail survey can be trusted as a source for making strategic decisions.

The main objective of this paper was to investigate nonresponse bias in a mail survey addressing driver behaviour in accident situations. Insurance data, available for both respondents and nonrespondents were used to analyze, as well as adjust for nonresponse. Response propensity was investigated by using descriptive statistics and logistic regression analyses. The survey data was then weighted by using inverse propensity weights. Two specific examples of survey estimates are addressed, namely driver vigilance and driver's distraction just before the accident. The results from this paper reveal that driver age and accident type were the most influential variables for nonresponse weighting. Driver gender and size of town where the driver resides also had some influence, but not for all survey variables investigated.

The main conclusion of this paper is that nonresponse weighting can increase confidence in accident data collected by a mail survey, especially when response rates are low. Weighting has a moderate influence on this survey, but a larger influence may be expected if applied on a more diverse driver population. The development of auxiliary data collection can further improve accident mail survey methodology in future.

1. Introduction

1.1. Background, objectives and research questions

Statistical accident data plays an important role for traffic safety development and has three main applications in road safety development: (1) making priorities bases on how frequently different safety issues occur and the consequences of when they occur, (2) making effect analysis of safety improvements that are yet to be implemented, and (3) verifying real world performance measured by the change in number of accidents or personal injuries once changes are implemented in real traffic (Isaksson-Hellman and Norin, 2005; Vaa et al., 2007). During the last few decades, real world accident data has formed an important part of traffic safety development around the world. Based on this data, improvements have been made to infrastructure, vehicles, and driver education. Accident data collected by mail surveys and in-depth investigations have successfully been used within the Swedish vehicle industry since the early 1970s as a part of the product development process, which has since substantially reduced the amount of personal injuries (Isaksson-Hellman and Norin, 2005). Accident data is a driving force in vehicle product development and guides strategic priorities, requirements and physical or virtual verification methods. Historically, vehicle safety development has mainly been concerned with injury prevention during collisions. In more recent years, product development has been directed towards driver behaviour and accident causation as well. In addition, low response rates has lead to concerns on whether estimates from a mail survey can be trusted as a source for making strategic decisions. This brings on new methodological challenges when collecting statistical accident data by using a questionnaire.

Analysis and compensation for nonresponse bias is a well established part of mail surveys in general. However, when using mail surveys to collect statistical accident data, nonresponse analysis is commonly not included as an integral part of the analysis. Sagberg (1999, 2001) presented an interesting study where accident data based on mail surveys was analyzed specifically for mobile phone use and tired drivers. Nonresponse analysis was not performed since background data for nonrespondents was not

* Corresponding author at: Vehicle Safety Division, Department of Applied Mechanics, Chalmers University of Technology, Lindholmspiren 3, SE-412 96 Göteborg, Sweden. Tel.: +46 (0)31 325 1743.
E-mail addresses: etiveste@volvocars.com, emma.tivesten@chalmers.se (E. Tivesten).

0001-4575/$ – see front matter
doi:10.1016/j.aap.2012.02.017

available. Sagberg points out that accident questionnaires with guaranteed anonymity, distributed by another party than the police or insurance companies, hold some advantages to police reported accident data that tends to suffer from underreporting. Performing nonresponse analysis and adjustments, if required, strengthens mail surveys as a tool for collecting accident data.

The main objective of this paper is to investigate nonresponse bias in a survey sent to drivers that have been involved in a car accident by using auxiliary variables. The auxiliary variables contain information (e.g. driver age, gender, etc.) on all mail survey recipients and are obtained from another source than the mail survey questionnaire. Two specific examples of survey estimates are addressed, namely driver vigilance and driver's distraction just before the accident. These are hot topics in the area of road safety, and research has been conducted based on both accident data and driving data (Doherty et al., 1998; Horne and Reyner, 1999; Sagberg, 1999, 2001; Lam, 2002; Klauer et al., 2006).

This paper addresses the following questions in the context of accident mail surveys:

(1) What is the relationship between response propensity and the auxiliary variables that may be important for the survey estimates?
(2) What is the influence on survey estimates from the auxiliary data when the received data is weighted?
(3) What individual auxiliary variables are influential for nonresponse weighting when considering both response propensity and the relationship to the survey data?

In addition, the propensity weight calculation was also investigated in terms of how well the models generalise to a different sample within the same population.

1.2. General knowledge about nonresponse in mail surveys

Mail survey is a widely used method for collecting statistical data. A large number of persons can be reached over a wide geographical area at a relatively low cost (Dillman, 1991; Macdonald et al., 2009). However, the response rate for surveys has declined in developed countries during the last decades, causing a growing concern for survey error (De Leeuw and De Heer, 2002).

Four different sources of error need to be considered before survey estimates can be generalised to the population of interest, (1) sampling error, (2) noncoverage error, (3) nonresponse error, and (4) measurement error (Dillman, 2007). While the first two deal with the sampling procedure, and the last with how respondents interpret and answer the questions, this paper is focused on the third type of error, i.e. nonresponse error.

Nonresponse error is described by Dillman (2007) to occur when "a significant number of the people in the survey sample do not respond to the questionnaire and have different characteristics from those who do respond, when these characteristics are important to the study". Similarly, Groves (2006) states that: "nonresponse bias occurs as a function of how correlated response propensity is to the attributes the researcher is measuring". Nonresponse can lead to bias which in turn can severely harm the quality of the calculated statistics, resulting in biased survey estimates (Dillman et al., 2002; Särndal and Lundström, 2005).

There is a vast body of literature on nonresponse in mail surveys, including nonresponse adjustment techniques (Särndal and Lundström, 2005). There are two strategies for reducing nonresponse bias, either by reducing the nonresponse rates, or by post survey adjustments such as weighting (Moore and Tarnai, 2002; Macdonald et al., 2009). Although the most common advice to decrease nonresponse bias in survey research is to increase the response rates, recent research suggests that there is not a strong relationship between response rates and nonresponse bias (Groves, 2006; Olson, 2006).

A method of calculating nonresponse weights by using logistic regression is described by Brick and Kalton (1996) and Iannacchione et al. (1991) and is applied to this paper. Auxiliary variables need to be available for both the respondents and nonrespondents, and are obtained from another source than the questionnaire. A logistic regression model estimates response probability for each case in the dataset which is then applied as inverse probability weight for the same case. There is, however, a risk of overfitting the data with this approach (Babyak, 2004). The model may represent the dataset perfectly, but have limited prediction if applied on another sample from the same population.

Weighting is usually described as efficient in reducing unit nonresponse bias while commonly increasing variance in survey estimates (Brick and Kalton, 1996; Little et al., 1997; Kreuter et al., 2010). However, if the auxiliary data is highly related to both response propensity and the survey variables, nonresponse bias can be reduced without increasing the variance in survey estimates (Little and Vartivarian, 2005; Särndal and Lundström, 2005; Kreuter et al., 2010; Kreuter and Olson, 2011). Nonresponse weighting assumes that data is missing at random, i.e. respondents and nonrespondents have the same response propensity and distribution of survey measures within groups sharing the same values for the auxiliary variables (Brick and Kalton, 1996; Groves, 2006). In reality, this assumption can usually not be verified due to a lack of additional data.

Many variables which influence response propensity have been identified in other research areas, e.g. gender, age, urbanicity, socioeconomic status, questionnaire volume, trust in the sponsor, interest in the topic, or feeling threatened/embarrassed by the topic (Etter and Perneger, 1997; Krosnick, 1999; Durrant, 2006; Groves and Peytcheva, 2008). To our knowledge, the question of which auxiliary variables are suitable for developing nonresponse weights for traffic accident surveys is yet to be fully addressed. While many variables have been identified as important to accident investigations, such as light conditions, state of the road, weather, speed limit, driver's age and gender, little is known about their relevance in terms of influencing response propensity and survey estimates in accident data collection.

2. Material and methods

This chapter contains information about the mail survey structure and distribution followed by a description of the auxiliary variables obtained from another source than the mail survey. Two sections describe the data analysis. First, the analysis of response propensity is described. Second, the analysis of nonresponse weighing of mail survey estimates for driver vigilance and distraction is presented. The relationship between the mail survey estimates and different auxiliary variables is also included in the analysis.

2.1. The accident mail survey

Volvo Cars' Accident Research Team collected accident data by using a mail survey questionnaire as well as performing in-depth investigations. The questionnaire was distributed to the owners of Volvo cars that had vehicle repair costs above 45,000 SEK following an accident, and were insured by Volvia insurance company. In Sweden, Volvia covers 100% of all Volvo cars up to three years of age, due to the new vehicle guarantee, and approximately 40% of all Volvo cars older than three years. The questionnaire was distributed to owners of vehicle models that were produced during 1990 or later, since reasonably modern vehicles are of interest in this survey. Rental vehicles and non-traffic damage, e.g. parking