Accident Analysis and Prevention 62 (2014) 294–302

Contents lists available at ScienceDirect

Accident Analysis and Prevention

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

Crash reconstruction and crash modification factors

Gary A. Davis*

Dept. of Civil Engineering, University of Minnesota, 122 CivE, 500 Pillsbury Drive SE, Minneapolis, MN 55455, USA

ARTICLE INFO

Article history:
Received 24 June 2013
Received in revised form
23 September 2013
Accepted 25 September 2013

Keywords:
Crash reconstruction
Crash modification factors
Crash simulation

ABSTRACT

This paper addresses the following question: Under what conditions can reconstructed road crashes be used to estimate the effect of a safety-related countermeasure? Results developed by Pearl and his associates are used to draw two main conclusions. First, when one can (1) identify a structural equation describing a type of crash, (2) identify an additional structural equation describing the countermeasure's impact, and (3) estimate the initiating conditions for a set of reconstructed crashes, then a lower bound for a crash modification factor can be estimated by simulating whether or not each of the reconstructed crashes would still have occurred had the countermeasure been present. If the countermeasure's effect is monotonic this bound becomes tight. Second, in situations where it is not possible to reliably identify the structural equations needed for simulation, but where one can (1) identify a set of crash inputs which, when given, make the crash outcome conditionally independent of the countermeasure, and (2) predict how the distribution of these inputs will change in response to the countermeasure, then nonparametric estimation of the countermeasure's crash modification factor is possible. When it is not possible to predict the countermeasure's effect on the conditioning variables it may still be possible to identify constraints or specifications which the countermeasure should satisfy in order to realize a target crash modification.

1. Introduction

It is well-known that planning and designing engineered systems often requires quantitative predictions of how the systems will perform under different design options. Observed statistical associations are not generally sufficient, what is required is "...causal theory sufficiently developed as to permit prediction." (Webber, 1983). For highway safety, important performance measures are the frequency and severity of crashes, which are affected by the designs of roads, by the designs of the vehicles on the roads, and by the behavior of road users. Shinar (2007, 2012) described three general approaches to investigating causation in road safety, distinguished by the type and resolution of the data employed. The first, which Shinar called "theory-based clinical case analysis," relies on detailed investigation and reconstruction of individual crashes, followed by expert assessment of whether or not specific features could be considered causal factors. The exemplar for this is the Tri-Level Study (Treat et al., 1979). The second approach, which Shinar called "statistical 'theory-free' data-base analysis," uses statistical methods to identify associations among variables recorded in crash databases, such as those compiled from police crash reports. The extent to which an observed association is accepted as causal then depends on the extent to which alternative explanations for the association can be rejected. Shinar's third approach, "prospective in-vehicle monitoring of driver behavior," relies on data collected in naturalistic driving studies, such as the 100-Car Study (Dingus et al., 2006). Here, volunteers allow their driving to be monitored continuously, leading to more complete records of the events preceding crashes and near crashes than are usually available from retrospective studies. Shinar also makes an interesting point relevant to clinical and naturalistic studies. "The analyses often reveal the inappropriate behaviors that preceded the crash and made it inevitable, but they do not necessarily identify the most appropriate countermeasure. They can, however, indicate whether a potential countermeasure... would eliminate or reduce the effect of a specific cause..." (2007, p. 723).

Shinar's three approaches, clinical, statistical, and naturalistic, are all observational, and their support for causal conclusions will generally be weaker than what could be obtained with experimental control. Still, observational studies do on occasion produce reliable causal conclusions, and a substantial effort has been devoted to identifying conditions which justify causal conclusions from observational data. The resulting literature is extensive, but arguably the most comprehensive treatment is that by Pearl and his associates (Pearl, 2000/2009). Applying concepts developed in Pearl (2000/2009), Davis (2002) outlined a framework within which clinical and statistical safety studies could be seen as addressing the same underlying phenomena, and two recent developments suggest that this topic is worth revisiting. First, the *Highway Safety Manual* (HSM) (AASHTO, 2010) now offers a set of

* Corresponding author. Tel.: +1 612 625 2598; fax: +1 612 626 7750.
E-mail address: drtrips@umn.edu

0001-4575/$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.aap.2013.09.027

tools for predicting how crash frequency or severity might respond to changes in road design. Second, advances in data collection technology now make it possible to collect information on individual traffic events at much finer levels of detail. These include both site-based methods (Davis and Swenson, 2006; Laureshyn et al., 2010; Saunier et al., 2010) and the SHRP 2 Naturalistic Driving Study (NDS). This leads to the interesting question of how detailed, individual data might be used to support enhancements to the HSM, and the assessment of safety countermeasures more generally. Arguably, this requires understanding how the macroscopic notion of causation used in the HSM relates to causation as encountered in individual events.

The remainder of this paper attempts to answer the following question: given a sample of individual crashes for which detailed information is available, how might this be used to estimate a crash modification factor (*CMF*), as defined in the HSM? The focus is on warrants for what can be called the identifiability of the *CMF*, by which is meant sufficient conditions implying that an estimator of the *CMF* converges to its objective, as the sample size becomes arbitrarily large. The warrants thus identify situations where estimation of the *CMF* is possible, at least in principle. This should be regarded as a first step towards an operational methodology, the next step being to develop, for those situations where estimation is possible, methods for characterizing the uncertainty associated with the estimated *CMF*s and for testing hypotheses about the *CMF*s. The results presented here draw heavily on Pearl (2000/2009). Section 2 reviews statistical and clinical treatments of causation while Section 3 contains the paper's main results. Section 4 then illustrates these results using simulated crashes between vehicles and pedestrians, and Section 5 presents conclusions and suggestions for further work. Although this paper focuses on crashes, the results can apply to other well-defined events, such as near crashes, or critical events in freeway shockwaves (Davis et al., 2012).

2. Statistical and clinical approaches to crash causation

2.1. Statistical estimation of crash modification factors

The methods used to develop the HSM fall under Shinar's "statistical, 'theory-free' data-base analysis," with Chapter 9 of the HSM providing guidance on how to estimate the causal effect of a safety-related countermeasure. The recommended approach, called empirical Bayes before–after analysis, uses generalized linear models to control for at least some possible confounding factors, such as changes in traffic volumes, and uses special adjustments to correct for the bias that can occur when crash history is used to select sites for treatment. The result is an estimate of a multiplier, called a crash modification factor (*CMF*), which can then be used to predict the effect of the countermeasure at new locations.

The statistical methods used to develop the HSM rely on the computerized crash and roadway databases compiled by government agencies, and can produce useable results when little is known about how crashes occur or how a countermeasure affects crashes. This statistical approach is justified by assuming that crash events are "random and unpredictable" (AASHTO, 2010, pp. 3–5). In this view, conditions on a particular road generate a propensity, which can be modeled as a probability, for road users to have crashes (Hauer, 1982; Lord et al., 2005). This propensity is assumed to produce stable relative frequencies of crash and non-crash events, but explaining why a specific road event resulted in a crash is no more meaningful than explaining any other inherently random event, such as the decay of a particular atom of Uranium-238 at a particular time. This is a powerful assumption, which not only justifies the statistical tools used to develop the HSM, it also justifies using the mathematics of random variables as the theoretical language of road safety.

2.2. Clinical assessment of causation

Unlike subatomic events, it is sometimes possible to investigate and reconstruct individual crashes, leading to Shinar's "theory-based clinical case analysis." As noted above, a well-known example of the clinical approach is the Tri-Level Study conducted in the 1970s (Treat et al., 1979). In this study, individual crashes were investigated and reconstructed, and an interdisciplinary team which included a human factors psychologist, an automotive engineer, and a crash reconstruction specialist then reviewed a list of possible causal factors, rating each with respect to ". . .had the factor not been present in the accident sequence the accident would not have occurred" (Treat et al., 1979, p. A-5). That is, the team assigned an ordinal plausibility to a counterfactual conditional, for each crash and relevant factor, where the assignment was based on subjective, albeit expert judgment.

More recently, researchers at the University of Adelaide's Center for Automobile Safety Research (CASR) have used a clinical approach to study vehicle speed as a causal factor in fatal and severe crashes (McLean et al., 1994; Kloeden et al., 1997). To illustrate, Fig. 1 shows a scene diagram for case 89-H002 from McLean et al. (1994 vol. 2, p. 207), where a child pedestrian came from behind a parked car, attempted to cross a street, and was struck by a moving vehicle. The assumed scenario can be described as follows. The car's driver, traveling at an initial speed denoted by s, noticed the encroaching pedestrian when at a distance d from the point of collision. After a perception/reaction time (r) the driver began braking at rate fg, where g is the gravitational attraction and f is a braking drag factor which expresses braking deceleration in units of g. In this case the stopping distance exceeded the initial distance, i.e. $sr + s^2/(2fg) > d$, and the crash occurred.

A primary emphasis of the 1994 CASR study was on the potential for speed reductions to lower the incidence of pedestrian fatalities. Referring to Fig. 1, the investigators were able to identify the point at which the pedestrian was stuck (A), the point at which the pedestrian's body came to rest (B), and a skidmark left by the vehicle's tires as the driver braked to a stop. The measured skidmark length was 23.5 m and equating the drag factor to a tire/pavement friction coefficient of $f = 0.72$, together with an assumption that 20% of the vehicle's kinetic energy was dissipated before the tires began making skidmarks, led to an estimated initial speed for the vehicle of 73 km/h. Next, assuming that the driver's perception/reaction time was $r = 1.5$ s and that the vehicle traveled 9.6 m between the application of the brakes and the start of the skidmark, put the driver's point of perception at $d = 55.4$ m from the collision point. Had the driver been traveling at the 60 km/h speed limit instead of 73 km/h the stopping distance would have been about 45 m, and so one could say that speeding was a causal factor for this crash.

Several comments are in order. First, as in the Tri-Level study, conclusions regarding the importance of causal factors were based on the plausibility of counterfactual conditionals, such as "If the driver had been traveling 60 km/h, other things being equal, the crash under consideration would not have occurred." However, the Tri-Level Study's subjective assessment of plausibility was replaced by a deterministic equation, and the counterfactual conditional was judged either true or false depending on an evaluation of that equation. Expert judgment was still needed, most obviously in the selection of the deterministic equation used to describe the crash. In addition, the information available about the crash was not sufficient to identify values for all of the equation's input variables, and the missing information was provided as point values, based on expert judgment.