SciencDirect Journals & Books Sign in Register
COVID-19 coverage: discover and explore the growing or relevant literature. Access to subscribed content

Transport Policy Volume 99 Pages 1-9 2020
Analytic Methods in Accident Research
Review Article Open access

Big data, traditional data and the tradeoffs between prediction and causality in highway-safety analysis

Fred Mannering, Chris R. Bhat, Naveen N. Pinjari, Venu M. Garikapati

Cite this article as:
https://doi.org/10.1016/j.tranpol.2020.06.011 Get rights and content

Highlights

- Prediction and causality tradeoffs are considered.
- Four distinct modeling approaches to safety data evaluated.
- Concerns with using data from instrumented accelerated and non-accelerated vehicles are discussed.
- The consequences of prediction/causality tradeoffs on safety policy are considered.

Abstract

The analysis of highway accident data is largely dominated by traditional statistical methods (standard regression based approaches), advanced statistical methods (such as machine learning and Bayesian approaches), and more recently, big data analytics (artificial intelligence, natural networks, machine learning, and so on). These methods have been applied mostly using data from observational sources (such as police accident reports, roadway inventories, and so on) that are inherently rich in data but incomplete as a whole due to not being represented in the data. In addition, when and where individuals choose to operate their vehicles (and the manner in which they drive) is influenced by unobserved factors (such changes over time. This here, the nature of the data, and the implementation target of the analysis imply that analysts must often make tradeoffs between prediction and causality. This paper discusses the concerns the underlying causal nature of such tradeoffs may have. The selection of this analysis method is often made without full consideration of the implications that the tradeoff between prediction and causality has on the development of safety countermeasures and policies. This paper provides a discussion of the issues involved in this tradeoff with regard to specific modeling approaches, and it provides a discussion of the impact of understanding of the trade-offs when being informed by results from analysis.

Previous article in issue Next article in issue

Keywords

Self-selectivity; Endogeneity; Highway safety; Accident Unobserved Accident risk

1. Introduction

The implicit assumption in traditional statistical analyses is that an appropriately estimated model will both uncover cause-and-effect and the underlying causal mechanisms. Although this is a reasonable assumption in the ideal case of data collection, and in some of causality in traditional statistical modeling, suggest that analysts must often, even if not always, be concerned with the tradeoff between prediction and causality. Specifically, causality, that is, the cause-and-effect, may not be the best or even strong causality, and results that are of a strong causality may not be the best or even strong prediction.

There are three general methodological approaches that are potentially suitable for the analysis of transportation safety data: traditional statistical methods, endogeneity/heterogeneity models, data driven methods, and causal inference methods. Each of these methods has its own strengths and weaknesses in prediction accuracy and their ability to uncover causality in transportation safety. Traditional statistical methods, such as those in the Highway Safety Manual (AASHTO, 2010), are well-established and well-understood methods, and they are readily available to most analysts (sometimes in ways that allow for various safety improvements to accident risk. The traditional literature (such as the Highway Safety Manual) has provided a wealth of information on causality and causal explanations, but generally lacks fundamental support that these claims to assessments of prescriptive (but, for example, potential biases in the data and unobserved factors). Although, the well-established nature of traditional highway-safety modeling has significantly found in assessment of aggregate crash (total count of accidents for example), and there is ample evidence that such analysis has been useful in the development of safety countermeasures and policies. This paper provides a discussion of the issues involved in this tradeoff with regard to specific modeling approaches, and it provides a discussion of the impact of understanding of the trade-offs when being informed by results from analysis.

Endogeneity and heterogeneity models, such as those that might be available in the analysis of transportation safety data, are well-established methods (such as endogeneity/heterogeneity models, data driven methods, and causal inference methods).