

CAUSAL INFERENCE IN TRANSPORTATION SAFETY STUDIES: COMPARISON OF THE POTENTIAL OUTCOMES AND CAUSAL BAYESIAN NETWORKS

BY VISHESH KARWA^{†,*}, ALEKSANDRA B. SLAVKOVIĆ^{*} AND ERIC T. DONNELL[†]

Department of Statistics, The Pennsylvania State University^{*}
Department of Civil Engineering, The Pennsylvania State University[†]

The research questions that often motivate transportation safety studies are causal in nature. Safety researchers typically use observational studies to answer such questions. Elucidation of causal relationships from observational studies must be shaped by assumptions about how the data are generated. The field of causal inference presents several modeling frameworks for formulating these assumptions, and tools for probing empirical data to assess causal relations. This paper focuses on exploring the applicability of two such modeling frameworks - Causal Bayesian Networks and Potential Outcomes - for a specific transportation safety problem. Nighttime crash data from three districts in North Carolina, appended to pavement marking retroreflectivity data, are used. The basic assumptions of the two causal modeling frameworks are outlined, and causal effects of pavement marking retroreflectivity on safety of a road segment are estimated. The results of both modeling frameworks are generally consistent with each other. The effect of increased pavement marking retroreflectivity was generally found to reduce the probability of target nighttime crashes.

1. Introduction. An estimated 2.2 million people suffered some kind of transportation-related injury in 2007. About 87 percent of these injuries resulted from highway crashes ([Bureau of Transportation Statistics, 2007](#)). Transportation safety management aims at identifying causes of such crashes, developing countermeasures to mitigate crashes, and evaluating the effectiveness of a safety countermeasure. It is well known that causal propositions of this kind, and effect sizes of countermeasures, are best estimated from randomized experiments.

Experimental study is a tool that is not always accessible to transportation safety researchers. The types of data available in transportation safety studies are primarily observational, which makes it difficult to consistently estimate the causal effects of countermeasures. In this paper, we evaluate

Keywords and phrases: Causal Inference, Potential Outcomes, Causal Bayesian Networks, Observational Studies, Transportation Safety, Nighttime Crash Data.

and compare the application of two commonly used causal inference frameworks (one that is commonly applied in computer science and another that is commonly applied in statistics) to transportation safety. In particular, the aim of this paper is two-fold:

- To introduce a unique transportation safety dataset, created from multiple sources, and to highlight the problems associated with the data used in safety studies.
- To explore the application of causal inference methods in transportation safety studies and document the issues associated with the analyses. We do this by estimating the causal effect of pavement marking retroreflectivity on nighttime accidents using the causal modeling settings of the Potential Outcomes (PO) and the Causal Bayesian Networks (CBN), and then compare the results.

Causal inference methods in transportation safety studies have received very little attention. [Davis \(2000\)](#) provides a review and notes that the assignment mechanism must be included in statistical models to consistently estimate the effect of any countermeasure on accidents. [Davis \(2004\)](#) uses Pearl’s causal Bayesian networks for crash reconstruction and examines token causal claims to answer single-event causation questions; see [Eells \(1991\)](#) for a review of token and type causes. In contrast to single-event causation, our work examines the application of causal inference methods to population level causal effects in transportation safety studies. Population causal claims are more applicable to transportation safety management since they reflect the effect of countermeasures in a population as opposed to singular causal claims which are geared more towards accident reconstruction and liability issues. We examine the hypothesis that low Pavement Marking Retroreflectivity (*PMR*) levels are causative agents for an increase in the risk of nighttime accidents. Causal effects are estimated using the potential-outcomes framework and causal Bayesian networks, and the results are compared.

It has been shown that the causal effects estimated by the potential-outcome model (PO) and the causal Bayesian networks (CBN) are mathematically equivalent to each other, under certain conditions (see [Pearl \(2000, Chap 7\)](#), [Sfer \(2005\)](#) and [Fienberg and Sfer \(2006\)](#)). Yet, both methods offer several different paths to estimate the effect, and in practice the results may not be similar. Apart from the common assumptions relating to the estimation of a causal effect from an observational database (which will be reviewed in the respective sections), each method requires additional parametric assumptions inherently tied to the effect estimation procedures. For instance, the CBN method requires the data to be either discrete or Gaus-

sian. While the CBN setting enables one to work with the complete data set, the PO approach could lead to elimination of a part of the data set. Sfer (2005) and Fienberg and Sfer (2006) also show on a simple simulated logistic regression example that there is an implicit agreement between PO and CBN frameworks. However, we are not aware of a study that compares the results of both modeling frameworks on real data, and examines the advantages and disadvantages associated with applying each method in a practical context.

The remainder of the paper is organized as follows. Section 2 introduces the problem statement. Section 3 introduces the dataset and the design of the hypothetical experiment to estimate the causal effects. Sections 4 and 5 present the analyses and results of the Potential Outcomes framework and the Causal Diagrams framework. Section 6 compares the results and Section 7 concludes with a discussion.

2. Problem Description. The fatality rate by traffic accident fatality rate increases by almost 75 percent during the period of time between 9 p.m. and 6 a.m. (National Highway Traffic Safety Administration, 2007). This raises a question of what can be done about the fact that there is a greater rate of accidents at night than during the day? Because the ratio of night-to-day traffic fatalities is nearly 2:1, the specific causes for this unbalanced ratio should be found. It is hypothesized that low pavement marking visibility may be one cause of the increased rate of nighttime crashes. One of the objectives of the paper is to examine this hypothesis.

Pavement markings delineate the limits of the traveled way and provide drivers navigation and control guidance. During the daytime, drivers are likely to use a combination of pavement markings, other traffic control devices (e.g., signs), and visual cues along the roadside (e.g., utility poles, vegetation, etc.) to navigate a roadway. Pavement markings have an important role at night. Apart from delineating the road, pavement markings reflect the light shone from a car’s headlamps back to the driver, thus enabling the driver to see the limits of the traveled way. This is known as retroreflectivity and is measured in millicandelas per square meter per lux ($mcd/m^2/lux$). Retroreflectivity in pavement markings is provided by glass spheres that are dropped-on or premixed with a wet pavement marking material. *Pavement marking retroreflectivity (PMR)* degrades over time because of fatigue to the material and its bond strength with glass spheres or the pavement surface. State transportation agencies typically re-stripe pavement markings after the end of their useful service life, defined as the time when retroreflectivity falls below a minimum threshold level.

A considerable amount of research has been carried out regarding the safety benefits of *PMR*. To answer the question if improving *PMR* has any effect in reducing the number of traffic crashes, most of the published literature used regression models with observational data, ignoring the treatment assignment mechanisms; for a literature review, refer to [Bahar *et al.* \(2006\)](#) and [Donnell, Karwa and Sathyanarayanan \(2009\)](#). Also, as mentioned in [Donnell, Karwa and Sathyanarayanan \(2009\)](#), none of the studies explicitly relate the in-situ *PMR* levels to the crash event. This is due to the fact that *PMR* levels and crash data are obtained from separate sources and merging them is difficult. The problems associated with merging the databases have been described in [Karwa \(2009\)](#). [Donnell, Karwa and Sathyanarayanan \(2009\)](#) was the first study that explicitly combines the *PMR* data (which is representative of real life degradation patterns of *PMR*) with crash data to develop a comprehensive database. This work did show that there were statistical associations between *PMR* and nighttime accidents. We use this database to examine, for the first time, the nature of the effect of *PMR* on traffic safety (defined in Section 3) using the potential-outcome model and causal Bayesian networks. The results are compared with a discussion of the application of the two methods, in an attempt to determine their possible broader application in transportation safety studies.

3. Description of the Data and Design of the Study. The fundamental unit of operation in this paper is a homogeneous road segment; homogeneous refers to having uniform geometric characteristics such as number of lanes, lane width, and shoulder width along a roadway segment. A segment within a fixed time period is considered to be different from the same segment at any other time period. A fixed time period of one month was selected to ensure homogeneity of *PMR* levels and other characteristics of a segment. For instance, the *PMR* level and monthly traffic volumes can be assumed to be reasonably uniform within this period.

Crash and pavement marking retroreflectivity data were collected retrospectively from three districts in North Carolina for a period of 2.5 years. As noted in Section 2, the data were obtained from two different sources. The *PMR* data were measured by a private contractor using a mobile retroreflector with a 30-meter geometry. These data were collected on two-lane and multi-lane highways in North Carolina. All pavement markings were of thermoplastic material. Since retroreflectivity estimates were not measured at the exact time and place of occurrence of the crash, a neural network model was used to interpolate the values of retroreflectivity on the segments where crashes were observed. The details of the model are provided in [Karwa](#)

and Donnell (2009).

The roadway inventory and crash event data were obtained from the Highway Safety Information System (HSIS) data files, maintained by the Federal Highway Administration (FHWA). These data were collected for the 19 roadway sections where retroreflectivity data were collected. There were 192 total segments that corresponded to the 19 sections of roadway where PMR estimates were computed based on the degradation model. Table 1 shows the locations where roadway inventory, crash, and PMR data could be linked. There are a total of 5916 segments (segments are a subset of sections).

TABLE 1
Roadways with Pavement Marking and Crash Data Available for Safety Analysis

County	Route	District	Begin MP	End MP	Number of Lanes	Total Length (miles)	Nighttime Target Crashes
Bertie	US13	1	0.00	11.07	4	11.07	8
Gates	US13	1	0.00	14.78	2	14.78	14
Northampton	US158	1	12.35	24.04	2	11.69	4
Washington	US64	1	10.54	19.67	2	9.13	2
Durham	I-85	5	7.88	14.19	4	6.31	5
Durham	US15	5	3.66	6.56	4	2.90	9
Durham	NC98	5	0.00	11.06	2	9.44a	15
Durham	NC157	5	0.70	3.98	2	3.28	2
Granville	I-85	5	0.00	23.73	4	1.80b	5
Person	US158	5	0.00	22.36	2	16.22	5
Vance	I-85	5	0.00	14.47	4	12.47c	46
Vance	US158	5	0.00	8.96	2	5.94d	1
Wake	I-40	5	6.47	20.19	4/6/8	13.72	67
Wake	NC98	5	0.00	4.55	2	4.55	1
Warren	I-85	5	0.00	9.88	4	9.88	39
Warren	US158	5	12.38	22.93	2	10.55	0
Catawba	I-40	12	13.13	19.67	4	6.54	10
Iredell	I-40	12	0.00	22.76	4	22.76	63
Iredell	I-77	12	14.75	23.75	4	9.00	17
Total						182.03	313

Crashes that satisfy the following criteria, referred to as *target crashes*, were used in the analysis: no weather contributing circumstances; occurred during dusk, dawn, or at night; non-work zone area; no alcohol involvement; dry roadway surface conditions; no roadway contributing circumstances; ran-off-the-road crashes; fixed object crashes (off-road); and opposite- or same-direction sideswipe crashes. It must also be noted that the data are sparse

due to rarity of accidents. Only 6 percent of the total number of segments had more than one target crash during the study period.

TABLE 2
Definition of Variables used in the Study

Variable	Definition
Right Shoulder Width (Rt. Shoulder)	Outer shoulder width in feet on right side of the traveled way
Adjusted Traffic Volume (ADT)	Annual average daily traffic adjusted for the period of a month, vehicles per day
Percentage of Trucks (Pct. Truck)	Percentage of the ADT that consists of heavy vehicles
Average Retroreflectivity (PMR)	Average value of PMR of all markings on a segment ($mcd \setminus m^2 \setminus lux$)
AGE (AGE)	Time (in months) elapsed since the application of pavement markings on a segment
Multilane Indicator (Multilane)	1 if there is more than 1 lane in each direction 0 if there is 1 lane in each direction
Median indicator (Median)	1 if the segment contains a median, 0 if the segment has no median
Safety indicator (Safety)	1 if at least 1 target crash occurred in the segment, during the month, 0 otherwise
Urban indicator (Urban)	1 if the segment is located in an urban area, 0 if the segment is located in a rural area
Speed limit indicator (Low speed)	1 if speed limit less than or equal to 35 mph, 0 if the speed limit is greater than 35 mph
Terrain indicator (Terrain)	1 if the segment is on flat terrain, 0 if the segment is on a rolling terrain
District	0 if the segment is located in District 1 1 if the segment is located in District 5 2 if the segment is located in District 12
PMR	0(Low) if $PMR < 139$ 1(Medium) if $139 < PMR \leq 200$ 2(High) if $200 < PMR \leq 446.53$

Safety of a road segment is defined as a *Bernoulli* random variable, taking the value 1 if there was at least one target crash in the segment during the treatment (or control) application period, and 0 if there was no target crash in the segment. The safety of a segment is stochastic and each segment has a fixed probability p of witnessing at least one target crash, which is assumed to be an inherent property of the road segment. This definition was chosen to ensure the absence of confounders between safety and PMR , based on the past PMR related safety literature. For instance, if it was clear according to the police crash report, that a particular crash occurred due to driving under the influence, such a crash would have been deemed to

occur because of human error, and hence excluded from the current analyses. Similarly, crashes in which weather was a contributing factor (such as heavy snow, or icy road conditions) were also excluded from the analyses. Weather conditions, human errors, etc. are stochastic factors that may cause crashes and not an inherent property of the segment; hence, any crash occurring due to such conditions would fall into the error term of the observed safety of a road segment.

Treatment variable on a segment is defined as the application of *PMR* with levels $\{Low, Med, High\}$; the exact range of *PMR* levels for each class is specified in Table 2. *Control* is defined as application of pavement markings at one of the two remaining levels of retroreflectivity. Out of the total sample size ($N = 5916$), about 36 percent of the segments had *Low* levels of *PMR*, about 46.5 percent had *Med* levels and the remaining segments had *High* levels of *PMR*. The assignment of *PMR* levels is clearly not random.

TABLE 3
Descriptive Statistics of Data

Continuous Variables				
Variable	Mean	Standard Deviation	Min	Max
Lt.Shoulder Width	2.84	4.24	0	13
Rt. Shoulder Width	9.39	4.03	0	14
Adjusted ADT	30383.31	27580.07	1615	114400
Percentage Trucks	14.77	8.54	0	83
Average <i>PMR</i>	227.36	64.46	139.64	446.53
AGE (in Months)	15.52	8.63	1	30
Categorical Variables				
Variable	Percentage			
Multilane	65.38 %			
Median	67.92 %			
Safety	6.36 %			
Urban	45.3 %			
Low Speed	14.84 %			
Terrain	22.31 %			
District 1	21.81 %			
District 5	55.54 %			
Low <i>PMR</i>	36.07 %			
Med <i>PMR</i>	46.45 %			
High <i>PMR</i>	17.48 %			
Number of Observations = 5916				

Apart from the data on *PMR* and the crash counts per month, data on 13 other covariates were collected. Information on the attributes of a segment such as the shoulder width, number of lanes, presence of a median, level of access control and posted speed limit, traffic flow characteristics such

as monthly traffic volumes (hereafter referred to as *ADT*), percentage of trucks, location related variables such as the geographic district in which the segment is located, the urban or rural setting of the segment location, and the terrain type were collected. Definitions of all variables are provided in Table 2 and the summary statistics are provided in Table 3. The data are very sparse which is typical of safety data. For instance, in the five way cross-classification of the entire sample with respect to the discrete variables *District*, *Terrain*, *PMR*, *Multilane* and *Safety*, 58 percent of the cells have sampling zeros.

As per Rubin (2008) and Maldonado and Greenland (2002), we conceptualize our problem as a hypothetical experiment to make the problem statement clear. Consider a population of homogeneous road segments. We wish to examine the effect of increased *PMR* on the safety of a road segment. Ideally, we would like to apply treatment (e.g., *PMR = Low*) and control (e.g., *PMR = High*) to the same population and observe the expected safety outcome to measure the causal effect. The causal effect is defined as the risk ratio of expected safety outcome under the treatment and controls, for the same population. Since this is not possible in practice, we use analytical simulations of this process. Sections 4 and 5 describe conceptualization of this hypothetical experiment under two different frameworks.

4. Potential Outcomes Framework. In this section we present the potential outcomes framework as applicable to the current study as well as the results of the analysis. Section 4.1 defines the causal estimands (the “science”, (see Rubin, 2005)) and the corresponding assumptions. Section 4.2 is concerned with the assignment mechanism and balancing of the data using propensity score matching. Section 4.3 explains the estimation of average causal effect, after matching, using regression adjustment method, and Section 4.4 presents the results of the analysis.

4.1. *Treatment and Potential Outcomes.* Let the homogeneous segments be indexed by the letter i . We focus on one hypothetical experiment at a time, introduced in Section 3, and estimate the effect of a binary *PMR* treatment on safety from a sample of segments. Extension to the case of three levels of treatment of *PMR* is performed using the method proposed by Rubin (1998) which involves creating a separate propensity score model for each two-level treatment comparison, equivalent to conducting three hypothetical experiments. Thus, in the present case, three separate propensity score models would be estimated. This method is followed since it is difficult to simultaneously balance all three treatment groups on all covariates.

The PO model uses potential outcomes as the fundamental element to

estimate the causal effects. We denote the treatment variable by T_i , where $T_i = 0$ denotes no treatment or the baseline condition for unit i , and $T_i = 1$ denotes the treatment condition. For instance, if we wish to estimate the effect of changing the *PMR* levels from *Med* to *Low*, the treatment would be application of *PMR = Low* and the control would be application of *PMR = Med*. Associated with each segment are two potential outcomes: *Safety*(S) of the segment at the end of a month after the treatment has been applied, $S_i(T = 1)$, and *Safety* of the same segment at the end of the month if there was no treatment (i.e., the baseline condition was applied, $S_i(T = 0)$). Covariates that represent the attributes of a segment are denoted by the vector $X_i = (x_{i1}, x_{i2}, \dots, x_{ip})$ for unit i . The causal effect of the treatment relative to the baseline for segment i is then defined as the ratio of $E[S_i(1)]$ and $E[S_i(0)]$ where $E[\cdot]$ denotes expectation. We assume that $E[S_i(0)] > 0$, and drop the T in the notation for simplicity.

For any particular segment, only one of the two values of $S(0)$ and $S(1)$ can be observed. This has been termed the “fundamental problem of causal inference” (Rubin, 1978; Holland, 1986), because of which unit level causal inferences are not possible¹. However the average causal effect of the treatment on a population can be estimated consistently from a randomized experiment under certain assumptions. The assumptions required for estimating average causal effect are given below.

Stable unit treatment value assumption (SUTVA) (Rubin, 1990). This assumption states that the treatment applied to one unit does not affect the outcome of any other unit and that there are no hidden versions of the treatment (i.e., no matter what mechanism was used to apply the treatment to the unit, the outcome would be the same). The last part is sometimes referred to as the consistency assumption (Cole and Frangakis, 2009). We make this assumption in the current study even though the treatment has been applied in groups (several segments along a particular route may have the same value of *PMR*). The following example illustrates a scenario where this assumption could be violated. Consider two consecutive segments on the same route. A vehicle traveling on this road could end up in a crash in segment 2 because of low visibility on segment 1. Such scenarios are not uncommon, but when accidents are reported, the reporting officer estimates the approximate segment location where the crash was initiated (based on a subjective assessment of skid marks of the vehicle and other crash specific parameters) and the crash is attributed to that segment. Moreover, most segments are long enough to ensure that any crash attributed to that seg-

¹Except in cases where the mechanism of causation is known, this is called token causation, see Pearl (2000, chap. 7)

ment originated in the same segment. [Hong and Raudenbush \(2005\)](#) extend SUTVA to account for possible interference among segments, but we do not consider this extension here.

Positivity. The positivity assumption states that there is a nonzero probability of receiving every level of treatment for every combination of values of exposure and covariates that occur among individuals in the population ([Rubin, 1978](#); [Hernan and Robins, 2006](#)). We make the positivity assumption since, in principle, each segment can be assigned any level of *PMR* treatment.

Unconfoundedness. The treatment mechanism is said to be unconfounded given a set of covariates x_i , if the treatment is conditionally independent of the potential outcomes given the covariates:

$$(4.1) \quad t_i \perp\!\!\!\perp S(0), S(1) \mid x_i.$$

In a randomized experimental setting, t_i would be unconditionally independent of the potential outcomes by design. In the current setting this is not the case, but the treatment assignment can be made independent of the potential outcomes by balancing on observed covariates. This is done using propensity score matching and is discussed in [Section 4.2](#).

4.2. Assignment Mechanism, Propensity Score, Matching and Balancing. The second part of the PO framework is to formulate the propensity scores used in assignment of the treatment to the segments and to match on the propensity scores to achieve balance as follows:

Let $P(T_i = 1 \mid X_i)$ be the propensity score. Following [Rosenbaum and Rubin \(1983\)](#), treatment assignment is strongly ignorable given a vector of covariates X if unconfoundedness and common overlap hold:

$$(4.2) \quad S(0), S(1) \perp\!\!\!\perp T \mid X$$

$$(4.3) \quad 0 < P(T = 1 \mid X) < 1$$

In the current setting, the treatment assignment mechanism can be assumed to consist of two parts. In the first part, pavement markings are applied by transportation agencies at different segments with a similar level of retroreflectivity (usually falling into the category *High*). In the second part, the markings are left to deteriorate over a period of 2.5 years. The *PMR* levels decrease due to stress on the pavement marking material from vehicle passes and natural factors such as weather. Thus it can be assumed

that nature assigns a level of PMR based on the time elapsed since the initial application period (AGE) of the pavement marking, number of vehicle passes and weather conditions. The assignment of PMR levels for each segment depends on the AGE of the marking within the segment and the number of vehicle passages over the segment within that period. Apart from this, PMR levels may also depend on the location of the segment (due to differences in weather conditions), the percentage of trucks that compose the traffic volumes (stress on the marking material is generally greater due to the heavy vehicles when compared to a passenger car), the number of lanes in a segment (the PMR levels used are the average of the different pavement markings present in a segment, and multi-lane segments generally have at least one extra marking when compared to two lane segments). All of these variables are included to form a rich propensity score model that specifies the assignment mechanism.

As noted earlier, three different propensity score models are estimated using binary logistic regression by selecting two out of the three treatment levels $\{Low, Med, High\}$ of PMR . As per [Rosenbaum \(1995\)](#), all covariates with a p-value of less than 0.5 are retained in the propensity score model. Balance is achieved by matching the treatments to controls on the estimated propensity scores. [Rosenbaum and Rubin \(1983\)](#) demonstrate that balance on the distribution of covariates can be achieved by balancing on the scalar value of propensity scores. For variables used in three propensity score models see [Table 4](#).

TABLE 4
Variables in Propensity Score Models

Variable	Comparison Level		
	Med to Low	High to Med	High to Low
Dist1	✓		✓
Dist5	✓		
Urban	✓	✓	✓
Terrain	✓		✓
Median	✓		✓
Multilane		✓	
Rt. Shoulder	✓		✓
Pct. Truck	✓	✓	✓
AGE	✓	✓	✓
ADT	✓	✓	✓*
Sample Size	4882	3782	3168

*log transformation used

Nearest Neighbor matching ([Dehejia and Wahba, 2002](#)) on the logit propensity scores was carried out using a caliper of 0.25 and the groups are tested

for balance. The software developed by [Leuven and Sianesi \(2003\)](#) was used. In this method, if e is the estimated propensity score, then for each individual in the treatment group, a pool of potential matches in the control group is identified whose logit propensities are in the interval $e \pm c$, where c is the value of the caliper selected. Within this pool, the closest individual in terms of the Mahalanobis distance of propensity score is selected. Matching is done without replacement. If balance is not achieved, the propensity score model is re-specified and the process is repeated. The balance between treated and comparison groups for all measured covariates is tested using an unpaired t -test. The distribution of propensity scores across the two groups is also examined, since comparison between one variable at a time does not take into account the correlation among the covariates ([Cochran, 1965](#)). Observations that were out of the common support of the propensity scores were discarded to ensure sufficient overlap of distributions of key covariates for the treatment and control groups. Sufficient overlap is needed to produce stable estimates of the Average Causal Effect (ACE) ([Dehejia and Wahba, 1998](#); [Rubin, 1997](#)). Specific details are provided in the *Results* section 4.4.

4.3. *Estimation of the Average Causal Effect.* Once the samples are divided into control and treatment groups and balance is achieved, several methods exist to estimate the ACE of the treatment ([Schafer and Kang, 2008](#)). We follow the regression estimation with propensity score covariates strategy suggested by [Little and An \(2004\)](#) and [Schafer and Kang \(2008\)](#).

One can think of the use of regression estimation of the potential outcomes serving two different purposes. In the first scenario, regression is used as a tool to mimic stratification or conditioning on the covariates of a treated or control population. In this case, the conditioning must be performed on the same covariates for both the treatment and the control groups. In the second scenario, regression is used simply as a prediction tool. Thus, the policy to be followed when estimating the regression equations for each of the control and the treatment groups is to ensure that the equations possess generalization capabilities. In this scenario, it may be the case that including the same covariates in both regression models violates this policy due to overfitting of the data. In this study, we follow the second scenario, that is we consider parsimonious regression models that help predict the safety outcome under treatment or control.

We use binary logistic regression to estimate a probability distribution of the safety outcomes for the treated segments. Similarly, for the control segments. Only those segments in the matched sample are used that fall under the common support. We classify the segments into five strata with bound-

aries defined by the 20th, 40th, 60th, and 80th percentiles of the estimated propensity scores. Categorical variables representing the strata are included in the regression model along with other covariates. The regression models are given by the following equations:

$$(4.4) \quad \text{logit}(P(S_i(1) = 1)) = \mathbf{x}_i\boldsymbol{\beta}_1$$

$$(4.5) \quad \text{logit}(P(S_i(0) = 1)) = \mathbf{x}_i\boldsymbol{\beta}_0$$

The regression equations are used to predict the probability distributions of $S_i(1)$ and $S_i(0)$; see Table 5 for significant covariates. As noted in Section 3, the ACE is estimated as a causal risk ratio:

$$(4.6) \quad ACE_{0\ to\ 1} = \frac{E[S_i(1)]}{E[S_i(0)]}$$

where $E[S_i(\cdot)] = E[E[S_i(\cdot)|P(S_i(\cdot) = 1)]]$. According to Cochran (1968), Little and An (2004), and Schafer and Kang (2008), the inclusion of these categorical variables should remove about 90 percent bias in the regression estimate. Kang and Schafer (2007) and Schafer and Kang (2008) use a simulated dataset to show that this method produces the most unbiased estimate of ACE, but there has been much discussion on the simulated dataset used (Ridgeway and McCaffrey, 2007; Robins *et al.*, 2007).

TABLE 5
Significant Covariates for Safety Models

Variables	Med to Low		High to Med		High to Low	
	PMR		PMR		PMR	
Safety	Med	Low	High	Med	High	Low
Urban		✓	✓		✓	✓
Median	✓					
Multilane		✓				
Propensity1				✓		
Propensity2		✓				
Propensity3			✓			
Propensity4				✓		
Rt. Shldw		✓		✓	✓	✓
Pct. Truck			✓	✓	✓	✓
ADT	✓	✓	✓		✓	✓
Obs	629	693	314	317	119	224

The ACE of equation (4.6) was estimated in two ways, and the standard errors for these estimates were obtained by bootstrapping. The predictions of probability distribution of Safety (S) from the estimated regression models (Table 5) were used in both methods. In the first method (referred to

as *individual prediction* from here on), the prediction was restricted only to the respective groups that were used to estimate the logistic regression model. In the second method (referred to as *combined prediction*), each of the two estimated models was used to predict the safety distribution for the combined matched sample of treatments and controls. For instance, in the comparison of Low *PMR* and Med *PMR*, the regression equation for the safety of Low *PMR* segments was used to predict the safety distributions for both the Low *PMR* segments and well as the Med *PMR* segments. Similarly, the regression equation for the safety of Med *PMR* segments was used to predict the safety of both the Low *PMR* and the Med *PMR* segments. The ACE estimates from the combined prediction method are generally unstable due to the extrapolation of the potential outcomes. However, when the regression models of the potential outcomes are very close to the true model, the combined prediction method is reliable and offers more statistical power, since all of the matched data are used. The results of the two methods and their comparisons are presented in the next section.

4.4. *Results.* As noted previously, three separate propensity score models were estimated and matching was performed using each model. Table 4 shows the covariates used in the final propensity score models. All variables with a p -value < 0.5 were retained in the model.

TABLE 6
Matches based on Propensity Score for Treatment = Medium and Control = Low

Treatment Assignment	Support		
	Off Support	On support	Total
Untreated	1,441	693	2,134
Treated	2,119	629	2,748
Total	3,560	1,322	4,882

4.4.1. *Comparison of Safety between Low and Medium *PMR* levels.* The subsample for the Low and Med *PMR* levels consists of 4882 observations, of which about 56 percent have Med *PMR* level ($T = 1$) and the remaining 44 percent have the Low *PMR* level ($T = 0$). Table 6 illustrates the results of matching and distributional support. A total of 693 controls were matched to 629 treated segments, which had a common support. Table 7 illustrates the balance before and after the matching. It includes the standardized bias (Rosenbaum and Rubin, 1985) before and after matching and the percentage decline in the bias. As per Table 7, there was a considerable improvement in the balance of key covariates in the treatment and the control groups. It also shows that the sample contains no control segments with Low speed =

TABLE 7
t-test before and after matching for Treatment = Medium and Control = Low

Variable	Sample	Mean			% decline bias	t-test	
		Treated	Control	%bias		t	p>t
Multilane	Unmatched	.525	.899	-91.0		-30.64	<0.001
	Matched	0.731	0.715	4.1	95.5	3.38	0.005
Urban	Unmatched	.425	.534	-21.8		-7.56	<0.001
	Matched	.4213	.445	-4.9	77.7	-1.08	0.279
Terrain	Unmatched	.266	.112	40.3		13.70	<0.001
	Matched	.273	.327	-14.0	65.2	-0.73	0.466
Median	Unmatched	.561	.890	-82.5		-27.80	<0.001
	Matched	.731	.714	4.1	95.0	2.47	0.01
Low Speed	Unmatched	.225	0	74.6		24.96	<0.001
	Matched	0.11	0	37.6	50.8	6.07	0.00
Rt. Shoulder	Unmatched	8.699	10.774	-56.1		-18.85	<0.001
	Matched	10.245	10.282	-1.0	98.2	1.9	0.058
Pct. Trucks	Unmatched	12.326	19.469	-85.5		-30.56	<0.001
	Matched	15.124	14.449	8.1	90.6	3.18	0.002
ADT	Unmatched	25694	44222	-68.9		-23.65	<0.001
	Matched	35663	34147	5.6	91.8	2.48	0.013
AGE	Unmatched	16.197	19.751	-47.5		-16.33	<0.001
	Matched	18.383	18.605	-3.0	93.7	0.81	0.416

1. Since for such segments, the counterfactual outcome cannot be estimated, the ACE estimates are only for a sub-population of segments where posted speed limit is greater than 35 mph (Low speed = 0).

To estimate the causal effect, separate binary logistic regression models of safety were estimated for the treatment and the control groups. The variables used in the safety models are shown in Table 5. The estimates of ACE of changing *PMR* from Med to Low on Safety are shown in Table 8. Based on the results of prediction from both groups, the risk of a target crash on segments with Low *PMR* is 1.54 times that of Med *PMR*. Similarly, based on the results from the prediction from individual groups, the risk of a target crash on segments with Low *PMR* is 1.46 times that of Med *PMR*. The point estimates of the ACE are close to each other. Also, there is considerable overlap in the confidence intervals from both methods.

4.4.2. *Comparison of Safety between Medium and High PMR levels.* The subsample for the Med and High *PMR* levels consists of 3782 observations, of which about 27 percent of the segments have the High *PMR* ($T = 1$) and the remaining 73 percent have Med *PMR* ($T = 0$). The summary of matched samples and the balance between the covariates before and after matching are shown in Tables 9 and 10, respectively. A total of 317 controls

TABLE 8
Average Causal Effect for change in *PMR* from Med to Low

Prediction using both Groups					
Variable	Obs	Mean	Std. Dev.	Min	Max
Safety for Control(Low)	1322	0.1037	0.0849	0.0472	0.5494
Safety for Treatment(Med)	1322	0.0673	0.0407	0.0162	0.1799

$$ACE_{MedtoLow} = \frac{E[S(T = Low)]}{E[S(T = Med)]} = \frac{0.1037}{0.0673} = 1.54$$

Confidence Interval = [1.43,1.65]

Prediction using individual Groups					
Variable	Obs	Mean	Std. Dev.	Min	Max
Safety for Control(Low)	693	.0995	.00831	.01499	.517
Safety for Treatment(Med)	629	.0683	.0454	.0162	.1799

$$ACE_{medtoLow} = \frac{E[S(T = Low)]}{E[S(T = Med)]} = \frac{0.0995}{0.0683} = 1.46$$

Confidence Interval = [1.11, 1.99]

were matched to 314 treatment segments. The variables used in the safety model are shown in Table 5 and the ACE of changing *PMR* level from High to Med is shown in Table 11.

TABLE 9
Matches based on Propensity Score for Treatment = High and Control = Medium

Treatment Assignment	Support		
	Off Support	On support	Total
Untreated	2431	317	2748
Treated	720	314	1043
Total	3151	631	3782

Based on the results of prediction from both groups, the risk of a target crash on segments with Med *PMR* is 1.36 times that of High *PMR*. Based on the results from the prediction from individual groups, the risk of a target crash on segments with Med *PMR* is 1.32 times that of High *PMR*. There is less overlap in confidence intervals based on the two methods in comparison to the Med to Low case. Furthermore, the confidence interval based on individual predictions indicates that the effect of change in *PMR* from High to Med is not significantly different from 1 which is consistent with hypotheses found in safety literature.

4.4.3. *Comparison of Safety between Low and High PMR levels.* The subsample for the Low and High *PMR* levels consists of 3168 observations, of

TABLE 10
t-test before and after matching for Treatment = High and Control = Medium

Variable	Sample	Mean			% decline bias	t-test	
		Treated	Control	%bias		t	p>t
Multilane	Unmatched	0.489	0.525	-7.1		-1.94	0.052
	Matched	0.513	0.559	-9.20	-30.20	-0.480	0.629
Urban	Unmatched	0.360	0.425	-13.5		-3.67	<0.001
	Matched	0.366	0.429	-12.800	5.000	-1.710	0.087
Terrain	Unmatched	0.338	0.266	15.7		4.38	<0.001
	Matched	0.169	0.244	-16.400	-3.90	-2.430	0.016
Median	Unmatched	0.538	0.561	-4.7		-1.29	0.197
	Matched	0.532	0.571	-8.0	-69.0	-0.22	0.826
Low Speed	Unmatched	0.249	0.226	5.3		1.47	0.143
	Matched	0.169	0.181	-2.8	47.3	-0.980	0.327
Rt. Shoulder	Unmatched	8.360	8.699	-7.7		-2.08	0.037
	Matched	8.796	8.887	-2.1	73.40	0.140	0.887
Pct. Truck	Unmatched	11.580	12.326	-12.4		-3.4	0.001
	Matched	13.392	14.000	-10.1	18.5	-0.380	0.705
ADT	Unmatched	14284	25694	-51.2		-12.27	<0.001
	Matched	20002	23503	-15.7	69.3	-1.620	0.106
AGE)	Unmatched	4.997	16.197	-187.5		-44.380	<0.001
	Matched	6.541	4.987	26.0	86.10	6.720	<0.001

which about 33 percent of the segments have High *PMR* level ($T = 1$), and the remaining 67 percent have the Med *PMR* level ($T = 0$). A total of 224 control segments were matched to 119 treated segments. Several segments were discarded due to lack of overlap. The summary of matched samples and the balance between the covariates before and after matching are shown in Tables 12 and 13, respectively. Table 12 shows zero control segments with Low speed = 1. Since for such segments the counterfactual outcome cannot be estimated, the ACE estimates are only for a sub-population of segments where the posted speed limit is greater than 35 mph (Low speed = 0) The ACE of changing *PMR* level from High to Low is shown in Table 14.

Based on the results of prediction from both groups, the risk of a target crash on segments with Low *PMR* is about 1.83 times that of High *PMR*. Similarly, based on the results from the prediction from individual groups, the risk of a target crash on segments with Low *PMR* is 2.13 times that of High *PMR*. The results show the same trend, and have considerable overlap in the confidence intervals.

5. Causal Diagrams. In this section we present the Causal Diagrams framework to estimate ACEs. Section 5.1 provides a brief introduction to Causal Bayesian Networks (CBN), different parts of the causal model, and

TABLE 11
Average Causal Effect for change in PMR from High to Medium.

Prediction using both Groups					
Variable	Obs	Mean	Std. Dev.	Min	Max
Safety for Treatment(High)	631	0.0365	0.0568	0.0007	0.3494
Safety for Control(Med)	631	0.0495	0.0640	0.0018	0.3558

$$ACE_{Hightomed} = \frac{E[S(T = Med)]}{E[S(T = High)]} = \frac{0.0495}{0.0365} = 1.36$$

Confidence Interval = [1.04, 1.67]

Prediction using individual Groups					
Variable	Obs	Mean	Std. Dev.	Min	Max
Safety for Treatment(High)	314	0.0382	0.0639	0.0007	0.3494
Safety for Control(Med)	317	0.0505	0.0637	0.0020	0.3312

$$ACE_{Hightomed} = \frac{E[S(T = Med)]}{E[S(T = High)]} = \frac{0.0505}{0.0382} = 1.32$$

Confidence Interval = [0.55, 2.27]

TABLE 12
Matches based on Propensity Score for Treatment = High and Control = Low

Treatment Assignment	Support		
	Off Support	On support	Total
Untreated	1963	171	2134
Treated	763	271	1043
Total	2726	442	3168

the required assumptions. Section 5.2 briefly illustrates the algorithms used to recover a causal model from observational data in light of the assumptions made in Section 5.1. Section 5.3 provides estimation procedures of the ACE and Section 5.4 presents the results of the analysis.

5.1. *Causal Diagrams and components of a Causal Model.* In the Causal Diagrams setting, conditional independence relations between variables are used as fundamental elements to estimate casual effects, in contrast with the PO model, where potential outcomes are the fundamental quantities. Let V denote the set of variables representing the attributes of a road segment which includes both the treatment assigned to a segment and its safety outcome. A Bayesian Network describes the conditional independence relations between the variables in V . The qualitative part of the conditional independence relations are represented by a graph using a set of nodes and edges, and the quantitative part by a set of conditional probability distributions associated with each node in the graph. For more details on graphs and

TABLE 13
t-test before and after matching for Treatment = High and Control = Low

Variable	Sample	Mean			% decline bias	t-test	
		Treated	Control	%bias		t	p>t
Multilane	Unmatched	0.489	0.899	189.1		-28.7	<0.001
	Matched	0.782	0.812	-7.400	93	-0.580	0.563
Urban	Unmatched	0.359	0.534	20		-9.31	<0.001
	Matched	0.529	0.462	13.900	61	1.040	0.299
Terrain	Unmatched	0.338	0.112	-47.2		16.01	<0.001
	Matched	0.227	0.188	9.700	83	0.730	0.464
Median	Unmatched	0.538	0.899	185		-25.35	<0.001
	Matched	0.723	0.812	-21.700	75	-1.62	0.106
Low Speed	Unmatched	0.248	0	4.8		26.56	<0.001
	Matched	0.109	0.00	35.7	56	3.77	<0.001
Rt. Shoulder	Unmatched	8.359	10.774	42.9		-19.87	<0.001
	Matched	9.740	10.342	-17.2	75	-1.34	0.181
Pct. Truck	Unmatched	11.58	19.469	5		-23.08	<0.001
	Matched	15.084	16.983	-22.8	76	-2.04	0.042
log(ADT)	Unmatched	9.225	10.479	126.9		-41.19	<0.001
	Matched	10.086	10.236	-18.4	88	-1.38	0.168
AGE	Unmatched	4.997	19.751	-272.8		-64.6	<0.001
	Matched	6.681	7.086	-7.5	97	-0.81	0.418

graphical models, see [Lauritzen \(1999\)](#).

We restrict our analysis to the class of directed acyclic graphs (DAGs) with discrete variables referred to as discrete Bayesian Networks (BNs). The probability distributions associated with each node is then a conditional probability table (CPT). Figure 1 shows an example of such a graph, where for instance W is a *child* of Y . These restrictions are made to enable the use of efficient algorithms for the generation of and inference in BNs. BNs with continuous variables are possible, but algorithms for handling arbitrary continuous distributions are not well developed. Also, many BN algorithms cannot handle mixed BNs (mixed here refers to the combination of continuous and discrete variables) that have continuous parents of discrete children. To avoid such issues, the data are discretized. The discretization rules for the continuous covariates are given in Table 15.

A BN is an efficient way to represent the joint probability distribution of V and the interpretation of conditional independence in the network does not necessarily imply causation. A Causal Bayesian Network (CBN) is a model that allows such interpretations. It is a DAG in which the predecessors of a node are interpreted as directly causing the variable associated with that node. Formally this the Causal Markov Assumption (CMA) and must be satisfied by all of the variables in the network.

TABLE 14
Average Causal Effect for change in PMR from High to Low.

Prediction using both Groups					
Variable	Obs	Mean	Std. Dev.	Min	Max
Safety for Treatment(High)	343	0.0281	0.0131	0.0072	0.0584
Safety for Control(Low)	343	0.0514	0.0325	0.0043	0.3083

$$ACE_{HightoLow} = \frac{E[S(T = Low)]}{E[S(T = High)]} = \frac{0.0514}{0.028} = 1.83$$

$$\text{Confidence Interval} = [1.59, 2.07]$$

Prediction using individual Groups					
Variable	Obs	Mean	Std. Dev.	Min	Max
Safety for Treatment(High)	119	0.0252	0.012	0.007	0.052
Safety for Control(Low)	224	0.0535	0.028	0.004	0.172

$$ACE_{HightoLow} = \frac{E[S(T = Low)]}{E[S(T = High)]} = \frac{0.0535}{0.0252} = 2.13$$

$$\text{Confidence Interval} = [1.88, 2.37]$$

TABLE 15
Discretization rules for continuous covariates for Bayesian Networks

Variable	Definition
Rt. Shoulder	0 if shoulder width < 7 feet 1 otherwise
ADT	0 if ADT < 30,000 vehicles per day, 1 otherwise
Pct. Trucks	0 if Pct. Trucks < 18, 1 otherwise
AGE	0 if AGE < 10 months, 1 if 10 < AGE < 20 months, 2 otherwise

The CMA states that given the values of a variable's immediate causes (i.e., its parents), the variable is independent of its non-descendants (Pearl, 2000, chap. 1). This assumption implies that we must include in the model every variable that is a cause of two or more other variables. It also implies Reichenbach's (Reichenbach, 1956) common cause assumption, which states that, if any two variables are dependent, then one is a cause of the other or there is a third variable causing both. These implications are important both when using algorithms to recover causal networks from observational data and when estimating the ACE.

The problem of causal inference involves learning the causal structure, represented by a DAG and a CPT, from data. The ACE of a treatment

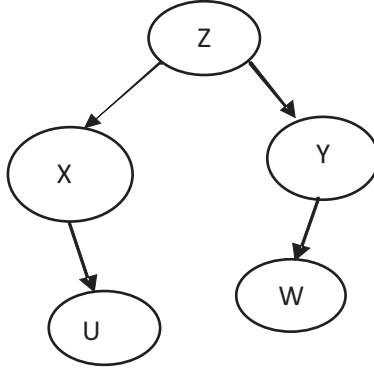


FIG 1. An Example Directed Acyclic Graph

under intervention is estimated using *intervention theory*, as explained in the next section.

5.1.1. *Causal Diagrams as models of intervention.* According to Pearl (2000), CBNs can be regarded as models of interventions if it is assumed that a DAG models the causal mechanism which generated the data. The edges in a directed graph represent a deterministic functional relation coupled with mutually independent error terms. These edges can be used to specify the changes in the joint distribution of variables V due to external intervention. For instance, in Figure 2, forcing the node PMR to take a particular value, say Low , amounts to lifting the existing mechanism on PMR and putting it under the influence of a new mechanism whose action is to force PMR to the value Low , keeping everything else constant. This action is mathematically represented by $do(PMR = Low)$. The effect of ‘setting’ a node to a fixed value corresponds to applying the low PMR treatment to all the segments in the sample.

As explained in Pearl (2000, chap. 3), such interventions, called atomic interventions are modeled in a DAG G by creating a new mutilated DAG G_{PMR} from G . In G_{PMR} , the links between Pa and PMR are removed (where Pa represent the set of parents of PMR), keeping the rest of the graph the same, simulating the effect of an external intervention. The distribution imposed by the new graph G_{PMR} under the condition $PMR = Low$ represents the effect of intervention and is called the post-intervention distribution. An example is shown in Figure 2.

5.1.2. *Causal Effect.* ACE is defined using the interventional notation. Given the safety outcome S and the treatment variable PMR , the ca-

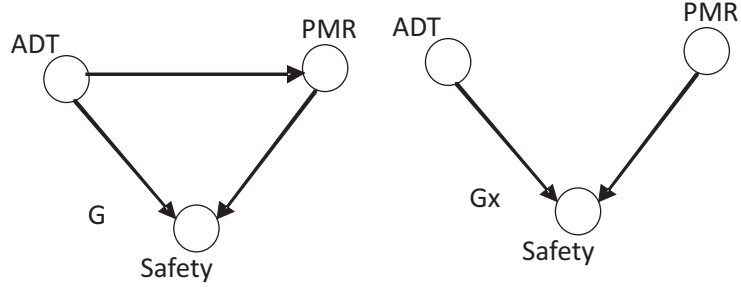


FIG 2. An example of the interventional distribution. The graph G represents the original DAG. The mutilated graph G_x under the intervention of forcing PMR to take a particular value is obtained by deleting the arcs between PMR and its parents (e.g., ADT)

sual effect of PMR on S , denoted by $P[S|do(PMR = i)]$, where $i \in \{Low, Med, High\}$, is a function from PMR to the space of probability distribution on S . For each realization of PMR , $P[S|do(PMR = i)]$ gives the probability of $S = s$ induced from the mutilated graph G_{PMR} and substituting the value of PMR as i in this graph. Mathematically, in the case of atomic intervention $do(PMR = i)$, the joint probability distribution of variables in V can be obtained by the following equation:

$$(5.1) \quad P(x_1, \dots, x_n | do(PMR = i)) = \begin{cases} \frac{P(x_1, \dots, pmr, x_n)}{P(pmr|pa)} & \text{if } pmr = i \\ 0 & \text{if } pmr \neq i \end{cases}$$

Based on equation 5.1, given a causal diagram in which all the parents of manipulated variables are observed, the casual effect can be estimated from passive or non-interventional data. However, when some parents of a child node ch are not observed, $P(ch|pa_i)$ may not be estimable in all cases. A graphical test has been provided by Pearl (2000, chap. 3) to find out when $P[S|do(PMR)]$ is estimable from the observed data. In the present case, we make the assumption that all potential confounders are included in the analysis. This is a strong assumption and must come from subject matter experts. These assumptions are reviewed in Section 5.2.

The ACE of PMR on S is estimated by marginalizing the joint probability distribution in equation 5.1. When PMR has two possible states (Low and Med), the ACE is given by the following equation

$$(5.2) \quad ACE = \frac{P(S = 1 | do(PMR = Low))}{P(S = 1 | do(PMR = Med))},$$

where $P[S = 1 | do(PMR = Low)]$ is the marginal probability of $S = 1$ under the atomic intervention $PMR = Low$; notice the similarity to equa-

tion (4.6) from the PO framework. The marginal probabilities are obtained from the mutilated graph G_{PMR} by intervening on PMR and stratifying on a set of covariates and marginalizing the probability distribution of S over each stratum². In the next section, we describe the algorithms that were used to learn the components of the causal model.

5.2. *Learning Causal Bayesian Networks from data.* Structure learning algorithms are used to recover a DAG G and parameter learning algorithms are used to estimate the CPTs.

5.2.1. *Structure Learning.* Learning the structure of causal networks from observational data has received a thorough treatment in the literature, see Pearl and Verma (1991); Heckerman (1996); Spirtes and Glymour (1991); Tsamardinos, Brown and Aliferis (2006). The most common strategies fall into two different classes called *constraint based learning* and *score based learning*. We adopt a simple combination of both approaches to learn the structure of the BN. Our approach is similar in principle to Tsamardinos, Brown and Aliferis (2006). The PC algorithm (Spirtes, Glymour and Scheines, 2001) is used as a constraint based strategy to recover a DAG structure from the data. This DAG is supplied as an initial structure to the score based learning strategy, which then attempts to find an optimum DAG structure. The simulated anneal strategy of Hartemink (2005) is used to search for optimum scored BN. The scoring function proposed by Heckerman, Geiger and Chickering (1994) is used. The scoring search is implemented in Java using the BANJO library (Hartemink, 2005) and the constraint search is performed using the BNT toolbox in Matlab (Murphy, 2001). Since the scoring method need not produce the globally optimum structure of the BN, we use the 10 best networks recovered by the algorithm and perform Bayesian model averaging to estimate the average causal effect. For details on the averaging, refer to Madigan and Raftery (1991), Hoeting, Adrian and Volinsky (1998) and Heckerman, Geiger and Chickering (1994).

Irrespective of the strategy used, the structure of a DAG can be recovered from observational data (up to d-separation equivalence, (see Pearl, 2000, chap. 1)) only if the three assumptions briefly outlined below are satisfied. Causal interpretation of BNs is possible only because of these assumptions, which are in general untestable from observational data and must come from subject matter experts.

²Conditioning on so called colliders in a DAG can actually introduce bias in the ACE, see Pearl (2003). In simple terms, a variable is a collider if it has two arrows into it. In the present case, there are no such colliders on the path between Safety and PMR

Causal Markov Assumption: The Causal Markov assumption (CMA) is an assertion that each variable in the model is independent of its non-descendants, given its parents or immediate causes. The natural question that arises is what are the immediate factors that affect the safety of a segment? For instance, is driving at a high speed considered an immediate cause of reduced safety? To understand the CMA in light of safety, it is important to first consider the factors that cause a crash. Factors causing crashes can be divided into the following three broad categories: road user (driver), roadway characteristics (geometric elements, environment conditions, roadway volume, etc.), and the vehicle. Analysis using drivers as fundamental units create several problems. The immediate causes of crashes (75 percent of which are due to human error (Stanton and Salmon, 2009)) become very specific to a particular crash and are governed by complex human behavior which is difficult to model and predict. Generally, information on the factors related to drivers and the vehicle is available only for vehicles involved in a crash, and not for non-crash vehicles. Thus, in a driver level analysis, most of the data would be missing. To avoid these issues, analysis is done at the segment level. Only stable attributes of a roadway segment are included in the analysis; specific human factors are included in the disturbance or error terms considered to be stochastic in nature. Thus, the Markov assumption is treated as a guiding principle rather than an assumption, where it defines the granularity of the model being considered ensuring that all relevant causes, as defined by subject matter experts and previous knowledge, are included in the analysis.

Faithfulness: The faithfulness assumption ensures that the population that generated the DAG model has exactly those independence relations specified by the DAG structure and no additional independencies. If there are any independence relations in the population that are not a consequence of the Causal Markov condition, then the population is unfaithful. By assuming Faithfulness we eliminate all such cases from consideration.

Latent Variables: This assumption states that there are no hidden variables in the model that violate the causal Markov condition. That is, all of the variables that effect more than two variables in the model are observed and included in the database. Again, this is a strong assumption, whose validity could be ensured by verification from subject matter experts. For instance, the definition of safety ensures that the causes due to driver and weather factors do not influence the outcome, else these would have to be entered into the model as latent variables.

5.2.2. *Parameter Learning.* The parameters of the CPT are modeled using *Dirichlet distributions* and the usual assumptions of parameter independence are made. For details on parameter learning, see [Heckerman, Geiger and Chickering \(1994\)](#). The Bayesian Dirichlet Equivalent Uniform Priors (BDEU) were used to compute the parameters of the CBN.

The Dirichlet hyper parameters α_{x_i, π_i} are specified by following equation:

$$(5.3) \quad \alpha_{x_i, \pi_i} = \alpha \times p(x_i, \pi_i),$$

where $\alpha_{x_i | \pi_i}$ pertains to variable X_i in a state x_i given that its parents are in joint state π_i , for $i = 1, \dots, n$ where α is the number of pseudo-counts, and p is a (marginal) prior distribution of pseudo-counts; this ensures the likelihood-equivalence of Markov equivalent structures ([Heckerman, Geiger and Chickering, 1994](#)). The value of α is taken to be 1. The distribution p is chosen to be uniform between 0 and 1 for all variables (representing non-informative prior), i.e. for any CPT, each parent-child combination is given an equal probability.

The expected value of the parameters ($\theta_{x_i | \pi_i}$) has the simple expression:

$$(5.4) \quad E_{p(\theta_{x_i | \pi_i} | D, G)}[\theta_{x_i | \pi_i}] = \frac{k_{x_i, \pi_i} + \alpha_{x_i, \pi_i}}{k_{\pi_i} + \alpha_{\pi_i}}$$

where k_{x_i, π_i} is the number of segments in the database with attributes x_i and parents are in joint state π_i .

5.3. *Estimation of ACE.* The ACE of *PMR* on safety is computed using equation 5.5. The intervention probability distributions of the safety variable, under each of the interventions of $do(\text{PMR} = i), i \in \text{Low}, \text{Med}, \text{High}$ are obtained by marginalization in the mutilated DAG.

$$(5.5) \quad \begin{aligned} ACE_{\text{High to Low}} &= \frac{P(\text{Safety} = 1 | do(\text{PMR} = \text{Low}))}{P(\text{Safety} = 1 | do(\text{PMR} = \text{High}))} \\ ACE_{\text{High to Med}} &= \frac{P(\text{Safety} = 1 | do(\text{PMR} = \text{Med}))}{P(\text{Safety} = 1 | do(\text{PMR} = \text{High}))} \\ ACE_{\text{Med to Low}} &= \frac{P(\text{Safety} = 1 | do(\text{PMR} = \text{Low}))}{P(\text{Safety} = 1 | do(\text{PMR} = \text{Med}))} \end{aligned}$$

Computing the marginal probability of a variable from the conditional probability distributions specified by a DAG under an evidence or intervention is not a trivial task when the number of variables is large. This is because the joint probability table increases exponentially with an increase

in the number of variables (Jensen, 2001). There are several methods in the literature (Cowell, 1998) to efficiently perform inference in a BN. We compute the marginal probability of $Safety(S)$ by using the *junction tree algorithm* that performs exact inference. The ACE of PMR on S is estimated as the ratio of the expected value of safety under the intervention level corresponding to the treatment and the expected value of safety under the intervention level corresponding to control; compare this to equation (4.6). A full bayes model of the Dirichlet parameters is specified and the confidence in the value of the ACE is estimated by computing a 95 percent Bayesian credible interval.

5.4. *Results.* Table 16 summarizes the 10 final models selected for the estimation of ACE of PMR on safety. In Table 16, each cell shows the parents of the corresponding variable. The first row represents the parents common to all DAGs, and the subsequent rows are numbered in decreasing order of the score. As per these DAGs, PMR levels in a segment are affected by the AGE of the marking, the percentage of trucks and the ADT in the segment. The safety of a segment is affected by the right shoulder width, PMR level, and the number of lanes in the segment. Most of the relations in the DAGs are in line with the common understanding of traffic safety research.

The DAG with the highest score is shown in Figure 3. In this DAG, the *Low speed* variable does not appear. This could be because given the combination of variables like ADT , *Median* and *Multilane*, the value of *Low speed* is completely determined, and the discretization of ADT into two levels makes it highly collinear with *Low speed*. It was surprising to see the safety of a segment unaffected by ADT in this particular DAG, since it is commonly observed that the higher the ADT , the higher the probability of a target crash on a segment. However, two of the top 10 graphs show that ADT does indeed affect safety. Other possible reasons could be the fact that ADT has been discretized into two levels, and hence possess high correlation with the *Multilane* variable. Similar problems were encountered in the PO framework. Segments with more than two lanes generally have high ADT . This could be the reason why the *Multilane* indicator affects safety in 8 out of the 10 highest scoring models.

Figure 4 shows the mutilated DAG used to model the effect of intervention on the PMR levels. As noted earlier, the mutilated DAG is formed by deleting all the edges from the original DAG that direct into the PMR variable, and fixing the value of PMR at a particular level. The marginal probability distribution of safety in such a DAG represents the effect of

TABLE 16
 10 Best DAGs Used for Bayesian Model Averaging

Variables								
DAG	Mlane	Urb	RShld	Med	ADT	Ptrk	<i>PMR</i>	Safety
Con-sensus	Dist Urb	Dist	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i>
1	Dist Urb	Dist	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i> RShld
2	Dist	Dist Mlane	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i>
3	Dist Urb	Dist	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane Med <i>PMR</i>
4	Dist	Dist Mlane	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane Med <i>PMR</i>
5	Dist Urb	Dist	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Med ADT Ptrk <i>PMR</i>
6	Dist	Dist Mlane	Mlane Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Med ADT Ptrk <i>PMR</i>
7	Dist	Dist Mlane	Mlane Trn	Mlane RShld Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i>
8	Dist Urb	Dist	Mlane Trn	Mlane RShld Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i>
9	Dist	Dist Mlane	Mlane Med Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i>
10	Dist Urb	Dist	Mlane Med Trn	Mlane Trn	Mlane Urb	Mlane Urb Trn	ADT Ptrk Dist	Mlane <i>PMR</i>

Dist = District

Urb = Urban

Med = Median Presence

Mlane = Multilane

Trn = Terrain

Ptrk = Percentage Truck

RShld = Rt. Shoulder Width

Terrain and District variables do not have any parents

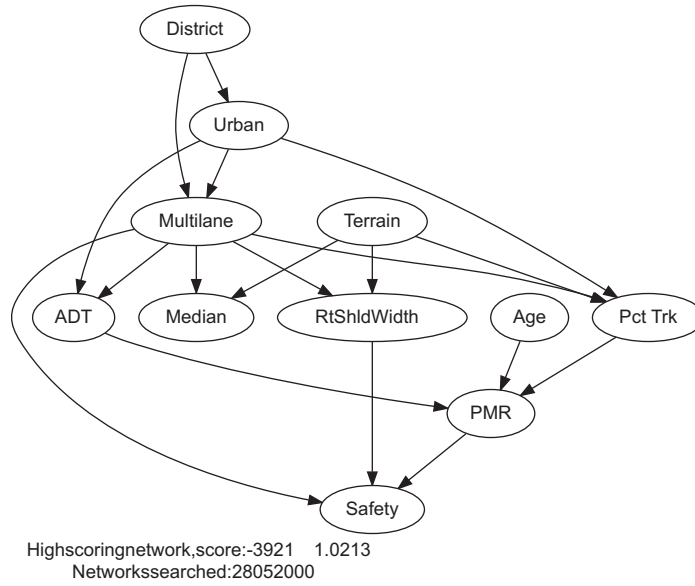


FIG 3. The best scoring DAG recovered by the search algorithm

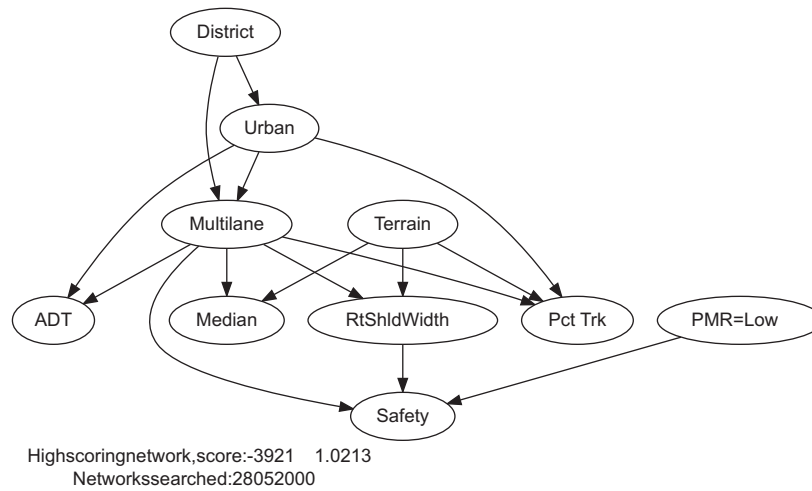


FIG 4. The Mutilated Graph under the Manipulation of PMR = Low

TABLE 17
Estimate of ACE from Causal Model

Change in Visibility Level	Point Estimate	95 % limits
High to Low	3.12	[2.32, 4.11]
Medium to Low	1.79	[1.31, 2.28]
High to Medium	1.86	[1.60, 2.17]

manipulating the PMR variable on safety. We compute the full Bayesian posterior of ACE using Monte Carlo simulation, averaging over the 10 best selected networks. Table 17 shows the final results of effect of PMR on safety, computed via the Causal Diagrams approach. Based on the results, higher PMR levels correspond to lower probability of a target crash. In particular, the expected safety of a segment decreases 3.12 times under the Low PMR conditions when compared to the High PMR conditions; decreases 1.79 times under the Low PMR conditions when compared to Med PMR conditions, and decreases by 1.86 times under Med PMR conditions when compared to High PMR conditions.

6. Comparison of Results. Figure 5 shows the estimates of means and confidence intervals of ACEs. It compares the results from both modeling methods and shows the overlap amongst the estimates of ACEs. It can be seen that the casual effect of increased PMR levels is to generally reduce the risk of a target crash. The results of the analysis follow transportation engineering intuition and the expectation of nighttime driver experts in terms of trend. The trend is shown in the figure as dotted lines. In general, the results show increasing magnitudes of risk of a target crash with an increase in the PMR level variation, with estimates from the PO framework typically smaller in magnitude when compared to the CBN estimates.

In the PO results, the ACE of PMR change from High to Low is the largest, followed by the ACE of PMR change from Med to Low and Med to High. The results from the different PO models are generally consistent with each other except in the case of High to Med comparison. In this case, the individual estimates of ACE assert that the risk of a target crash does not change due to changes in PMR from High to Med. This result is in line with the hypothesis of safety experts, which states that drivers may be incapable of identifying differences in changes in PMR levels defined by Med and High in the present study. On the other hand, the ACE estimate from the combined predictions asserts that there can be an increased risk of target crashes in Med PMR , when compared to High PMR , the magnitude of which could be between 4 and 67 percent. However, as pointed out earlier such estimates maybe less accurate due to possible extrapolation. In the

CBN setting, the ACE of *PMR* change from High to Low is also the largest, about 3.12, and then decreases to 1.79 for Med to Low, and to 1.85 for the High to Med comparison; the last two are not significantly different. But, the CBN results imply that there is a significant difference in the effect of *PMR* between High and Med levels which is in contradiction to the PO results and engineering intuition.

There could be several reasons for the small differences in the results from the PO model and the CBN. One is distributional support. The samples used in the estimation of Causal effect from the Bayesian Network were discretized to ensure the use of efficient algorithms in relation to BNs. On the other hand, there was no such discretization performed in the PO model. Secondly, the PO model discards samples for which no matches were obtained, thus reducing the sample size, although aiming to establish necessary balance. Furthermore, crashes are rare events. In modeling such a response with binary logistic regression, it is likely that the $P(S = 1)$ is underestimated, thus leading to risk estimates in the PO framework which are overall smaller in magnitude.

Thirdly, the differences could also be due to the Bayesian estimation procedure. The Bayesian Dirichlet Equivalent Uniform Priors (BDEU) were used to compute the parameters of the CBN. Although the results of the ACE were not sensitive to the strength (pseudo sample size) of the priors, the priors assign non-zero values to those combination of cells in the CPT with sampling zeros to avoid degenerate probabilities. Zero counts happen due to the exponential increase in the number of cells of a CPT with the increase in the levels of a categorical variable as well as the increase in number of parents. Since all of the variables are discrete, the size of the CPT can be very large. In the present case, the largest CPT had 3×2^4 cells. The estimates from the PO framework are also affected by this problem, although in part mitigated by the presence of continuous variables. However, the problem of degenerate cells was encountered during the matching procedures. We were unable to check the balance between the interaction terms of the categorical covariates of the segment due to sampling zeros. Ensuring balance between the marginal value of the variables itself proved to be very difficult and forced us to discard a significant proportion of the original samples. An attempt to ensure balance between the interactions of the variables was even more difficult. The resulting sample had very few observations left to enable any useful estimation of ACE.

In the present case, since the true ACE is not known, it is difficult to examine which method provides the closest estimate to the true value of ACE. Although, some recent studies indicate that the Bayesian network models

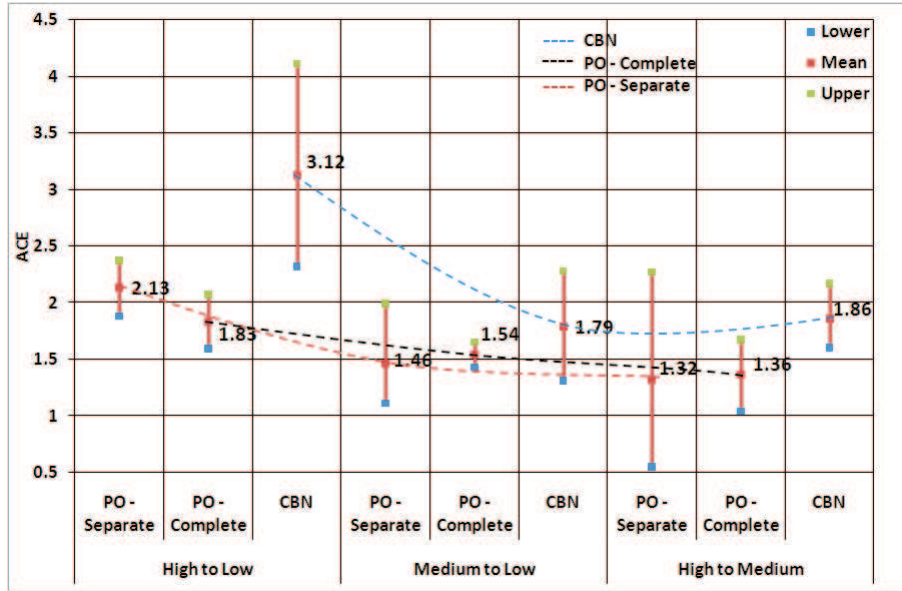


FIG 5. *Overlap Between the Confidence Intervals of ACE from both the methods*

are more sensitive to changes to model parameters than previously thought (see [Druzdzel and Onisko, 2008](#)). Future studies should look at a simulated dataset and compare both the CBN and the PO methods, including full Bayesian PO analysis, and compare the results.

7. Conclusion and Discussion. The examination of causal inference methods to transportation safety reveals that there is considerable scope of their application to estimate safety effects of a countermeasure. The purpose of this paper was to examine the applicability and perform a comparison of two popular causal inference methods in the context of an important transportation safety issue. A comparison was made between the Potential Outcomes framework and the Causal Diagrams framework. In this exploratory analysis, it was found that increased *PMR* levels in general improve the nighttime safety of a segment.

In this analysis we modeled safety as a binary outcome. Alternatively, one could model the number of crashes in a segment with Poisson or Negative Binomial regressions. Sampling zeros were encountered in both the PO model as well as the CBN setting. In the PO framework, such sampling zeros created problems in matching and achieving balance over interactions of covariates. In the CBN setting, the use of Bayesian Inference in part addressed the estimation problem. A similar approach in the PO framework

would be to use a full Bayes model of both the propensity scores as well as safety outcomes (Rubin *et al.*, 2008), which we hope to do as part of further investigations.

Also, to obtain a better comparison of the methods, future studies should aim at using data from simulation. Crash data could be simulated to mimic real life data. The true casual effect of the population would be known a priori, and the quality and size of data can be controlled. Other advances on this exploratory work can be made by using more complex causal modeling methods and improving upon the limitations of this study. For instance, the *PMR* treatment variable was discretized in this study. This can be avoided by estimating ACE using a dose-response model (Hirano and Imbens, 2005). The *PMR* treatment can also be modeled as a time varying treatment (Gill, van der Vaart and Robins, 2004). In this study, the effect of a simple treatment variable on crashes was considered. Specification of the assignment mechanism for the *PMR* treatment variable is convenient when compared to other possible countermeasures, such as those related to geometric properties of a segment. For instance, a railroad crossing installed on a segment as a countermeasure may create additional challenges for estimation of causal effects. The assignment mechanism for a railroad crossing is generally influenced by factors such as local design policies, complaints from residents, or even arbitrary discretion of the traffic engineer. These factors may be difficult to measure and account for. The assignment mechanism may also be related to past crash history. Such assignment mechanisms may prove difficult to model and may require the use of latent variables. These explorations are left to the scope of future work.

A parallel can be drawn between the assignment mechanism introduced in the potential-outcome framework and interventional notation used in the CBN. The potential outcomes framework models the probabilistic mechanism by which the value of a treatment attribute is assigned to each unit. The DAG structure of a CBN models the probabilistic mechanism by which the values of all attributes are assigned to a unit. The CBN approach provides an intuitive and graphical way to represent casual assumptions and make them more explicit. However, inference in CBN is computationally expensive and requires strong distributional assumptions. Learning and inference algorithms for BN structures that can handle combinations of arbitrary probability distributions are limited. The potential-outcome framework is less limited by these problems, and considerable statistical literature has shown the importance of careful study design and balancing. A combination of the best features from both methods can provide a rich framework to estimate the safety effects of countermeasures in transportation studies.

References.

- BAHAR, G., MASLIAH, M., ERWIN, T., TAN, E. and HAUER, E. (2006). Pavement marking Materials and Markers: Real-World Relationship between Retroreflectivity and Safety over Time. *NCHRP Web Only Document 92. Transportation Research Board, National Research Council, Washington D.C.*
- BUREAU OF TRANSPORTATION STATISTICS, U. S. (2007). National transportation statistics Technical Report, US Department of Transportation. Available at <http://www.bts.gov/publications/nationaltransportationstatistics>
- COCHRAN, W. G. (1965). The Planning of Observational Studies of Human Populations. *Journal of the Royal Statistical Society. Series A(General)* **128** 234-266.
- COCHRAN, W. G. (1968). The effectiveness of adjustment by subclassification in removing bias in observational studies. *Biometrics* **24** 295-313.
- COLE, S. R. and FRANGAKIS, C. E. (2009). The consistency statement in causal inference: a definition or an assumption? *Epidemiology (Cambridge, Mass.)* **20** 3-5. Available at <http://dx.doi.org/10.1097/EDE.0b013e31818ef366>
- COWELL, R. (1998). Introduction to inference for Bayesian networks. In *Proceedings of the NATO Advanced Study Institute on Learning in graphical models* 9-26. Kluwer Academic Publishers, Norwell, MA, USA.
- DAVIS, G. A. (2000). Accident reduction factors and causal inference in traffic safety studies: a review. *Accident Analysis & Prevention* **32** 95 - 109. <http://www.sciencedirect.com/science/article/B6V5S-3XTDTJH-E/2/3496a300559086bcd2c24417d4f8e005>. 10.1016/S0001-4575(99)00050-0
- DAVIS, G. A. (2004). Possible aggregation biases in road safety research and a mechanism approach to accident modeling. *Accident Analysis & Prevention* **36** 1119-1127.
- DEHEJIA, R. H. and WAHBA, S. (1998). Causal Effects in Non-Experimental Studies: Re-Evaluating the Evaluation of Training Programs NBER Working Papers report No. 6586, National Bureau of Economic Research, Inc. Available at <http://ideas.repec.org/p/nbr/nberwo/6586.html>
- DEHEJIA, R. H. and WAHBA, S. (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics, 2002 - MIT Press* **84** 151-161.
- DONNELL, E. T., KARWA, V. and SATHYANARAYANAN, S. (2009). A Methodology to Explore the Relationship between Pavement Marking Retroreflectivity and Traffic Crash Frequency on Highways in North Carolina: Application of Artificial Neural Networks and the Generalized Estimating Equations. *Transportation Research Record: Journal of the Transportation Research Board* **Accepted**.
- DRUZDZEL, M. J. and ONISKO, A. (2008). The Impact of Overconfidence Bias on Practical Accuracy of Bayesian Network Models: An Empirical Study. *Proceedings of the 6th Bayesian Modelling Applications Workshop, UIA-2008*.
- EELLS, E. (1991). *Probabilistic Causality. Cambridge Studies in Probability, Induction, and Design Theory Series*. Cambridge University Press.
- FIENBERG, S. E. and SFER, A. M. (2006). Randomization, Models, and the Estimation of Causal Effects. *Statistics in Medicine, submitted*.
- GILL, R., VAN DER VAART, A. and ROBINS, J. (2004). Estimating the causal effect of a time-varying treatment on time-to-event using structural nested failure time models. *Statistica Neerlandica* **58** 271-295.
- HARTEMINK, A. J. (2005). Banjo: Bayesian Network Inference with Java Objects. Available at <http://www.cs.duke.edu/~amink/software/banjo/>
- HECKERMAN, D. (1996). A Tutorial on Learning With Bayesian Networks Technical Re-

- port, Learning in Graphical Models. Available at <ftp://ftp.research.microsoft.com/pub/tr/tr-95-06.pdf>
- HECKERMAN, D., GEIGER, D. and CHICKERING, D. (1994). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. In *Machine Learning* 197–243.
- HERNAN, M. A. and ROBINS, J. M. (2006). Estimating causal effects from epidemiological data. *J Epidemiol Community Health* **60** 578–586. Available at <http://jech.bmj.com/cgi/content/abstract/60/7/578>. 10.1136/jech.2004.029496
- HIRANO, K. and IMBENS, G. W. (2005). The Propensity Score with Continuous Treatments. In *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives* (X.-L. M. Andrew Gelman, ed.) 73–84. Department of Statistics and Department of Political Science, Columbia University, New York, USA; Department of Statistics, Harvard University, USA.
- HOETING, J., ADRIAN, D. M. and VOLINSKY, C. T. (1998). Bayesian Model Averaging. In *In Proceedings of the AAAI Workshop on Integrating Multiple Learned Models* 77–83. AAAI Press.
- HOLLAND, P. W. (1986). Statistics and causal inference. *Journal of the American Statistical Association* **81** 945–960.
- HONG, G. and RAUDENBUSH, S. W. (2005). Effects of Kindergarten Retention Policy on Children’s Cognitive Growth in Reading and Mathematics. *Educational Evaluation and Policy Analysis* **27** 205–224. Available at <http://epa.sagepub.com/cgi/content/abstract/27/3/205>. 10.3102/01623737027003205
- JENSEN, F. V. (2001). *Bayesian Networks and Decision Graphs*. Information Science and Statistics. Springer. Available at <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0387952594>
- KANG, J. D. Y. and SCHAFER, J. L. (2007). Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Statistical Science* **22** 523–539.
- KARWA, V. (2009). Safety Effects of Pavement Marking Retroreflectivity: An Application of Causal Bayesian Networks Master’s thesis, The Pennsylvania State University.
- KARWA, V. and DONNELL, E. T. (2009). Models of Pavement Marking Retroreflectivity: Comparison of Artificial Neural Networks to Linear Statistical Models. *ASCE Journal of Transportation Engineering*.
- LAURITZEN, S. L. (1999). Causal Inference from Graphical Models. In *In Complex Stochastic Systems* 63–107. Chapman and Hall/CRC Press.
- LEUVEN, E. and SIANESI, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. Statistical Software Components, Boston College Department of Economics. Available at <http://ideas.repec.org/c/boc/bocode/s432001.html>
- LITTLE, R. and AN, H. (2004). Robust Likelihood-based Analysis of Multivariate Data with Missing Values. *Statistica Sinica* **14** 949–968.
- MADIGAN, D. and RAFTERY, A. E. (1991). Model Selection and Accounting for Model Uncertainty in Graphical Models using Occam’s Window. *Journal of the American Statistical Association* **89** 1535–1546.
- MALDONADO, G. and GREENLAND, S. (2002). Estimating causal effects. *International Journal of Epidemiology* **31** 422–429. Available at <http://view.ncbi.nlm.nih.gov/pubmed/11980807>
- MURPHY, K. P. (2001). The Bayes Net Toolbox for MATLAB. *Computing Science and Statistics* **33** 2001.
- NATIONAL HIGHWAY TRAFFIC SAFETY ADMINISTRATION, U. S. (2007). Traffic Safety Facts 2007 Technical Report, US Department of Transportation. Available at <http://>

www-nrd.nhtsa.dot.gov/Pubs/811002.PDF

- PEARL, J. (2000). Causality: Models, Reasoning, and Inference.
- PEARL, J. (2003). Statistics and Causal Inference: A Review. *Test* **12**. Available at <http://www.springerlink.com/content/r411t0qktvnm5005>
- PEARL, J. and VERMA, T. (1991). A theory of inferred causation. In *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference* **11** 441-452.
- REICHENBACH, H. (1956). *The Direction of Time*. Univ. of California Press, Berkeley.
- RIDGEWAY, G. and MCCAFFREY, D. F. (2007). Comment: Demystifying Double Robustness: A Comparison of Alternative Strategies for Estimating a Population Mean from Incomplete Data. *Statistical Science* **22** 540-543.
- ROBINS, J., SUED, M., LEI-GOMEZ, Q. and ROTNITZKY, A. (2007). Comment: Performance of Double-Robust Estimators When Inverse Probability Weights Are Highly Variable. *Statistical Science* **22** 544-559.
- ROSENBAUM, P. R. (1995). *Observational studies*. New York: Springer-Verlag.
- ROSENBAUM, P. and RUBIN, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* **70** 41-55.
- ROSENBAUM, P. R. and RUBIN, D. B. (1985). Constructing a Control Group Using Multivariate Matched Sampling Methods That Incorporate the Propensity Score. *The American Statistician* **39** 33-38.
- RUBIN, D. B. (1978). Bayesian Inference for Causal Effects: The Role of Randomization. *The Annals of Statistics* **6** 34-58.
- RUBIN, D. B. (1990). Formal mode of statistical inference for causal effects. *Journal of Statistical Planning and Inference* **25** 279-292. Available at [http://dx.doi.org/10.1016/0378-3758\(90\)90077-8](http://dx.doi.org/10.1016/0378-3758(90)90077-8)
- RUBIN, D. B. (1997). Estimating Causal Effects from Large Data Sets Using Propensity Scores. *Ann Intern Med* **127** 757-763.
- RUBIN, D. B. (1998). Estimation from nonrandomized treatment comparisons using subclassification on propensity scores. *Annals of Internal Medicine* **8** 757 - 763.
- RUBIN, D. B. (2005). Causal Inference Using Potential Outcomes: Design, Modeling, Decisions. *Journal of the American Statistical Association* **100** 322-331. Available at <http://dx.doi.org/10.1198/016214504000001880>
- RUBIN, D. B. (2008). For objective causal inference, design trumps analysis. *Annals of Applied Statistics* **2** 808. Available at [doi:10.1214/08-AOAS187](https://doi.org/10.1214/08-AOAS187)
- RUBIN, D. B., WANG, X., YIN, L. and ZELL, E. R. (2008). Bayesian causal inference: Approaches to estimating the effect of treating hospital type on cancer survival in Sweden using principal stratification. In *Handbook of Applied Bayesian Analysis* (T. OHagan and M. West, eds.) Oxford Univ. Press, Oxford.
- SCHAFFER, J. L. and KANG, J. (2008). Everage Causal Effects from Nonrandomized Studies: A Practical Guide and Simulated Example. *Psychological Methods* **13** 279-313.
- SFER, A. M. (2005). Randomization and Causality Technical Report, PhD Thesis, Universidad Nacional de Tucuman.
- SPIRITES, P. and GLYMOUR, C. (1991). An Algorithm for Fast Recovery of Sparse Causal Graphs. *Social Science Computer Review* **9** 62-72. [10.1177/089443939100900106](https://doi.org/10.1177/089443939100900106)
- SPIRITES, P., GLYMOUR, C. and SCHEINES, R. (2001). *Causation, Prediction, and Search*, Second ed. The MIT Press, Cambridge, MA, USA. Available at <http://www.amazon.com/exec/obidos/redirect?tag=citeulike07-20&path=ASIN/0262194406>
- STANTON, N. A. and SALMON, P. M. (2009). Human error applied to driving: A generic driver error taxonomy and its implications for intelligent transport systems. *Safety Science* **47** 227-237.

TSAMARDINOS, I., BROWN, L. E. and ALIFERIS, C. F. (2006). The max-min hill-climbing Bayesian network structure learning algorithm. *Mach. Learn.* **65** 31–78. <http://dx.doi.org/10.1007/s10994-006-6889-7>

ADDRESS OF THE FIRST AND SECOND AUTHORS
DEPARTMENT OF STATISTICS
PENN STATE UNIVERSITY
UNIVERSITY PARK, PA, 16802
E-MAIL: vk106@psu.edu
sesa@stat.psu.edu

ADDRESS OF THE THIRD AUTHOR
DEPARTMENT OF CIVIL ENGINEERING
PENN STATE UNIVERSITY
UNIVERSITY PARK, PA, 16802
E-MAIL: edonnell@engr.psu.edu