CAUSE-EFFECT RELATIONSHIPS IN ANALYTICAL SURVEYS: AN ILLUSTRATION OF STATISTICAL ISSUES

GARY L. GADBURY[1*] and HANS T. SCHREUDER[2]
[1] *Department of Mathematics and Statistics, University of Missouri, Rolla, Missouri, U.S.A.;*
[2] *Rocky Mountain Research Station, Fort Collins, Colorado, U.S.A.*
([*] author for correspondence, e-mail: gadburyg@umr.edu)

(Received 23 October 2001; accepted 18 May 2002)

Abstract. Establishing cause-effect is critical in the field of natural resources where one may want to know the impact of management practices, wildfires, drought, etc. on water quality and quantity, wildlife, growth and survival of desirable trees for timber production, etc. Yet, key obstacles exist when trying to establish cause-effect in such contexts. Issues involved with identifying a causal hypothesis, and conditions needed to estimate a causal effect or to establish cause-effect are considered. Ideally one conducts an experiment and follows with a survey, or vice versa. In an experiment, the population of inference may be quite limited and in surveys, the probability distribution of treatment assignments is generally unknown and, if not accounted for, can cause serious errors when estimating causal effects. The latter is illustrated in simulation experiments of artificially generated forest populations using annual plot mortality as the response, drought as the cause, and age as a covariate that is correlated with mortality. We also consider the role of a vague unobservable covariate such as 'drought susceptibility'. Recommendations are made designed to maximize the possibility of identifying cause-effect relationships in large-scale natural resources surveys.

Keywords: bias, causality, counterfactual, ignorability, potential response

1. Introduction

Controversy exists concerning the extent to which observational studies can be used to establish cause-effect relationships (Feinstein, 1988; Schreuder and Thomas, 1991), though it has been remarked that the objective of conducting such studies is, in fact, to elucidate cause-effect relationships (Cochran, 1965). There is a fundamental difference between observational studies and experimental studies. In an experimental study, the assignment of experimental units to treatments is controlled by the experimenter who ensures that units receiving different treatments are comparable (Rosenbaum, 1995, page 2). This control is absent in an observational study for various reasons. In studies concerning the effect of environmental factors on natural resources, the control is absent since it is beyond the reach of the investigator, e.g., assignment of rainfall to forest stands is determined by 'nature' and may not be a random assignment.

A considerable literature is available on cause-effect relationships but clear guidelines on how to establish them, in practice, are unavailable. We attempt to

Environmental Monitoring and Assessment **83:** 205–227, 2003.

remedy this situation by using a forestry example to illustrate guidelines for implementing a cause-effect study, the assumptions that are needed, and the pitfalls that may occur when estimating causal effects. Various statistical methods can be used for estimation of which our focus is likelihood estimation.

Two key issues are involved in establishing cause-effect. The first concerns identification of a possible causal hypothesis. Once this is done, the second can be addressed via manipulating the causal variable in some nonrandom pattern and observing the corresponding changes in the response variable.

2. Literature Review

Often in studies involving survey data, some effect is observed and investigators postulate various causes to which the effect might be attributed. An example is a decline in growth rate (basal area increment) observed in the Southeastern United States forests from 1961 to 1982 (see Bechtold *et al*., 1991). Sheffield *et al*. (1985) reported several possible causes for this decline including atmospheric deposition of pollutants, aging of stands, increased stand density, drought, lower water tables, disease, and combined effects. Most of these 'possible causes' were not observed in the Forest Inventory and Analysis (FIA) data (Bechtold *et al*., 1991), so the plausibility of these 'causal hypotheses' rested on other knowledge. Rubin (1990) states that it is generally impossible to observe an outcome and then find 'the' cause. Some argue that establishing causation depends on manipulating a defined causal treatment and then observing the associated effect (Holland, 1986).

Checklists of criteria could be helpful in formulating a causal hypothesis (Hill, 1965; Susser, 1988) and guiding the investigator through a process of identifying plausible causal hypotheses. It is a process of formulating an 'elaborate theory' (Cochran, 1965) where one can consider a number of explanations for some observed effect and select the most plausible for further investigation. Rubin (1991) notes that the key issues with any of the approaches to causal inference are the formulation of a clear conceptual framework in which causal relationships can be defined. Within this framework, necessary and sufficient conditions to establish cause-effect require satisfying two of the three criteria listed in Mosteller and Tukey (1977): consistency, responsiveness and mechanism. Consistency implies that the presence and magnitude of the effect is always associated with a minimal level of the suspected causal agent; responsiveness is established by experimental exposure to the suspected causal agent and reproducing the symptoms; and mechanism is established by demonstrating a biological or ecological process that causes the observed effect.

Once a plausible hypothesis has been defined, the suspected causal variable must be defined and its measurability assessed. Then an analytical survey can be designed with a purpose to collect the necessary data and estimate the causal effect of the variable in question on some response of interest. An inference technique

is used to provide a 'best' estimate of the causal effect. This process is an observational study since assignment of levels of the causal variable to population units was not controlled by the scientist.

The randomized experiment has generally been considered the 'gold standard' for assessing the causal effect of a treatment on a population (Feinstein, 1988). In such a setting, the treatments are assigned to experimental units at random. However, randomization of treatments rarely occurs in surveys, so resulting inferences and conclusions about a particular treatment or exposure variable can be sensitive to hidden effects from an unknown treatment assignment mechanism (Rosenbaum, 1995). Even in a randomized experiment, inferences regarding the causal effect of a particular treatment on a larger population beyond that of the experimental units may be questioned. The units in the experiment are often not representative of a larger population because they are not a random sample from the population. An example is a clinical trial to evaluate the efficacy of a new drug. Even though treatments may be randomly assigned, patients frequently are volunteers and a larger population model can only be invoked as an untestable assumption (Lachin, 1988). In this setting, the mechanism that samples units from the population may be unknown to the investigator and conclusions regarding a treatment's effect on the larger population may be subject to bias. In other words, the sampled population may behave very differently from the target population.

Much has been written regarding the theory of sampling and assignment mechanisms (Rubin, 1976; Rosenbaum, 1984; Smith and Sugden, 1988) and the conditions under which the unknown mechanisms can be ignored in the inference process (referred to as conditions of ignorability). That is, inferences will be accurate even though one or both of the mechanisms that produced the data were unknown to the scientist and, hence, ignored. Smith and Sugden (1988) presented sufficient conditions for ignoring the mechanisms in a likelihood framework. These involved statistical conditional independence between variables and included conditions that pertain to the possible effect of a lurking variable (i.e., a variable that is not observed in the study but that influences the inference). However, the role that these conditions might play in a forestry context might still be vague to a forest researcher. Furthermore, in many cases, the conditions for ignoring the mechanisms cannot be checked from observed data and, so, must be assumed. In practice, these assumptions are often made without a clear understanding of how violations might alter conclusions from a study.

To pinpoint the problem, suppose that there are N units in a population, denoted by $u_1, u_2, \ldots, u_N$, with a response of interest Y. Each of the N units is exposed to an agent that is represented by the variable Z, which is assumed to have, say, k levels. A unit is 'exposed' to the agent if it has been assigned a positive level of Z. Otherwise it is unexposed. If we are only comparing two treatments, for example, then Z may take on the values of 0 or 1, depending on whether the unit was exposed to the agent or not. If Z is continuous, then there are essentially an infinite number of exposure levels. When Z is dichotomous, one of the two bivariate outcomes

can be observed for a unit u_i, $(Y_i, Z_i) = (Y_i^{(1)}, 1)$ or $(Y_i^{(0)}, 0)$, where $Y_i^{(j)}$ is the response the unit would have when assigned $Z = j (j = 0, 1)$. The bivariate outcomes, $(Y_i, Z_i) = (Y_i^{(1)}, 0)$ or $(Y_i^{(0)}, 1)$ have been called counterfactual (Glymore, 1986) since they cannot be observed in nature simultaneously. For simplicity, we will not use the bivariate response notation above and assume that if a unit is exposed to the agent, i.e., $Z = 1$, we observe a response $Y^{(1)}$, and if it is not exposed, i.e., $Z = 0$, we observe, instead, a response $Y^{(0)}$. Ideally both responses, $Y^{(1)}$, $Y^{(0)}$, could be measured at the same time. But it is clear that only one of these responses can be observed for any given unit depending on whether the unit was exposed to the agent. This is the heart of the problem in establishing cause-effect relationships. The $Nx2$ matrix of 'potential responses' (Rubin, 1974) defines a true population $\underline{Y} = \underline{Y}^{(1)}, \underline{Y}^{(0)}$, where $\underline{Y}^{(1)}$ and $\underline{Y}^{(0)}$ are $Nx1$ response vectors if all units in the population have $Z = 1$ or $Z = 0$, respectively. The true causal effect of the variable Z on the ith population unit u_i is the unobservable difference, $D_i = Y_i^{(1)} - Y_i^{(0}$. We assume there is no interference between units (Cox, 1958a, p. 19). A population average causal effect of Z (Smith and Sugden, 1988) is the difference between the average of $\underline{Y}^{(1)}$ and $\underline{Y}^{(0)}$, that is,

$$\bar{D} = \frac{1}{N} \sum_{i=1}^{N} D_i \ . \tag{1}$$

This is the quantity we would like to estimate unbiasedly and efficiently in natural resource cause-effect studies using maximum likelihood inference. For this method of inference, it is useful to extend the potential response framework to one that facilitates model-based inference and at the same time, accommodates covariates. These extensions were also used in Smith and Sugden (1988). Likelihood inference has wide acceptance in science and has the further advantage that it is easy to incorporate prior distributions as done in Bayesian statistics. This would make it especially attractive for forest managers having to make decisions based on their knowledge (to be used in specifying the prior distribution) and the data collected.

Assume the following: Let $\underline{X}$ be an Nxp matrix of p covariates observable on all population units and $\underline{W}$ be an Nxq matrix of covariates 'unobservable' on any units. Covariate values are not influenced by levels of Z. The potential responses $(\underline{W}, \underline{X}, \underline{Y})$ are generated from a superpopulation. The joint distribution of potential responses is:

$$f(\underline{W}|\underline{Y}, \underline{X}, \underline{\psi}) g(\underline{Y}|\underline{X}, \underline{\theta}) h(\underline{X}, \underline{\phi}) \ , \tag{2}$$

with distinct parameters $\underline{\psi}$, $\underline{\theta}$, and $\underline{\phi}$ (Smith and Sugden, 1988). Note that $|X, \theta$ is shorthand notation for: conditional on X, θ. 'Distinct' refers to the idea that the space of one parameter is unrestricted by that of another. In a Bayesian framework, distinctness of parameters means that the joint prior distribution is the product of the marginal prior distributions. Using the notation in Equation (2), the parameter

$\underline{\theta}$ is of interest, or some function of it that we denote as $\tau(\theta)$. As noted earlier, we cannot observe all potential responses, so we must use the joint distribution in Equation (2) along with some mechanisms that produce observable data. The first mechanism is a treatment assignment mechanism denoted by,

$$P(\underline{Z}|\underline{Y}, \underline{W}, \underline{X}) , \tag{3}$$

where $\underline{Z}$, if dichotomous, is an $Nx1$ vector of ones and zeros and $Z = 1$ corresponds to a unit exposed to the agent. Thus, the outcome from Equation (3) will produce observable values in $\underline{Y}$. We assume that treatment assignment of the level of Z to population units occurs before sampling from the population. This is consistent with the idea that 'nature' exposes units in a forest population to levels of Z and the scientist then samples exposed (or unexposed) units via some survey sampling design. We define a sampling mechanism by,

$$P(\underline{S}|\underline{Z}, \underline{Y}, \underline{W}, \underline{X}) , \tag{4}$$

where $\underline{S}$ is an $Nx1$ vector of ones and zeros and $S_i = 1$ corresponds to unit u_i being selected for the sample. For a simple random sample of size n, an outcome of $\underline{S}$ will be a vector containing n ones and $N - n$ zeros for each of M possible outcomes where

$$M = \begin{pmatrix} N \\ n \end{pmatrix} .$$

Ideal studies comprise experiments conducted following surveys or surveys conducted following experiments because the scientist both selects sample units and assigns the treatments. Suppose that sample selection follows assignment of treatments. Examples of a treatment assignment that is controlled by the scientist are a completely randomized design where the observable covariate $\underline{X}$ is not used, that is, $P(\underline{Z}|\underline{Y}, \underline{W}, \underline{X}) = P(Z)$, or a matched-pairs or block design where $\underline{X}$ is used to match or block units, i.e., $P(\underline{Z}|\underline{Y}, \underline{W}, \underline{X}) = P(\underline{Z}|\underline{X})$. Examples of a sampling design that is controlled by the scientist are a simple random sample where neither $\underline{Z}$ nor $\underline{X}$ are used, that is, $P(\underline{S}|\underline{Z}, \underline{Y}, \underline{W}, \underline{X}) = P(\underline{S})$, or a stratified sampling design where $\underline{Z}$ and/or $\underline{X}$ (both are assumed observable for all population units) are used to stratify the population. In either case, the scientist is assumed to know the random mechanisms that produced the observed data.

Using the more general assignment mechanisms in Equations (3) and (4), the full likelihood of observable data can be written as (Smith and Sugden, 1988),

$$h(\underline{X}; \underline{\phi}) \int_{\underline{C}} \int_{\underline{W}} P(\underline{S}|\underline{Z}, \underline{Y}, \underline{W}, \underline{X}) P(\underline{Z}|\underline{Y}, \underline{W}, \underline{X})$$

$$f(\underline{W}|\underline{Y}, \underline{X}; \underline{\psi}) g(\underline{Y}|\underline{X}; \underline{\theta}) dw dy , \tag{5}$$

where the first integration is with respect to the unobservable covariate, $\underline{W}$, and the second is with respect to unobservable potential responses in $\underline{Y}$. That is, the region $\underline{C}$ is the union of the unsampled units and unobservable potential responses of sampled units, i.e. the union of (a) and (b) below where

(a) $Y_{\tilde{s}}^{(j)} j = 0, 1$; all potential responses for unsampled units ($\tilde{s}$);
(b) $Y_{s-\tilde{y}(j)}^{(j)} j = 0, 1$; potential responses for units in the observed sample s not observable since they are counterfactual to the actual treatment assigned.

The purpose of writing Equation (5) in this way is to highlight the role of the treatment assignment and sampling mechanisms in influencing inference results on $\tau(\theta)$ since most population values cannot be observed. As a simple illustration, consider a set of potential responses for $N = 3$ units. The outcomes for $\underline{Z}$ and $\underline{S}$ are also given.

$$(\underline{W}, \underline{Y}, \underline{X}) = \begin{pmatrix} W_1 & Y_1^{(1)} & Y_1^{(0)} & X_1 \\ W_2 & Y_2^{(1)} & Y_2^{(0)} & X_2 \\ W_3 & Y_3^{(1)} & Y_3^{(0)} & X_3 \end{pmatrix} \underline{S} = \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \underline{Z} = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} .$$

The integration in Equation (5) consists of integrating over all three W variables and the region $\underline{C}$ consists of both Y values in the third row, and then $Y_2^{(1)}$ and $Y_1^{(0)}$ in rows 2 and 1, respectively.

Let us assume a simple random sampling mechanism where for every outcome $\underline{s}$ of the random variable $\underline{S}$, $P(\underline{S} = \underline{s}) = 1/M$. When this is the case, the sampling mechanism is not affected by either of the two sets of integrations in Equation (5) and, so, can be taken outside of the integrals. If the sampling mechanism depended on either the unobservable $\underline{W}$ values, the unobservable $\underline{Y}$ values, and/or $\underline{Z}$, then it would play a role in the integrations. When the sampling mechanism can be taken outside of the integrals and when all parameters are distinct as stated earlier, then the sampling mechanism is said to be ignorable for likelihood based inference (Smith and Sugden, 1988). When this is the case, the maximum likelihood estimate (MLE) for $\tau(\theta)$ is not affected by the sampling mechanism though the value of the likelihood at the MLE for $\tau(\theta)$ is scaled by the probability of observing that particular outcome of $\underline{S}$, which is $1/M$ for simple random sampling. We consider this further in discussing the treatment assignment mechanism.

Consider the two conditions for a treatment assignment (Smith and Sugden, 1988),

$$\begin{aligned} &\text{(i)} \quad \underline{Z} \perp \underline{Y} | \underline{X}, \underline{W}; \\ &\text{(ii)} \quad \underline{Y} \perp \underline{W} | \underline{X}; \underline{\psi} \ , \end{aligned} \tag{6}$$

where $\perp$ indicates independence. When conditions (i) and (ii) are met, the treatment assignment mechanism is ignorable for likelihood inference. To see this,

suppose the sampling mechanism is $P(\underline{S})$, and that conditions (i) and (ii) are met. The full likelihood in Equation (5) can then be written as,

$$P(\underline{S})h(\underline{X};\underline{\phi})\int_{\underline{W}} P(\underline{Z}|\underline{W},\underline{X})f(\underline{W}|\underline{X};\underline{\psi})dw$$

$$\int_{\underline{C}} g(\underline{Y}|\underline{X};\underline{\theta})dy = K(\cdot)h(\underline{X};\underline{\phi})\int_{\underline{C}} g(\underline{Y}|\underline{X};\underline{\theta})dy\ , \tag{7}$$

where $K(\cdot)$ is a function of observed values of $\underline{Z}$, $\underline{X}$, and the unknown parameter $\underline{\psi}$, but since $\underline{\psi}$, $\underline{\phi}$, and $\underline{\theta}$ are distinct, $K(\cdot)$ is only a proportionality constant for the expression inside the integral. The quantity,

$$h(\underline{X};\underline{\phi})\int_{\underline{C}} g(\underline{Y}|\underline{X};\underline{\theta})dy\ , \tag{8}$$

has been called a face value likelihood (Smith and Sugden, 1988) since it only contains variables that are observable in the population, and it assumes that the treatment assignment and sampling mechanisms are unknown to the investigator. When these mechanisms are ignorable for likelihood inference, the face value likelihood will produce the same MLE for $\tau(\theta)$ as the full likelihood in Equation (5). For completeness, the sufficient condition for the sampling mechanism to be ignorable for likelihood inference is, in addition to condition (ii) in Equation (6),

$$\text{(iii)} \qquad \underline{S} \perp \underline{Y}|\underline{X},\underline{W},\underline{Z}\ . \tag{9}$$

When the scientist controls the treatment assignment (i.e., via an experimental design), condition (i) in Equation (6) is met, and when he/she controls the sampling mechanism (i.e., via an analytical survey with primary purpose a comparison between subgroups of the population sampled), condition (iii) in Equation (9) is met. Any effect from a violation of (ii) in Equation (6) would be controlled through a randomized experimental design and a random sampling design.

In summary, meeting condition (i) indicates that the treatment assignment variable $\underline{Z}$ is independent of the response $\underline{Y}$, given the measurable conditions $\underline{X}$ and the unobservable conditions $\underline{W}$. Meeting condition (iii) indicates that the sample selection variable $\underline{S}$ is independent of the response variable $\underline{Y}$, given the measurable conditions $\underline{X}$, the unknown conditions $\underline{W}$, and the treatment variable $\underline{Z}$. Condition (iii) is analogous to condition (i) for the treatment assignment mechanism except that levels of $\underline{Z}$ have already been assigned to population units. Meeting condition (ii) indicates that either $\underline{Y}$ is independent of $\underline{W}$ or the effect of $\underline{W}$ on $\underline{Y}$ can be removed by adjusting $\underline{Y}$ for observable $\underline{X}$ values. In this formulation, $\underline{W}$ has the potential to be the classic 'lurking variable' described in many introductory statistical texts. Lurking variables are unobserved variables that could alter conclusions from a study if they were observed and included in the analysis. Conditions (i) and (ii)

are sufficient for a treatment assignment mechanism to be ignorable. Conditions (ii) and (iii) together are sufficient for sample selection to be ignorable for inference. If all 3 conditions are satisfied, the face value likelihood in Equation (8) leads to the same inference for $\tau(\theta)$ as the full likelihood in Equation (5). Confidence intervals for $\tau(\theta)$ could then be constructed using distributional or large sample properties of MLE's.

3. An Example

3.1. General framework

Assume a population of size $N = 5000$ forest plots in a certain region of interest, a response variable Y being a measure of tree mortality for a plot, and a causal variable Z being an index of drought (inverse amount of rainfall during the months May–August). The true causal relationship is that Z causes a particular effect on Y. Graphically depicted, this is $Z \rightarrow Y$. Let us also assume that the functional relationship relating the two variables is $Y = 5Z + e$, where e is a standard normal random variable representing heterogeneity of potential responses across units at a fixed value of Z. This means, for our example, that plot mortality is 5 times the inverse of the rainfall in May–August of each year with a standard normal error term added. So population units at a fixed value of $Z = z$ will have a mean of $5z$ and vary normally about the mean with variance equal to 1. The treatment effect for the ith unit between two levels of Z, say z_j and $z_{j'}$, is given by,

$$D_i = Y_i(Z = z_j) - Y_i(Z = z_{j'}) = 5(z_j - z_{j'}) + e_i - e_i = 5(z_j - z_{j'}) \,. \quad (10)$$

The effect will be the same for all units $i = 1, 2, \ldots, N$, i.e., the effect of Z on Y is constant across population units, called unit-treatment additivity by others (Hinkelmann and Kempthorne, 1994; Cox, 1992). This means that the effect of drought index is the same on each plot regardless of whether they have different soil types, topography, genetic composition, or stand density. Assuming a constant effect, though unrealistic, helps to illuminate the role of nonrandom assignment mechanisms in drawing inferences. When the effect is constant, the mean treatment effect characterizes the true effect of Z on Y since this effect does not vary from plot to plot.

The mean treatment effect might even be misleading when the treatment effects vary widely from plot to plot in the population (Cox, 1958b), and one cannot directly estimate this variance using observable data (Gadbury and Iyer, 2000). If the variance is not zero, results from this example would still remain valid as long as the mean treatment effect remains the quantity of interest. For example, in a situation where the interest is in the total as in total mortality of forest plots in a population, mean mortality is the quantity of interest since it corresponds directly with it (mean mortality = total mortality/total number of plots). However, if there

is interest in identifying any subset of plots that have high mortality, one might be interested in how the effect of a treatment varies. This variability is a separate issue that is not considered in this paper, but some details regarding it can be found in Gadbury and Iyer (2000).

It is possible to observe either $Y_i(Z = z_j)$ or $Y_i(Z = z_{j'})$, but not both. So it is not possible to observe D_i as defined in Equation (10). However, since the effect of Z on Y is constant, D_i does not depend on i, and $D_i = \bar{D}$ (defined in Equation (1)) for all $i = 1, \ldots, 5000$. Using the notation of the earlier section, the parameter of interest is $\underline{\theta} = (\theta^{(1)}, \ldots, \theta^{(k)})$, where $\theta^{(j)}$, $j = 1, \ldots, k$ is the population mean treatment response if the entire population was exposed to a level of $Z = z_j$. The mean treatment effect between the two levels of Z is then $\tau(\theta) = \theta^{(j)} - \theta^{(j')}$. That is, we are interested in $\theta^{(j)} - \theta^{(j')}$, $j \neq j'$, to determine whether there is in fact an effect of Z on Y.

3.2. IDENTIFYING A CAUSAL HYPOTHESIS

Suppose drought index Z causes a particular effect on plot mortality (response Y) as described above, but is not observable. Instead, a variable V, rainfall at rain gauges established by the U.S. Weather Bureau, is observed that is correlated with levels of Z. This situation can be encountered in observational data when some effect has been observed but it is uncertain what variable has contributed to the effect. For example, suppose some plots have a high mortality and others a low mortality. If the former are close to a measured value of V that is low (i.e., low rainfall), one could conclude that it was lack of rainfall that caused the increased mortality. This conclusion would depend on the unobservable relationship between V and Z. Suppose that this relationship is given by $Z = a/V + b(1/V)^2 + e_2$, where e_2 is a random error term and a and b are constants. So V is producing some level of Z for a plot that causes a particular level of mortality. Graphically, this relationship is $V \rightarrow Z \rightarrow Y$. As noted above, the change in potential responses between two levels of Z for a plot i is again $D_i = 5(z_j - z_{j'})$, which is constant across all population units. The true effect of a change in V on potential responses in Y is then given by $D_i = 5a(1/v_j - 1/v_{j'}) + 5b(1/v_j^2 - 1/v_{j'}^2)$. Because Z is not observable for a particular plot i, the observable responses would be either $Y_i(v_j)$ or $Y_i(v_{j'})$, depending on whether the rainfall received is v_j or $v_{j'}$. The problem with evaluating the relationship between rainfall and mortality is that V is not measured at the plot level*. Because weather stations are usually spaced far apart, it might be assumed for a particular region of a forest near a weather station with, say n_R plots $i = 1, \ldots, n_R$, that the level of rainfall is the same for all plots in the region. In other words, suppose $V = v_j$ is observed at a weather station. It might be assumed that all plots in the region have that level of rainfall or that the drought index measure is $Z = 1/v_j$ for every plot. In this case, one would infer that any

* Even if rainfall could be measured on the plots, the relationship between rainfall and drought index is usually unknown.

difference between mortality levels at different plots is not due to drought index because it was assumed, possibly incorrectly, that the unobserved drought index was the same for every plot. The only way to associate changes in rainfall with changes in mortality would be to compare plots in different regions, where the measured rainfall V is different. But this increases the chances that the observed difference in mortality across the two regions is due to some other unobserved variable – not just lack of rainfall. At this point, one is faced with a search for a cause for some observed effect-something that Rubin (1990) cautioned is nearly impossible.

Still, all is not lost. One can apply more subjective criteria to compare possible hypotheses for a difference in mortality across plots. Certainly, if high mortality was observed on forest plots where the rainfall measured at a weather station was low, it would be plausible to expect that some measure of drought index at the plot level would explain the observed mortality. In fact, 'plausibility' was one of several criteria used for evaluating a causal hypothesis discussed in Hill (1965) and in Susser (1988) and referred to as 'mechanism' by Mosteller and Tukey (1977). That is, lack of rainfall could lead to an ecological process that increases mortality. To proceed further, one would need to demonstrate responsiveness by experimental exposure of plots to various levels of drought index and then observe the associated effect. However, the scientist does not control the levels of drought to which forested plots in a natural population are exposed. We continue the example by assuming that a causal hypothesis has been identified and Z and Y are measured at the plot level. All else is as described in the section, General Framework.

3.3. Estimating a causal effect using likelihood inference

Assume that 'nature' uses some unknown mechanism to assign levels of Z to each of the 5000 units in the population. Without loss of generality, assume that levels of Z are between 0 and 1, where $Z = 0$ implies no drought and $Z = 1$ implies full exposure to drought. Then assume that the investigator obtains a simple random sample (SRS) of $n = 200$ plots from the population of 5000 plots. From the sample, an estimate, $\hat{\tau}(\theta)$, is calculated. Note that for all cases considered below, condition (iii) in Equation (9) is met since SRS is ignorable as discussed earlier.

The following simulations illustrate the effect of an unknown treatment assignment mechanism on conclusions from a study. In the cases considered, an ignorable treatment assignment mechanism was first used followed by a nonignorable one to assess the effect of the mechanism on inference results. Although results might be derived analytically in the simpler cases, to remain consistent throughout, we used simulations to estimate a sampling distribution of $\hat{\tau}(\theta)$. For each of 1000 simple random samples of size $n = 200$, we computed the face value estimate of the treatment effect using Equation (8) and a t-distribution based 95% confidence interval (L, U) for $\tau(\theta)$. Summary statistics for the 1000 values of L and U were computed. The simulated sampling distribution of $\hat{\tau}(\theta)$ was graphically compared

with the true treatment effect that was used to generate the population data. We did not compute the MLE for $\tau(\theta)$ using the full likelihood Equation (5) because in situations where the treatment assignment mechanism is not ignorable, this estimator can quickly become intractable (see Gadbury, 1998, for an example); nor did we concern ourselves with units of measurement. Rather, we assumed, without loss of generality, that units are measured on some scale that is meaningful and that original measurements have been transformed (if necessary) to a scale where normal distribution theory applies. Four cases are considered below. Z is dichotomous in Case 1 and continuous in Case 2. In Case 3, Z is again dichotomous but a covariate is observed, and the effect of an unobservable covariate is considered in Case 4.

3.3.1. *Case 1*

The causal variable Z has only two levels, $z_1 = 1$ and $z_2 = 0$, and the causal effect of Z on Y is constant across the population and equal to $D = 5$. For each plot $i, i = 1, \ldots, N$, there are two potential responses, $Y_i(Z = 1)$ and $Y_i(Z = 0)$, only one of which is observable depending on the outcome of the treatment assignment mechanism for the ith plot. So we estimate that the true effect of severe drought (high drought index) on a plot will be that five times more trees will die than if there was no drought. The estimated mean treatment effect is,

$$\hat{\tau}(\theta) = \bar{d} = \bar{y}_1 - \bar{y}_2$$

where $\bar{y}_1$ is the mean response of sample observations with $z = 1$, and $\bar{y}_2$ is the sample mean response of observations with $z = 0$. Three situations are studied (see below) that represent different probability distributions for Z (i.e., different treatment assignment mechanisms). Results for the three situations are shown in Table I for Case 1.

Situation (a): For each plot $i, i = 1, \ldots, N$, the distribution of Z_i (i.e., the treatment assignment mechanism) can be written as $P(Z_i = 1) = \frac{1}{2}$. Thus, it is independent of Y and should not influence the results of inference. That is, each plot in the population has either a low or high level of drought ($Z = 0$ or 1) each with probability $\frac{1}{2}$. Condition (i) in Equation (6) is met. Furthermore, there is no covariate to confound results so condition (ii) in Equation (6) is met. So sufficient conditions are met for ignorability and valid inferences about the true effect of drought on mortality can be made. Results in Table I show that the true effect of drought was correctly estimated 96% of the time using a 95% confidence interval.

Situation (b): For each plot, the distribution of Z_i is given by $P(Z_i = 0) = 0.25$ and $P(Z_i = 1) = 0.75$. In this situation, the assignment of treatments is biased but, since they are independent of Y, results of inference should be unbiased. This situation is quite similar to (a) except that an expected 75% of the plots were assigned a high level of drought and an expected 25% were not. We know what treatment a plot received, so we can again estimate the true treatment effect unbiasedly but less efficiently than if the probabilities of assigning treatments to plots was considered

TABLE I

Summary statistics for 1000 simulated 95% confidence intervals (L, U)[a] for $\tau(\theta)$ in Cases 1–3. The mean and standard deviation for 1000 values of L and U, respectively, are shown along with the proportional coverage of (L, U). Case 3 includes confidence intervals constructed from the naïve estimator (i) and the adjusted estimator (ii)

		L (mean, st. dev.)	U (mean, st. dev.)	(L, U) coverage
Case 1				
Situation (a)		(4.73, 0.13)	(5.28, 0.13)	0.963
Situation (b)		(4.44, 0.16)	(5.30, 0.16)	0.953
Situation (c)		(5.57, 0.13)	(6.09, 0.13)	0.000
Case 2				
Situation (a)		(4.47, 0.24)	(5.46, 0.24)	0.962
Situation (b)		(4.38, 0.31)	(5.63, 0.31)	0.952
Situation (c)		(6.64, 0.34)	(7.77, 0.34)	0.000
Case 3				
Situation (a)	$\hat{\tau}_\theta^{(i)}$	(4.73, 0.14)	(5.29, 0.13)	0.955
($\rho = 0$)	$\hat{\tau}_\theta^{(ii)}$	(4.73, 0.15)	(5.28, 0.15)	0.933
Situation (b)	$\hat{\tau}_\theta^{(i)}$	(5.08, 0.13)	(5.59, 0.12)	0.248
($\rho = 0.4$)	$\hat{\tau}_\theta^{(ii)}$	(4.77, 0.13)	(5.24, 0.13)	0.933
Situation (c)	$\hat{\tau}_\theta^{(i)}$	(5.44, 0.12)	(5.90, 0.11)	0.000
($\rho = 0.8$)	$\hat{\tau}_\theta^{(ii)}$	(4.90, 0.05)	(5.10, 0.05)	0.933

[a] Confidence intervals constructed using $\hat{\tau}_\theta^{(ii)}$ are approximate intervals so the proportional coverage may differ from nominal 95%.

(i.e., confidence intervals tend to be wider). In this situation we estimate the true treatment effect 95% of the time using a 95% confidence interval.

In situation (c), higher drought levels are assigned more often to plots with above average mortality but we assume that the probabilities of assignment are not known in the actual analysis of the data. For each plot i, the distribution of Z_i is given by,

$$P(Z_i = 1|Y) = 0.25I(e_i \leq 0) + 0.75I(e_i > 0) ,$$

where $I(a)$ is an indicator function equal to 1 if condition (a) is true. This assignment mechanism suggests that plots with an 'above average' potential response (i.e., $e_i > 0$) are more likely to receive the level of $Z = 1$ than those with below

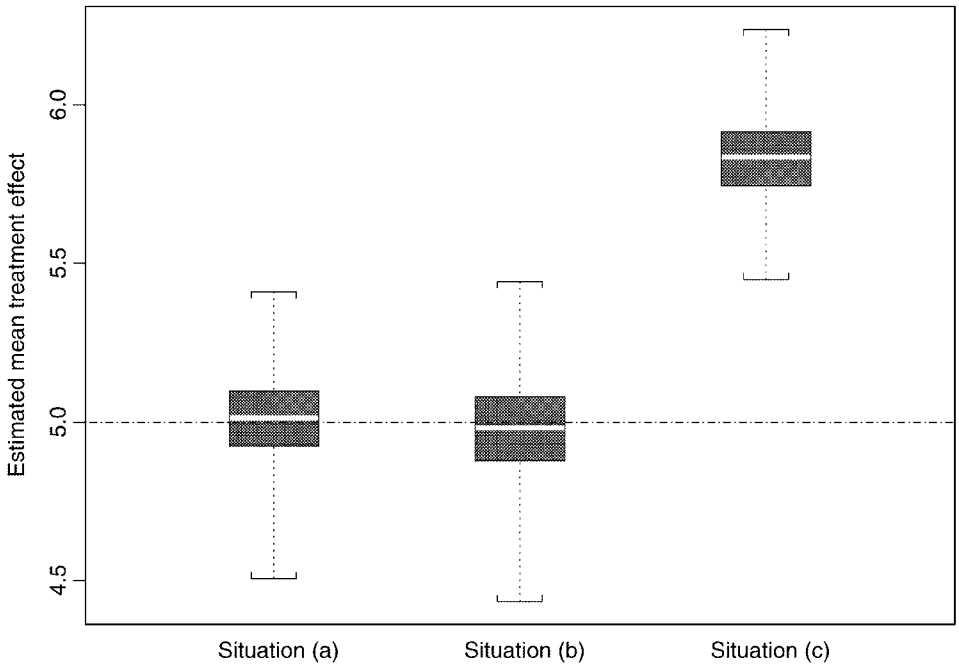

Figure 1. Boxplot of the sampling distribution for the estimator of the mean treatment effect ($\hat{\tau}(\theta)$) for each of the three situations from Case 1. The dotted horizontal line (value 5.0) is the true mean treatment effect.

average potential responses, which implies that treatment is not independent of the potential response, thus violating condition (i) in Equation (6). Results show that the effect is disastrous in inference (Table I, Case 1). The true effect of drought on mortality was not estimated correctly in any of the 1000 simulations.

In situations (a) and (b) we could estimate the average effect of drought (as measured by drought index), but not in (c). Figure 1 graphically depicts a simulated sampling distribution of the estimated average effect of drought on mortality for the three situations in Case 1 using 1000 simple random samples. As shown, the sampling distributions are closely centered on the true effect in situations (a) and (b), but not in (c).

3.3.2. *Case 2*

Case 1 is generalized by allowing Z to be a continuous exposure variable so there is an infinite number of potential responses, Y, each corresponding to a particular value of Z. We now have the more realistic situation that the level of exposure to drought is a continuous variable so that we have a continuous response Y according to the relationship plot mortality = $5 * \text{drought index} + e$. As in Case 1, drought index was assigned to plots with equal (situation (a)) and unequal (situation b) probabilities and then, in situation (c) it was dependent on Y. The treatment effect

for the *i*th plot between two levels of Z, say z_j and $z_{j'}$ is again given by Equation (10). The parameter of interest is $\tau(\theta)$ = the slope parameter, which is $\beta = 5$. The estimated slope, $\hat{\tau}(\theta)$, was obtained by regression of the observed y on z.

Situation (a): Z is distributed according to a continuous uniform distribution over the interval 0 to 1.

Situation (b): The probability distribution of Z is exponential with the mean parameter equal to $\frac{1}{4}$. To remain consistent with the previous case where Z was in the interval [0, 1], generated values of Z greater than 1 were set to $z = 1$.

Situation (c): Let P_z denote the distribution in Situation (b) above, and let P_N be a normal distribution with mean $\frac{1}{2}$ and standard deviation 0.1. Set to zero any values of Z less than zero from P_N and to one any values greater than one. Let P_{NT} denote this truncated normal distribution. The distribution of Z in this situation is a mixed distribution defined by,

$$P(z|Y) = P_z \cdot I(e \leq 0) + P_{NT} \cdot I(e > 0) \ .$$

Plots with a potential response, at any fixed level of Z, less than the population mean are distributed according to P_z, and those above the mean according to P_{NT}. Here, condition (i) in Equation (6) is violated.

Again, we obtained reliable inference in (a) and (b) (95–96% of samples), but inferences were severely biased in (c), not one sample covering the true value. Thus, under situations (a) and (b) one could correctly estimate the cause-effect relationship, but not in (c). Results for these situations are shown in Table I for Case 2 (statistics) and Figure 2 (slope of the regression coefficient *b*).

3.3.3. *Case 3*

Case 3 differs from the previous cases because there is a covariate, X = age of plot, that is correlated with the potential mortality responses. The value of the covariate, of course, is not influenced by the level of Z assigned to a plot. Only high or low drought level, as in Case 1, is considered for Z and the assignment of drought index to plots depends on the plot age X, but not directly on Y. So condition (i) in Equation (6) is met for Case 3. The age for a plot (after centering and scaling) is assumed to be normally distributed with zero mean and variance 1, and mortality was determined by the linear relationship: Mortality = 5 $*$ drought index + $\rho *$ plot age + error e, where e is distributed normally with mean 0 and variance $1 - \rho^2$ (that is, $Y = 5Z + \rho X + e$). So the marginal distribution of Y has mean $5Z$ and unit variance, and the correlation between potential response Y and X is equal to ρ. The true treatment effect of Z on the *i*th plot is,

$$D_i = Y_i^{(1)} - Y_i^0 = 5 + \rho(X_i - X_i) = 5 \ ,$$

so the parameter of interest is again $\tau(\theta) = 5$. Clearly inference depends on the size of ρ, which is set to values ρ = 0.0, 0.4, and 0.8, situations (a), (b), and (c), respectively.

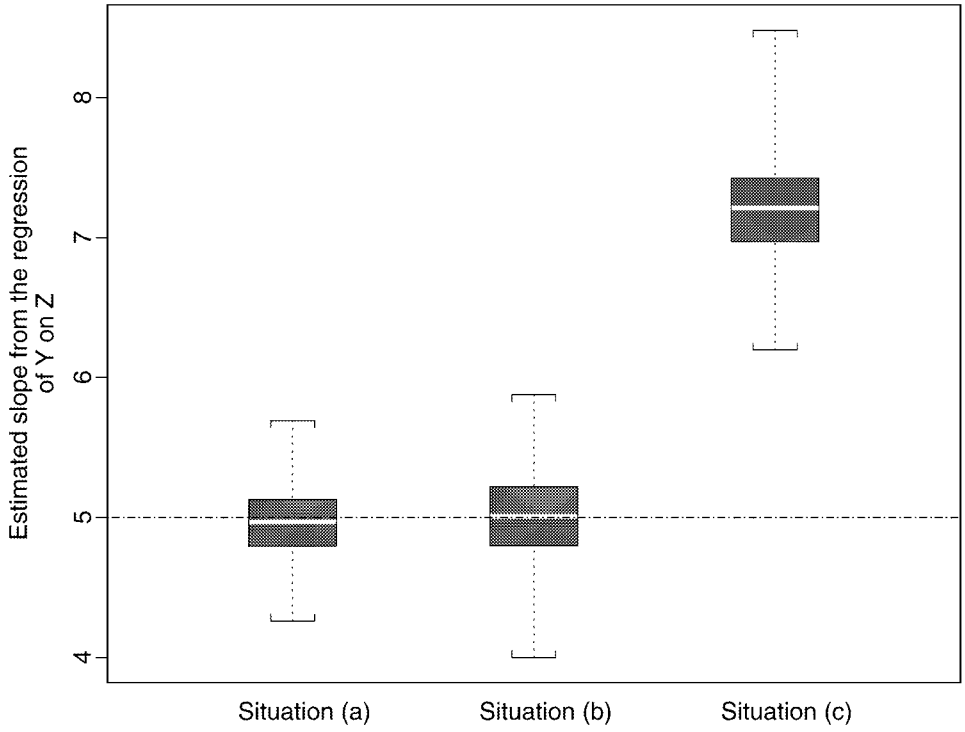

Figure 2. Boxplot of the simulated sampling distribution of the estimated slope (b) of the regression of observed Y on Z for the three situations from Case 2. The true slope is 5.0.

The probability distribution of Z for a plot i, $i = 1, 2, \ldots, 5000$, is given by,

$$P(Z_i = 1 | X_i = x) = 0.75I(x > 0) + 0.25I(x \leq 0) \; . \tag{11}$$

For each situation, two estimators are considered: (i) the naïve estimator of a mean treatment effect if X is not used in the analysis, i.e., $\hat{\tau}_\theta^{(i)} = \bar{y}_1 - \bar{y}_0$ and (ii) the adjusted estimator of the mean treatment effect that assumes X is observed and used in the analysis. The estimator (i) is the same as in Case 1. It can be shown (Lord, 1955) that the estimator (ii) is given by,

$$\hat{\tau}_\theta^{(ii)} = \bar{y}_1 - \bar{y}_0 - b_1(\bar{x}_1 - \bar{x}) + b_0(\bar{x}_0 - \bar{x}) \; , \tag{12}$$

where $\bar{y}_1$ and $\bar{x}_1$ are the observed sample means for observations with $Z = 1$, and b_1 is the sample regression coefficient relating Y to X in the sample with $Z = 1$. The quantities $\bar{y}_0$, $\bar{x}_0$, and b_0 are the analogous estimators for observations with $Z = 0$. Finally, $\bar{x}$ is the mean of the 200 values of X in the sample.

As noted, condition (i) in Equation (6) is met for all situations in Case 3, but condition (ii) is only met if either X is observed and used in the analysis, or X is not observed or not used but the correlation $\rho = 0$. In this latter event, condition

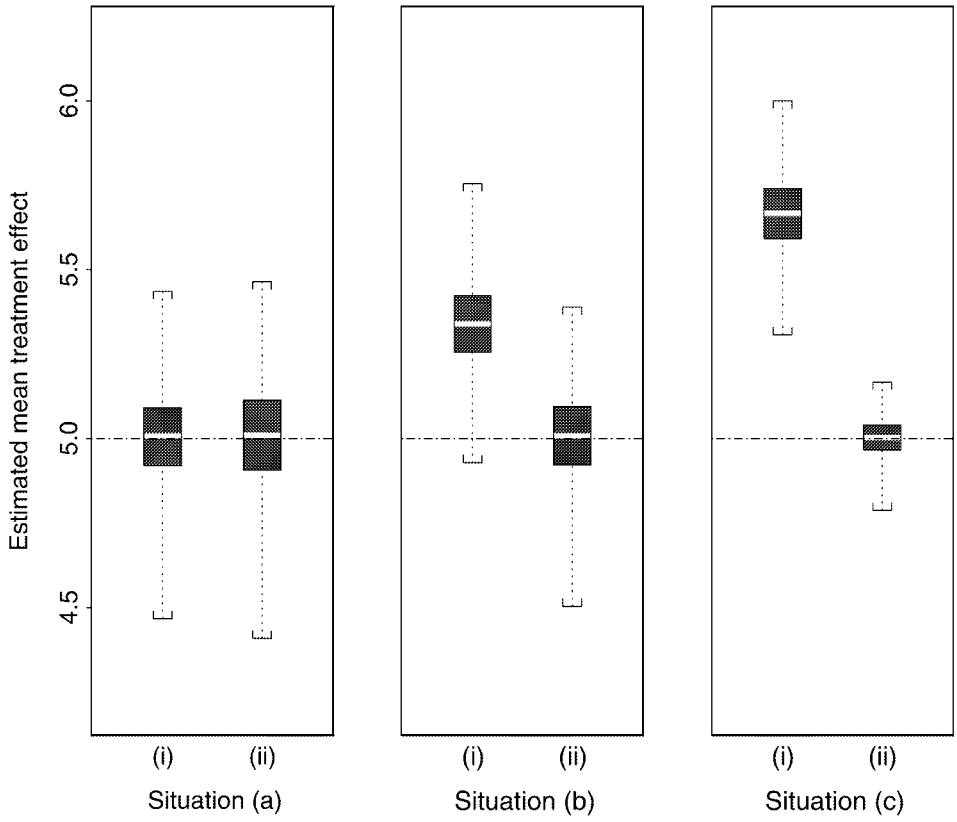

Figure 3. Boxplots of the simulated sampling distribution for the estimator of the mean treatment effect for Case 3. Situation (a), (b), and (c) have $\rho = 0.0$, 0.4, and 0.8, respectively. Plots (i) and (ii) represent the use of estimators $\hat{\tau}^{(i)}_{(\theta)}$ and $\hat{\tau}^{(ii)}_{(\theta)}$, respectively. The value of the true treatment effect is 5.0.

(ii) in Equation (6) is met since potential responses Y are independent of X. Results are shown in Table I for Case 3 (statistics) and Figure 3 (simulated sampling distributions of the two estimators).

In situation (a), there is no correlation between mortality and drought index level, so one would expect the same results as for Case 1. Both the naïve estimator and the adjusted estimator give reasonable results (95 and 93% of confidence intervals covered the true effect), although the second one less so since one is adjusting for X when it is not necessary.

In situations (b) and (c) ($\rho = 0.4$ and 0.8), the naïve estimator (i) gave incorrect inferences (25 and 0% of confidence intervals contained 5). So with this particular treatment assignment mechanism that meets condition (i) in Equation (6), inferences were severely biased since a covariate X is involved but not considered. With the adjusted estimator (ii) where X is used, inference was correct in 93% of samples in both situations. The adjusted estimator had the same coverage for all

three situations since the only parameter that changed was ρ (i.e., simulated values of X were the same for all three situations, and an adjustment for X produced the same coverage, regardless of ρ). Under (a) we estimate the cause-effect relationship correctly regardless of whether plot age is used in the analysis, but in situations (b) and (c), plot age must be used since, otherwise, condition (ii) in Equation (6) would be violated. So, if the covariate could be observed and used in the analysis, conditions for ignorability of the treatment assignment would be met.

3.3.4. *Case 4*

Drought index, Z, is dichotomous taking on values of 0 or 1. There is also a covariate W = drought susceptibility that cannot be observed for any plot in the population. The other variables are the same as Case 3. We assume now that the potential responses $(W, X, Y^{(0)})$ are jointly distributed according to a tri-variate normal distribution so that the three marginal distributions are standard normal (i.e., zero mean and unit variance). The correlation structure is given by,

$$\begin{pmatrix} 1 & \rho_{W,X} & \rho_{W,Y^{(0)}} \\ \rho_{W,X} & 1 & \rho_{X,Y^{(0)}} \\ \rho_{W,Y^{(0)}} & \rho_{X,Y^{(0)}} & 1 \end{pmatrix} = \begin{pmatrix} 1 & -0.5 & r \\ -0.5 & 1 & -0.5 \\ r & -0.5 & 1 \end{pmatrix}, \tag{13}$$

where $\rho_{W,X}$ is the correlation between W and X, and other correlations are similarly defined. The partial correlation between W and Y given X is,

$$\rho_{wy/x} = \left[\rho_{wy} - \rho_{wx}\rho_{yx}\right] \sqrt{[(1-\rho_{wx}^2)(1-\rho_{yx}^2)]}\,.$$

Equation (13) implies that drought susceptibility tends to be higher for younger plots, and that younger plots tend to have higher mortality. Age of a plot X is assumed to be observable for all plots in the population. The constant r is the correlation between drought susceptibility and mortality for plots with $Z = 0$, and its value will be varied in simulations below. The mortality for plots exposed to $Z = 1$ is $Y^{(1)} = Y^{(0)} + 5$, so that the true effect of Z on mortality is equal to 5. The treatment assignment mechanism is given below in the simulations. Once values of Z are assigned to the plots in the population, the observable mortality is $Y = Y^{(0)} + 5Z$ (note that this is the same as for Cases 1 and 3 since the marginal distribution of $Y^{(0)}$ is standard normal). The estimator for $\tau(\theta)$ is $\hat{\tau}_{(\theta)}^{(ii)}$, also used in Case 3 (Equation (12)), since X is observed and used in the analysis, but W is not. Two situations, (a) and (b), are considered.

Situation (a): The treatment assignment, for a plot i, is the same as Case 3, Equation (11), except an unobservable covariate is now present. First, r in Equation (13) equals 0.25 so that the partial correlation between Y and W, given X, is zero. That is, $\rho_{W,Y|X} = 0$ so condition (ii) in Equation (6) is satisfied in addition to condition (i). Second, r is equal to 0.8 so that $\rho_{W,Y|X} = 0.73$ indicating that condition (ii) in Equation (6) is violated. These two simulations are (a1) and (a2), respectively, in Table II.

TABLE II

Summary statistics for 1000 simulated 95% confidence intervals (L, U)[a] for $\tau(\theta)$ in Cases 4a and 4b. The mean and standard deviation for 1000 values of L and U, respectively, are shown along with the proportional coverage of (L, U)

	L (mean, st. dev.)	U (mean, st. dev.)	(L, U) coverage
Case 4a			
Simulation (a1)	(4.75, 0.13)	(5.23, 0.13)	0.943
Simulation (a2)	(4.78, 0.13)	(5.26, 0.13)	0.931
Case 4b			
Simulation (b1)	(4.73, 0.12)	(5.21, 0.12)	0.949
Simulation (b2)	(4.76, 0.12)	(5.24, 0.12)	0.955
Simulation (b3)	(4.88, 0.12)	(5.36, 0.12)	0.843

[a] Confidence intervals constructed using $\hat{\tau}_\theta^{(ii)}$ are approximate intervals so the proportional coverage may differ from nominal 95%.

Table II shows that the true effect of Z on mortality is correctly estimated in 94 and 93% of samples using 95% confidence intervals. In simulation a2, the violation of condition (ii) did not have any appreciable effect since W is not directly involved in the treatment assignment. However, it does appear that there is a slight imbalance in the distribution of W across the two treatment groups since confidence intervals covered the true effect slightly less than nominal coverage. This can always occur in randomized experiments, but its effect on conclusions is controlled through a random treatment assignment. As an extreme but unrelated example, consider two treatments that are randomly assigned to a group of male and female subjects. It is possible that a completely random treatment assignment would result in treatment 1 being assigned to male and treatment 2 to female subjects. So the treatment might be confounded with gender effects. This particular treatment assignment, though, would have a low probability of occurring. Recall that Smith and Sugden's conditions (1988) were sufficient for ignorability but not necessary. These simulations suggest that there could be lurking variables that are correlated with the response, but they might not significantly interfere with analysis results unless they are involved in the probabilities of treatment assignment. This latter condition is illustrated next, in situation (b).

Situation (b): The treatment assignment mechanism is now given by,

$$P(Z_i = 1 | X_i = x, W_i = w) = 0.5I(x > 0 \cup w > 0) + pI(x \leq 0 \cap w \leq 0),$$

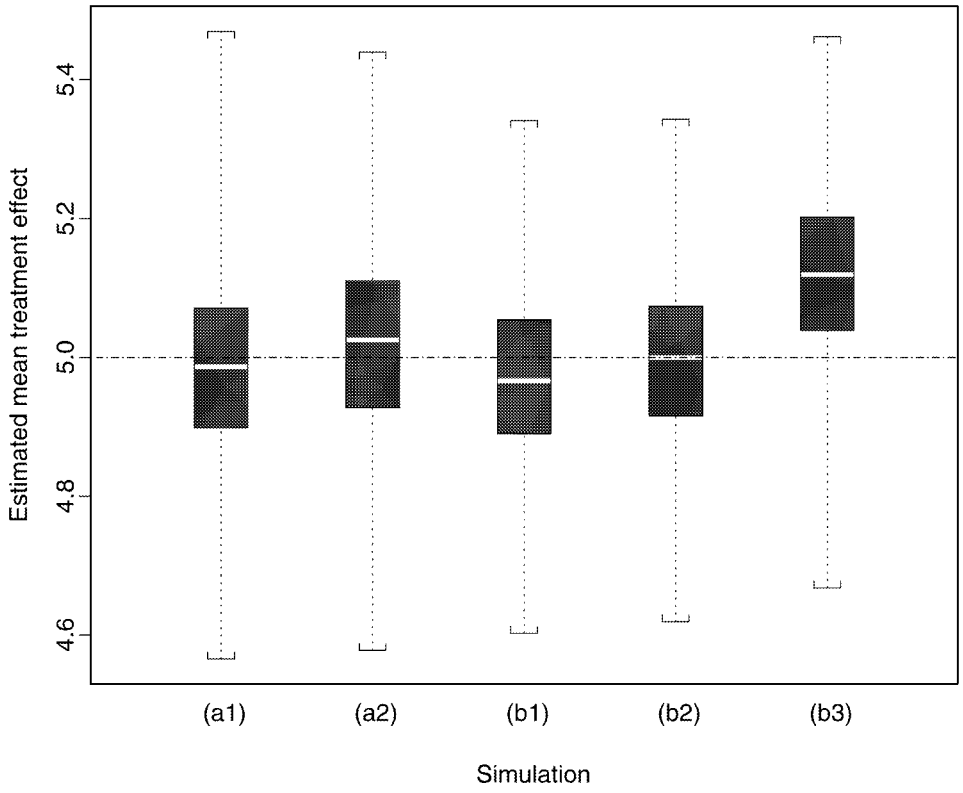

Figure 4. Boxplot of the simulated sampling distribution for the estimator of the mean treatment effect from Case 4. The value of the true treatment effect is 5.0.

where $\cap$ and $\cup$ represent 'and' and 'or', respectively. This treatment assignment mechanism indicates that for plots with either age *or* drought susceptibility greater than zero, assignment of the two levels of Z is done with equal probability. However, for plots with age *and* drought susceptibility less than or equal to zero, assignment of $Z = 1$ to a plot is done with probability equal to p and $Z = 0$ with probability $1-p$. Condition (i) in Equation (6) is met but condition (ii) is not unless the partial correlation of W and Y given X equals zero or $p = 1/2$. Two values of r equal to 0.25, and 0.80 are again considered so that the partial correlation between W and Y, given X, is 0 and 0.73, respectively. We considered values of $p = 0.25$ and 0.50. In simulation (b1) $r = 0.25$ and $p = 0.25$, in (b2) $r = 0.80$ and $p = 0.5$, and in (b3) $r = 0.80$ and $p = 0.25$. Results are shown in Table II for Case 4b.

Results show that when drought susceptibility is independent of mortality, given age, treatments can be assigned to plots with unequal probabilities ($p = 0.25$) and inference will be accurate (simulation (b1)). However, if drought susceptibility is correlated with mortality, given age, then if $p = 0.5$ inference will remain accurate (simulation (b2)).

In simulation (b3), drought susceptibility and mortality are correlated given age. Of the 5000 plots in the simulated population, treatment is assigned with equal probabilities to 4116 (4116 plots had either $W > 0$ or $X > 0$) and with unequal probabilities to the other 884 which had both X and W less than or equal to zero. Results of inference become biased for the true effect of Z on mortality. This bias is more easily seen in Figure 4 where the sampling distribution of $\hat{\tau}_{(\theta)}^{(ii)}$ is nearly centered on the true treatment effect in all simulations except (b3) as expected.

4. Summary and Conclusions

It is clear that establishing cause-effect is difficult, especially since the true cause and/or how to measure it is often unknown. Usually we only have measurable covariates that we believe are related to a possible causal agent. It is therefore not surprising that the final chapter in establishing cause-effect, the actual biological mechanism by which a cause results in an effect, usually takes a long time to discover as in, for example, how cigarette smoking causes cancer (Peifer, 1997).

Ideally, investigations into the causal connection between an exposure variable and a response in the population of interest would use an experimental design combined with a survey sampling design. This combination can assure that both the sampling mechanism and the treatment assignment mechanism are controlled by the scientist, and that results of inference should be free of systematic errors, i.e., bias. Without this control, bias can be present in results and conclusions may be wrong. Such designs assume that a suspected causal variable has been defined, and steps have been taken to measure its effect on a response of interest.

When the causal variable has not been identified or defined, one can resort to subjective checklists to form an elaborate theory that attempts to consider all possible variables that could have produced the observed effect on the response (see Olsen and Schreuder, 1997). Once this has been done, a study design can be developed that will measure the necessary variables on population units. However, in survey designs, one still does not know how treatments were assigned and this can bias results of inference.

This problem was illustrated in simulations where nature exposed forest plots to drought and, then, we randomly sampled from the exposed population. Our focus was the treatment assignment mechanism. The sufficient conditions for ignorability of this mechanism are statistical in nature, and the practical effect of a violation may not be well understood without the use of some constructed numerical examples. We showed that subtle violations of the ignorability conditions could have extreme consequences on the ability to correctly estimate a causal effect. In practice, one will not know for certain that these conditions are met; one must make assumptions about the structure of the data and, in particular, how observed data were produced.

We also defined a sampling mechanism but did not study the effect of a non-ignorable sampling mechanism in the simulations. A sampling mechanism that depends on potential responses or on unobserved covariates that are correlated with potential responses will introduce bias similar to that produced by non-random treatment assignment. A well-designed survey can help assure that the sampling mechanism is ignorable and will not bias results of inference. One ignorable sampling mechanism is simple random sampling, but unequal probability sampling, including stratified random sampling designs, would also be ignorable if covariate values that were used to assign sampling probabilities were observed.

We conclude with the following recommendations.

(1) Identifying cause-effect hypotheses

a. Have a clear understanding of what you can and cannot do with your data. Often, there are inadequate data to establish cause-effect, but there may be enough to identify interesting hypotheses. Making irresponsible claims is dangerous.
b. Detecting major effects is usually necessary to establish causation.
c. Practical steps needed:

1. Involve a statistician early in survey and experimental designs.
2. Collect high quality data.
3. Study all possible explanations for the data, i.e., construct elaborate theories, such as in Schulze (1989) for example.
4. Carefully identify all relevant covariates and measure them if possible.
5. Submit findings to critical reviewers for comments and seriously consider them in revisions.

(2) Establish cause-effect:

a. Expect a need to conduct experiments if the hypotheses were obtained from survey data.
b. Expect a need for analytical surveys if hypotheses were obtained from experimental data.
c. If causal effects are highly variable across plots, stratify plots into subpopulations where effects may be assumed constant. See Gadbury and Iyer (2000) for more information on plot-treatment interaction.
d. If experiments and/or analytical surveys are not possible, then be aware of the effects of nonignorability. It may be possible to assess the sensitivity of inference results to various degrees of nonignorability by using simulations or other analytical methods (e.g., Rosenbaum, 1995).

e. Avoid sources of bias (e.g., assuming ignorability when one cannot). Clearly indicate what is assumed and what the implications might be in terms of bias.
f. Know treatment assignment and sample selection probabilities if possible. If not available, expect inference to be quite limited in scope and clearly state this in any communication.
g. Submit findings to critical reviewers for comments.

Acknowledgements

The authors acknowledge constructive comments from Geoffrey B. Wood, Hari K. Iyer, Sally Lin, and Paul L. Patterson.

References

Bechtold, W. A., Ruark, G. A. and Loyd, F. T.: 1991, 'Changing stand structure and regional growth reductions in Georgia's natural pine stands. *For. Sci.* **37**, 703–717.

Cochran, W. G.: 1965, 'The planning of observational studies of human populations (with discussion)', *J. Roy. Stat. Soc. A.* **128**, 234–265.

Cox, D. R.: 1958a, *The Planning of Experiments*, John Wiley & Sons, New York.

Cox, D. R.: 1958b, 'The interpretation of the effects of non-additivity in the Latin square', *Biometrika* **45**, 69–73.

Cox, D. R.: 1992, 'Causality, some statistical aspects', *J. Roy. Stat. Soc. A.* **155**, 291–301.

Feinstein, A. R.: 1988, 'Scientific standards in epidemiological studies of the menace of daily life', *Science* **242**, 1257–1263.

Gadbury, G. L.: 1998, 'Causal Inference in Randomized Experiments and Observational Studies', *Ph.D. Dissertation*, Colorado State Univ., Fort Collins, CO, 210 p.

Gadbury, G. L. and Iyer, H. K.: 2000, 'Unit-treatment interaction and its practical consequences', *Biometrics* **56**, 882–885.

Glymore, C.: 1986, 'Statistics and metaphysics (comment on Holland 1986, 'Statistics and causal inference'), *J. Amer. Stat. Assoc.* **81**, 964–966.

Hill, A. B.: 1965, 'The environment and disease: Association or causation?', *Proc. Roy. Soc. Med.* **58**, 295–300.

Hinkelmann, K. and Kempthorne, O.: 1994, *Design and Analysis of Experiments*, Vol. 1, John Wiley & Sons, New York, 495 p.

Holland, P. W.: 1986, 'Statistics and causal inference (with discussion)', *J. Amer. Stat. Assoc.* **81**, 945–970.

Lachin, J. M.: 1988, 'Statistical properties of randomization in clinical trials', *Con. Clinical Trials* **9**, 289–311.

Lord, F. M.: 1955, 'Equating test scores – A maximum likelihood solution', *Psychometrika* **20**, 193–200.

Mosteller, F. and Tukey, J. W.: 1977, *Data Analysis and Regression*, Addison-Wesley Publishing Co., Reading, MA.

Olsen, A. R. and Schreuder, H. T.: 1997, 'Perspectives on large-scale natural resource surveys when cause-effect is a potential issue', *Env. Ecol. Stat.* **4**, 167–180.

Peifer, M.: 1997, 'Cancer-beta-catenin as oncogene: The smoking gun', *Science* **275**, 1752–1753.

Rosenbaum, P. R.: 1984, 'From association to causation in observational studies: The role of tests of strongly ignorable treatment assignment', *J. Amer. Stat. Assoc.* **79**, 41–48.

Rosenbaum, P. R.: 1995, *Observational Studies*, Springer-Verlag, N.Y., 230 p.

Rubin, D. B.: 1974, 'Estimating causal effects of treatment in randomized and nonrandomized studies', *J. Educ. Psych.* **66**, 688–701.

Rubin, D. B.: 1976, 'Inference and missing data', *Biometrika* **63**, 581–592.

Rubin, D. B.: 1990, 'Formal modes of statistical inference for causal effects', *J. Stat. Plann. Inference* **25**, 279–292.

Rubin, D. B.: 1991, 'Practical implications of modes of statistical inference for causal effects and the critical role of the assignment mechanism', *Biometrics* **47**, 1213–1234.

Schreuder, H. T. and Thomas, C. E.: 1991, 'Establishing cause-effect relationships using forest survey data (with discussion)', *For. Sci.* **37**, 1497–1525.

Schulze, E. D.: 1989, 'Air pollution and forest decline in a spruce (*Picea abies*) forest', *Science* **244**, 776–783.

Sheffield, R. M., Cost, N. D., Bechtold, W. A. and McClure, J. P.: 1985, 'Pine Growth Reductions in the Southeast', *USDA For. Serv. Res. Bull. SE-83*, Southeastern Forest Experiment Station.

Smith, T. M. F. and Sugden, R. A.: 1988, 'Sampling and assignment mechanisms in experiments, surveys, and observational studies', *Intern. Stat. Rev.* **56**, 165–180.

Susser, M.: 1988, 'Falsification, Verification, and Causal Inference in Epidemiology: Reconsiderations in the ' Light of Sir Karl Popper's Philosophy', in K. J. Rothman (ed.), *Causal Inference*, Epidemiology Resources Inc., Chestnut Hill, MA.