Published in final edited form as:

Stat Med. 2013 November 10; 32(25): 4319–4330. doi:10.1002/sim.5828.

Causal inference, probability theory, and graphical insights

Stuart G. Baker*

*Biometry Research Group, National Cancer Institute, Bethesda, MD 20892 USA. sb16i@nih.gov

Abstract

Causal inference from observational studies is a fundamental topic in biostatistics. The causal graph literature typically views probability theory as insufficient to express causal concepts in observational studies. In contrast, the view here is that probability theory is a desirable and sufficient basis for many topics in causal inference for the following two reasons. First probability theory is generally more flexible than causal graphs: besides explaining such causal graph topics as M-bias (adjusting for a collider) and bias amplification and attenuation (when adjusting for instrumental variable), probability theory is also the foundation of the paired availability design for historical controls, which does not fit into a causal graph framework. Second probability theory is the basis for insightful graphical displays including the BK-Plot for understanding Simpson's paradox with a binary confounder, the BK2-plot for understanding bias amplification and attenuation in the presence of an unobserved binary confounder, and the PAD-Plot for understanding the principal stratification component of the paired availability design.

Keywords

BK-Plot; causal graph; confounder; instrumental variable; observational study; Simpson's paradox

1. Introduction

Perhaps the greatest concern in the analysis of observational data in a clinical setting is the lack of complete knowledge as to why a person receives a particular treatment [1]. For this reason the causal analysis of observational studies is a challenge. Two common types of observational studies are multivariate adjustments with concurrent controls and analyses based on before-and-after studies.

Causal graphs are one approach to understanding bias with multivariate adjustment when estimating treatment effect in observational studies [2]. In these multivariate adjustments lack of knowledge as to why a person receives treatment can be viewed as an unobserved confounder that directly influences treatment received and outcome. An unobserved confounder leads to biased estimates of causal effect if not included in a multivariate adjustment. Examples of unobserved confounders are unrecorded information on treatment history, disease history, or symptoms.

The causal graph literature has contributed new ideas to understanding bias in multivariate adjustments, such as M-bias when incorrectly adjusting for a collider [3–10] and bias amplification or attenuation when incorrectly adjusting for an instrumental variable in the presence of an unobserved confounder [11–16]. Some of the causal graph literature claims

that probability theory is insufficient for understanding causal inference [2]. The view here is that probability theory is a desirable and sufficient basis for causal inference in multivariate adjustments if there is no adjustment for a variable that is a consequence of treatment. Adjusting for a variable that both affects outcome and is a consequence of treatment is well-known to yield biased estimates of treatment effect [17].

A probability theory viewpoint to understanding causal inference in observational studies has two desirable aspects. First probability theory is flexible: besides explaining M-bias and bias amplification and attenuation, probability theory is the basis for the paired availability design for historical controls [18, 19], a method outside the purview of causal graphs. Second probability theory leads to graphical insights: the BK-Plot for understanding Simpson's paradox with a binary confounder, the BK2-plot for understanding bias amplification and attenuation in the presence of an unobserved binary confounders, and the PAD-Plot for understanding the principal stratification [20] component of the paired availability design. Graphical approaches have a long history of providing insight in mathematics such as the "look-see" proof of the Pythagorean theorem [21] and the sum of an infinite geometric series [22].

To put both probability theory and causal graphs into perspective, it is worth noting that there are topics in causal inference for biology and medicine in which neither applies. Examples include downward causation [23] and simply distinguishing cause from effect as whether mutations cause cancer or cancer causes mutations [24–26].

2. Simpson's paradox: The BK-Plot

Simpson's paradox is perhaps best summarized by the iconic phrase of the noted statistician Thomas Louis "Good for men, good for women, bad for people" [27], which applies to the effect of treatment on outcome in causal graph (a) where X is treatment, Y is outcome, and U is sex. The correct approach is to adjust for U yielding the conclusion that treatment is beneficial.

There are two major types of discussions regarding Simpson's paradox. In both cases (essentially by definition) if the paradox is examined in the appropriate way, the paradox disappears. One type of discussion focuses why adjustment for U is correct in causal graph (a) while an unadjusted estimate is correct for causal graphs (b) and (c) [2, 28, 29]. In causal graph (b) U is a consequence of treatment, so adjustment is not appropriate. Causal graph (c) illustrates M-bias, a topic of considerable debate in *Statistics in Medicine* [4–9]. The key aspect of this graph is that U is a collider (two arrows point to it). Because U is a collider, X and Y are independent on the back-door path ($X \leftarrow Q \rightarrow U \leftarrow R \rightarrow Y$) but adjusting for U makes X and Y dependent on this back-door path. Therefore for causal graph (c) adjustment on U is not appropriate. Although this is a well-known result in the causal graph literature [2], a simple proof based solely on probability theory is given in the Appendix A. In more complicated causal graphs involving multiple colliders, the back-door criterion can be used to determine if X and Y are independent on the back-door path [2]. An open question is how often M-bias arises in practice, as an analyst would need to adjust for U but not R or Q . However, if Q were known, the analyst would adjust for Q so M-bias would not arise.

A second type of discussion of Simpson's paradox involves understanding why a reversal of signs occurs with the crude versus the adjusted risk difference for the scenario depicted in causal graph (a) when U is a binary variable. Table 1 presents a numerical example. Here the BK-Plot provides useful insight by graphically summarizing the probability calculations with mixtures of binary variables. The BK-Plot was developed independently by Jeon *et al.* [30] and Baker and Kramer [27] and given its name by Howard Wainer [31]. The BK-Plot has also been used to illustrate calculations involving missing data [32], the transitive fallacy in randomized trials [33], and binary surrogate endpoints [34].

The arrows in causal graph (a) imply a joint probability distribution (combining both front and back-door paths) that is initially factored as $\text{pr}(Y=1, U=u, X=x) = \text{pr}(Y=1 | U=u, X=x) \text{pr}(X=x | U=u) \text{pr}(U=u)$. The first step in formulating a BK-Plot is to rewrite this joint distribution as

$$\text{pr}(Y=1, U=u, X=x) = \text{pr}(Y=1 | X=x) \text{pr}(U=u | X=x) \text{pr}(X=x). \quad (1)$$

In this formulation the causal effect is summarized by a single parameter so that

$$\text{pr}(Y=1 | U=u, X=1) = \text{pr}(Y=1 | U=u, X=0) + \Delta. \quad (2)$$

The conditional risk difference is $\text{pr}(Y=1 | U=u, X=1) - \text{pr}(Y=1 | U=u, X=0) = \Delta$. The crude risk difference, which is biased because it does not adjust for U , is

$$\text{crudeRD} = \text{pr}(Y=1 | X=1) - \text{pr}(Y=1 | X=0), \text{ where } \text{pr}(Y=1 | X=x) = \sum_u \text{pr}(Y=1 | U=u, X=x) \text{pr}(U=u | X=x). \quad (3)$$

The adjusted risk difference equals the crude risk difference with $\text{pr}(U=u)$ substituted for $\text{pr}(U=u | X=x)$ in equation (3),

$$\text{adjustedRD} = \sum_u \text{pr}(Y=1 | U=u, X=1) \text{pr}(U=u | X=1) - \sum_u \text{pr}(Y=1 | U=u, X=0) \text{pr}(U=u | X=0). \quad (4)$$

The BK-Plots at the top of Figure 2 displays the above algebraic equations corresponding to Table 1. Each diagonal line plots $\text{pr}(Y=1 | x)$ as a function of $\text{pr}(U=1 | x)$. The diagonal lines are parallel because there is no interaction between treatment X and confounder U in equation (2). The conditional RD is $0.30 - 0.20 = 0.10$ for $U=0$ and $0.90 - 0.80 = 0.10$ for $U=1$.

The plot on the top left of Figure 2 graphically shows the computation of the crude RD using different dashed vertical lines for $\text{pr}(U=1 | X=0)$ and $\text{pr}(U=1 | X=1)$, where the crude RD is the difference between the horizontal dashed lines. Here the crude RD is $\{(0.30)(0.75) + (0.90)(0.25)\} - \{(0.20)(0.25) + (0.80)(0.75)\} = 0.45 - 0.65 = -0.20$.

The plot on the top right of Figure 2 shows the computation of the adjusted RD using a single dashed vertical line for $\text{pr}(U=1 | x) = \text{pr}(U=1)$. Graphically the adjusted risk difference is the vertical difference between the diagonal lines. Here the adjusted RD is $\{0.30 \text{pr}(U=0) + 0.90 \text{pr}(U=1)\} - \{0.80 \text{pr}(U=1) + 0.20 \text{pr}(U=0)\} = 0.10$.

The BK-Plot visually shows that the crude RD is biased when the distribution of U differs by treatment group, as indicated by the horizontal shift in the dashed vertical lines that leads to a reversal of signs for the crude RD versus the adjusted RD . If there were an interaction

between U and X in equation (2), the diagonal lines would not be parallel, but the reversal of signs could still occur.

3. An unobserved binary confounder: The BK-Plot

The BK-Plot also provides insight into bias from an unobserved binary confounder. Examples of unobserved binary confounders include presence or absence of previous treatments [1], complete or incomplete removal of tumors prior to therapy [1], presence or absence of intense pain in women in labor [18], or early or late treatment initiation relative to the development of disease. An unobserved binary confounder also summarizes a combination of factors for low or high risk of outcome that is related to choice of treatment.

Figure 2 can be viewed as BK-Plot for an unobserved binary confounder under three scales: RD , relative risk (RR) and the odds ratio (OR) [35]. It is well known that if there is no interaction between X and U , the conditional RD equals the adjusted RD , the conditional RR equals the adjusted RR , but the conditional OR does not equal the adjusted OR [35,36]. Consequently RD and RR , but not OR , are said to be collapsible [37]. Collapsibility of RD and RR , but not OR , is illustrated on the right side of Figure 2. In this scenario with no interaction between X and U , adjusted values of RD and RR , but not OR , are valid for other distributions of U , an important consideration in the meta-analysis of randomized trials when U is unobserved [35].

4. Bias amplification and attenuation: BK-2 Plot

In recent years there has been considerable interest in understanding how bias from an unobserved confounder is amplified or attenuated by adjustment for an instrumental variable, which is a variable that only influences outcome through treatment [11–16]. One example of an instrumental variable is the clinic where a patient is seen, when each clinic preferentially administers a given treatment but otherwise provides the same patient management. Another example is glaucoma diagnosis, which affects whether statin or glaucoma drugs are received, but is not thought to be directly to incidence of hip fracture [26]. In practice an investigator may be uncertain if a variable is instrumental, so includes it in the multivariate adjustment in case it may be a confounder. An important issue is how this incorrect adjustment affects bias.

Two motivating examples are presented in which the statistic of interest is RD . Variable U is the confounder and Z is the instrumental variable. The hypothetical data in Table 2 illustrate bias amplification. The conditional RD (for each stratum of U and Z) is 0.10. The crude RD (combining tables for U and Z) is -0.068 , which translates into a bias of magnitude 0.168. In contrast the Z -adjusted RD (after combining tables over U) is -0.11 which translates into a larger bias of magnitude 0.21—hence bias amplification. The hypothetical data in Table 3 illustrate bias attenuation. The conditional RD , which is the causal effect, is 0.10. The crude RD is -0.25 , which translates into a bias of magnitude 0.35. In contrast the Z -adjusted RD is -0.11 , which translates into a smaller bias of magnitude 0.21—hence bias attenuation. The BK2-Plots in Figures 3 and 4 visually explain these results.

We discuss bias amplification and attenuation in the situations depicted in the causal graphs (d) and (e) in Figure 1 where U is an unknown binary confounder. These causal graphs summarize a key assumption, namely Y is independent of Z given X and U . Causal graph (d) defines Z as an instrumental variable [27]. For causal graph (d) the joint distribution of the variables, as indicated by the arrows, is $\text{pr}(Y=1, U=u, Z=z, X=x) = \text{pr}(Y=1 | x, u) \text{pr}(X=x | u, z) \text{pr}(Z=1) \text{pr}(U=u)$. For causal graph (e), the joint distribution of the variables, as indicated by the arrows, is $\text{pr}(Y=1, U=u, Z=z, X=x) = \text{pr}(Y=1 | x, u) \text{pr}(X=x | u, z) \text{pr}(Z=1, U=u)$. We can rewrite the joint distribution of the variables in causal graphs (d) and (e) as

$$\text{pr}(Y=1, U=u, Z=z, X=x) = \text{pr}(Y=1 | x, u) \text{pr}(U=u | x, z) \text{pr}(Z=1 | x) \text{pr}(X=x) \quad (5)$$

Because the focus is the effect of X on Y , it is not a concern that equation (5) does not preserve the information in causal graph (d) that U and Z are independent.

To gain insight into how bias amplification and attenuation plays out, a particular parameterization is needed. The BK2-Plot is based on the following parameterization for equation (5),

$$\text{Model Y: } \text{pr}(Y=1 | x, u) = \beta + \beta_U u + \beta_X x,$$

$$\text{Model U: } \text{pr}(U=1 | x, z) = \alpha + \alpha_Z z + \alpha_X x + \alpha_{ZX} x z,$$

$$\text{Model Z: } \text{pr}(Z=1 | x) = \gamma + \gamma_X x,$$

where the binary variables u , x , and z are set equal either 0 or 1. In Model Y, the causal effect of X on Y is denoted by β_X . For simplicity, there is no modification of the causal effect by an interaction between X and U . The BK2-Plot implicitly requires that the parameters yield probabilities between 0 and 1. Because the purpose of the BK2-Plot is explanation, and not estimation, there is no concern about parameters lying outside admissible values.

The BK2-Plot involves the following two algebraic derivations of risk difference, which extend those for a single binary confounder [13, 14, 28] to include the additional variable Z

4.1 Derivation I

Let $f_{xz} = \text{pr}(Y=1 | x, z) = \sum_u \text{pr}(Y=1 | x, z, u) \text{pr}(U=u | x, z)$ so that, under the models,

$$f_{00} = \beta(1 - \alpha) + (\beta + \beta_U)\alpha, \quad (6)$$

$$f_{10} = (\beta + \beta_X)(1 - \alpha - \alpha_X) + (\beta + \beta_X + \beta_U)(\alpha + \alpha_X), \quad (7)$$

$$f_{01} = \beta(1 - \alpha - \alpha_Z) + (\beta + \beta_U)(\alpha + \alpha_Z), \quad (8)$$

$$f_{11} = (\beta + \beta_X)(1 - \alpha - \alpha_Z - \alpha_X - \alpha_{ZX}) + (\beta + \beta_X + \beta_U)(\alpha + \alpha_Z + \alpha_X + \alpha_{ZX}). \quad (9)$$

Let $f_X = \text{pr}(Y=1 | X=x)$. The crude risk difference is the risk difference not adjusting for any variable, namely,

$$RD_{crude} = f_1 - f_0 = \sum_z f_{1z} \text{pr}(Z=z|X=1) - \sum_z f_{0z} \text{pr}(Z=z|X=0) \quad (10)$$

$$= \beta_X + \beta_U \alpha_X + \beta_U \gamma_X \alpha_Z + \beta_U (\gamma + \gamma_X) \alpha_{ZX}. \quad (11)$$

The risk difference conditional on stratum z is

$$RD_{cond(z)} = f_{1z} - f_{0z} = \beta_X + \beta_U \alpha_X + \beta_U \alpha_{ZX} z. \quad (12)$$

The Z-adjusted risk difference is obtained by substituting $\text{pr}(Z=z) = \gamma$ for $\text{pr}(Z=z | X=x)$ in equation (10) to yield

$$RD_{Zadj} = \sum_z f_{1z} \gamma - \sum_z f_{0z} \gamma = \beta_X + \beta_U \alpha_X + \beta_U \gamma \alpha_{ZX}. \quad (13)$$

As required, $RD_{Zadj} = RD_{crude}$ when $\gamma_X = 0$.

4.2 Derivation II

Using the identity $\text{pr}(U=1|x) = \sum_z \text{pr}(U=1 | x, z) \text{pr}(Z=z | x)$, let

$$\phi = \text{pr}(U=1|X=0) = \alpha(1 - \gamma) + (\alpha + \alpha_Z)\gamma, \quad (14)$$

$$\phi + \phi_X = \text{pr}(U=1|X=1) = (\alpha + \alpha_X)(1 - \gamma - \gamma_X) + (\alpha + \alpha_Z + \alpha_X + \alpha_{ZX})(\gamma + \gamma_X), \quad (15)$$

where $\phi_X = \alpha_X + \gamma_X \alpha_Z + (\gamma + \gamma_X) \alpha_{ZX}$. The identity $f_X = \sum_u \text{pr}(Y=1 | x, u) \text{pr}(U=u | x)$ gives

$$\begin{aligned} RD_{crude} &= f_1 - f_0, \text{ where} \\ f_0 &= \beta(1 - \phi) + (\beta + \beta_U)\phi = \beta + \beta_U \phi, \\ f_1 &= (\beta + \beta_X)(1 - \phi - \phi_X) + (\beta + \beta_U + \beta_X)(\phi + \phi_X) = \beta + \beta_X + \beta_U \phi + \beta_U \phi_X. \end{aligned} \quad (16)$$

As will be shown, the BK2-Plot graphically derives RD_{crude} from equation (16) and graphically derives RD_{Zadj} by setting $\gamma_X = 0$ in equation (15) before substitution into equation (16).

4.3 Bias amplification and attenuation

Bias amplification (attenuation) refers to a larger (smaller) bias when estimating causal effect in the presence of the unobserved confounder when using RD_{Zadj} instead of RD_{crude} . These biases are

$$Bias_{crude} = RD_{crude} - \beta_X = \beta_U \alpha_X + \gamma_X \beta_U \alpha_Z + \beta_U (\gamma + \gamma_X) \alpha_{ZX}, \quad (17)$$

$$Bias_{Zadj} = RD_{Zadj} - \beta_X = \beta_U \alpha_X + \beta_U \gamma \alpha_{ZX}. \quad (18)$$

Appendix B presents the conditions for bias amplification and attenuation based on equations (17) and (18).

4.4 The BK2-Plot

The BK2-Plot provides insight into bias amplification and attenuation. To simplify the graphical display, α_{XZ} is set equal to 0. Consequently, a key determinant of whether bias amplification or bias attenuation occurs is the sign of the parameter γ_X (with α_X and α_{ZX} already specified). The BK2-Plot in Figure 3 shows bias amplification corresponding to $\gamma_X > 0$. The BK2-Plot in Figure 4 shows bias attenuation corresponding to $\gamma_X < 0$.

The four top panels, which are based on Derivation I, are identical BK-Plots stratified by Z . The horizontal axis is $\text{pr}(U=1 | x, z)$ and the vertical axis is $\text{pr}(Y=1 | x, u)$. The parallel diagonal lines represent Model Y for $X=0$ and $X=1$. The lower diagonal line plots $\text{pr}(Y=1 | X=0, z)$ from β to $(\beta + \beta_U)$ as a function of $\text{pr}(U=1 | X=0, z)$. The upper diagonal line plots $\text{pr}(Y=1 | X=1, u)$ from $(\beta + \beta_X)$ to $(\beta + \beta_X + \beta_U)$ as a function of $\text{pr}(U=1 | X=1, z)$. The causal effect, β_X , is the vertical distance between these parallel diagonal lines (line $X=1$ minus line $X=0$). To construct $f_{00} = \beta(1 - \alpha) + (\beta + \beta_U)\alpha$, a vertical line is drawn at α , the intersection with $X=0$ is determined, and a horizontal line from that intersection specifies f_{00} . To construct $f_{10} = (\beta + \beta_X)(1 - \alpha - \alpha_X) + (\beta + \beta_X + \beta_U)(\alpha + \alpha_X)$, a vertical line is drawn at $(\alpha + \alpha_X)$, the intersection with $X=1$ is determined, and a horizontal line from that intersection specifies f_{10} .

The middle panels in Figures 3 and 4 graphically link Derivations I and II by computing ϕ and $(\phi + \phi_X)$ using equations (15) and (16) with $\alpha_{XZ}=0$. The computation of ϕ as a fraction γ of the distance between α and $(\alpha + \alpha_Z)$ is indicated by a dashed blue vertical line. The computation of $(\phi + \phi_X)$ as a fraction $(\gamma + \gamma_X)$ of the distance between $(\alpha + \alpha_X)$ and $(\alpha + \alpha_X + \alpha_Z)$ is indicated by the dashed red vertical line for $\gamma_X = 0$, and the dashed green vertical line for $\gamma_X > 0$ (Figure 3) or $\gamma_X < 0$ (Figure 4).

The bottom panels, which are based on Derivation II, are BK-Plots involving Z -adjusted and crude risk differences; they graphically compute $Bias_{crude}$ and $Bias_{Zadj}$ based on the results from the middle panel. To compute f_0 a blue vertical line is drawn at ϕ , the intersection with $X=0$ is determined, and a horizontal line from that intersection specifies f_0 . To compute f_1 a red (for $\gamma_X = 0$) or green (for $\gamma_X < 0$) vertical line is drawn at $(\phi + \phi_X)$, the intersection with $X=1$ is determined, and a horizontal line from that intersection specifies f_1 .

The causal effect β_X is the vertical difference between the diagonal lines reproduced from the top panels, namely the distance between the blue and dashed black horizontal lines, a positive quantity. RD_{crude} is the difference between f_1 and f_0 when $\gamma_X = 0$, namely the distance between red and blue horizontal lines, a negative quantity. RD_{Zadj} is the difference between f_1 and f_0 when $\gamma_X < 0$, namely the distance between green and blue horizontal lines, a negative quantity. Bias is indicated by the colored arrows. In Figure 3 the red downward arrow for $Bias_{Zadj}$ is larger than the green downward error for $Bias_{crude}$, indicating bias amplification. In Figure 4 the red downward arrow for $Bias_{Zadj}$ is smaller than the green downward error for $Bias_{crude}$, indicating bias attenuation.

The BK2-Plot aids intuition by explaining bias amplification and attenuation as relative shifts in vertical lines when adjusting versus not adjusting for a variable that is directly

related to treatment but not directly related to outcome. In practice, it is difficult to decide *a priori* the direction and extent of the shift of vertical lines in a particular problem.

5. The paired availability design: The PAD-Plot

The paired availability design [18, 19] does not fit into a causal graph framework but can be formulated using probability theory. A brief summary is given here along with an improved version of a graphical display [18]. Let Z denote time period and Y denote outcome. Let $T0$ and $T1$ denote treatments. Treatment availability changes from time period $Z=0$ to time period $Z=1$. Under various assumptions that can be made more plausible by design, the causal effect of time period is

$$\Delta_{overall} = \text{pr}(Y=1|Z=1) - \text{pr}(Y=1|Z=0). \quad (19)$$

The goal of the paired availability design is to estimate the causal effect of treatment $T1$ instead of $T0$. Achieving this goal involves the following principal stratification model [20] and two plausible assumptions. The principal strata, denoted $R = r$, are based on the treatment a participant would receive if arrival was (sometimes hypothetically) in either time period are

$R = n$, if would receive $T0$ regardless of time period,

$R = c$, if would receive $T0$ in time period $Z=0$ and $T1$ in time period $Z=1$,

$R = i$, if would receive $T1$ in time period $Z=1$ and $T0$ in time period $Z=0$,

$R = a$, if would treatment $T1$ regardless of time period.

Based on the definitions, the probability of receiving $T1$ in a given time period is a function of the principal strata: $\text{pr}(T1 | Z=0) = \text{pr}(R=i) + \text{pr}(R=n)$ and $\text{pr}(T1 | Z=1) = \text{pr}(R=c) + \text{pr}(R=a)$. Therefore the effect of time period on the probability of receiving $T1$ is

$$\Delta_{treated} = \text{pr}(T1|Z=1) - \text{pr}(T1|Z=0) = \text{pr}(R=c) - \text{pr}(R=i). \quad (20)$$

The overall treatment effect is

$$\Delta_{overall} = \Delta_{stratum(a)} \text{pr}(R=a) + \Delta_{stratum(c)} \text{pr}(R=c) - \Delta_{stratum(i)} \text{pr}(R=i) - \Delta_{stratum(n)} \text{pr}(R=n). \quad (21)$$

where $\text{stratum}(r) = \text{pr}(Y=1 | Z=1, r) - \text{pr}(Y=1 | Z=0, r)$. The following two assumptions are invoked for identifiability.

Assumption 1. The probability of outcome does not change over the time periods for $R = n, a$. Mathematically this assumption is $\text{pr}(Y=1 | Z=0, R=n, T0) = \text{pr}(Y=1 | Z=1, R=n, T0) \equiv \text{pr}(Y=1 | n, T0)$ and $\text{pr}(Y=1 | Z=0, R=a, T1) = \text{pr}(Y=1 | Z=1, R=a, T1) \equiv \text{pr}(Y=1 | a, T1)$.

Assumption 2. Under fixed availability (the increase in availability of treatment occurs at a fixed time), $\text{pr}(R=i) = 0$. Under random availability (the increase in availability of treatment occurs at random times), receipt of $T1$ or $T0$ occurs by chance among principal strata $R=c$ and $R=i$. Mathematically this assumption translates to $\text{pr}(Y=1|Z=0, R=c, T0) = \text{pr}(Y=1|Z=1,$

$R=I, T0) \equiv \text{pr}(Y=1|c, T0)$ and $\text{pr}(Y=1|Z=I, R=c, T1) = \text{pr}(Y=1|Z=0, R=I, T1) \equiv \text{pr}(Y=1|c, T1)$.

Assumption 1 implies $\text{stratum}(a) = \text{stratum}(n) = 0$. The addition of Assumption 2 implies overall $= \text{stratum}(c) \text{pr}(R=c)$ for fixed availability and $\text{stratum}(c) = -\text{stratum}(i)$ so overall $= \text{stratum}(c) \{\text{pr}(R=c) - \text{pr}(R=i)\}$ for random availability. This yields the well-known result

$$\Delta_{\text{stratum}(c)} = \{\text{pr}(Y=1|Z=I) - \text{pr}(Y=1|Z=0)\} / \{\text{pr}(T1|Z=1) - \text{pr}(T1|Z=0)\} \quad (21)$$

The PAD-Plots in Figures 5 and 6 graphically display the above equations, adding insight to the calculations.

Acknowledgments

This work was supported by the National Institutes of Health. The author thanks Jessica Myers, and the reviewers and associate editor for helpful comments.

Appendix A

For the back-door path in causal graph (c) in Figure 2, conditioning on U makes Y and X dependent on the back-door path ($X \leftarrow Q \rightarrow U \leftarrow R \rightarrow Y$), namely $\text{pr}(Y, X|u) = \text{pr}(Y|u) \text{pr}(X|u)$. The proof comes from comparing the following equations,

$$\text{pr}(Y, X|U) = \text{pr}(Y, X, U) / \text{pr}(U) = \sum_r \sum_q \text{pr}(Y|r) \text{pr}(X|q) \text{pr}(U|q, r) \text{pr}(R) \text{pr}(Q) / \text{pr}(U),$$

$$\text{pr}(Y|u) = \text{pr}(Y, U) / \text{pr}(U) = \sum_r \sum_q \text{pr}(Y|r) \text{pr}(U|q, r) \text{pr}(R) \text{pr}(Q) / \text{pr}(U),$$

$$\text{pr}(X|u) = \text{pr}(X, U) / \text{pr}(U) = \sum_r \sum_q \text{pr}(X|q) \text{pr}(U|q, r) \text{pr}(R) \text{pr}(Q) / \text{pr}(U).$$

Appendix B

The ratio of the absolute values of the biases from equations (17) and (18) is

$$\text{BiasRatio} = |\text{Bias}_{\text{Zadj}}| / |\text{Bias}_{\text{crude}}| = |A| / |A+B|,$$

where $A = \alpha_X + \gamma \alpha_{ZX}$ and $B = \gamma_X (\alpha_Z + \alpha_{ZX})$. Bias attenuation ($\text{BiasRatio} < 1$) requires $|A+B| > |A|$, which holds under the following scenarios: S1: $A > 0$ and $B > 0$, S2: $A < 0$ and $B < 0$, S3: $A < 0$, $B > 0$ and $(A+B > 0 \text{ so } A+B > -A)$, and thus $B > -2A$. S4: $A > 0$, $B < 0$, and $(A+B < 0 \text{ so } -A-B > A)$, and thus $B < -2A$. Therefore bias amplification ($\text{BiasRatio} > 1$) arises in remaining scenarios, S5: $A < 0$ and $B > 0$ and $B < -2A$, and S6: $A > 0$ and $B < 0$ and $B > -2A$. Because B involves only products of parameters between 0 and 1 while A is the sum of a parameter between 0 and 1 and the product of parameters between 0 and 1, $|B|$ will generally be less than $|A|$ so S3 and S4 will not occur. Without scenarios S3 and S4, the determination of whether bias amplification or bias attenuation occurs depends only on the signs of A and B .

References

1. Byar DP. Why data bases should not replace randomized clinical trials. *Biometrics*. 1980; 35:337–342. [PubMed: 7407321]

2. Pearl, J. Causality: Models, Reasoning, and Inference. 2nd ed.. New York, NY: Cambridge University Press; 2009.
3. Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003; 14:300–306. [PubMed: 12859030]
4. Shrier D. Letter to the editor. *Statistics in Medicine*. 2008; 27:2740–2741. [PubMed: 18069729]
5. Rubin DB. Author's reply (to Ian Shrier's Letter to the Editor). *Statistics in Medicine*. 2008; 27:2741–2742.
6. Shrier D. Letter to the editor: Propensity scores. *Statistics in Medicine*. 2009; 28:1317–1318. [PubMed: 19288531]
7. Sjolander A. Letter to the editor: Propensity scores and m-structures. *Statistics in Medicine*. 2009; 28:1416–1420. [PubMed: 19340845]
8. Pearl J. Letter to the editor: Remarks on the method of propensity scores. *Statistics in Medicine*. 2009; 28:1420–1423.
9. Rubin DB. Author's reply: Should observational studies be designed to allow lack of balance in covariate distributions across treatment groups? *Statistics in Medicine*. 2009; 28:1420–1423.
10. Liu W, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S. Implications of M bias in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2012; 176:938–948. [PubMed: 23100247]
11. Bhattacharya J, Vogt WB. Do Instrumental Variables Belong in Propensity Scores? *International Journal of Statistics and Economics*. 2012; 9:A12.
12. Pearl, J. Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010). Corvallis: Association for Uncertainty in Artificial Intelligence; 2010. On a class of bias-amplifying variables that endanger effect estimates; p. 425-432.
13. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Joffe MM, Glynn RJ. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *American Journal of Epidemiology*. 2011; 174:1213–1222. [PubMed: 22025356]
14. Pearl J. Invited commentary: understanding bias amplification. *American Journal of Epidemiology*. 2011; 174:1223–1227. [PubMed: 22034488]
15. Myers JA, Rassen JA, Gagne JJ, Huybrechts KF, Schneeweiss S, Rothman KJ, Glynn RJ, Myers, et al. Response to "Understanding Bias Amplification". *American Journal of Epidemiology*. 2011; 174:1228–1229.
16. VanderWeele TJ, Shpitser I. A new criterion for confounder selection. *Biometrics*. 2011; 67:1406–1413. [PubMed: 21627630]
17. Breslow, NE.; Day, NE. Statistical methods in cancer research. Lyon: International Agency for Research on Cancer; 1980. p. 104
18. Baker SG, Lindeman KS. Revisiting a discrepant result: a propensity score analysis, the paired availability design for historical controls, and a meta-analysis of randomized trials. *Journal of Causal Inference*. 2013 in press.
19. Baker SG, Lindeman KS. The paired availability design: a proposal for evaluating epidural analgesia during labor. *Statistics in Medicine*. 1994; 13:2269–2278. [PubMed: 7846425]
20. Frangakis CE, Rubin DB. Principle stratification in causal inference. *Biometrics*. 2002; 58:21–29. [PubMed: 11890317]
21. Gardner, M. Martin Gardner's Sixth Book of Mathematical Games from Scientific American. San Francisco: W.H. Freeman and Company; 1971. p. 154
22. Maor, E. Trigonometric Delights. Princeton: Princeton University Press; 1998. p. 122-123.
23. Soto AM, Sonnenschein C, Miquel PA. On physicalism and downward causation in developmental and cancer biology. *Acta Biotheor*. 2008 Dec; 56(4):257–274. [PubMed: 18542843]
24. Prehn RT. Cancers beget mutations versus mutations beget cancer. *Cancer Res*. 1994; 54:5296–5300. [PubMed: 7923156]
25. Baker SG. Paradoxes in carcinogenesis should spur new avenues of research: An historical perspective. *Disruptive Science and Technology*. 2012; 1:100–107.
26. Baker SG. Paradox-driven cancer research. *Disruptive Science and Technology*. 2013 In press.

27. Baker SG, Kramer BS. Good for women, good for men, bad for people: Simpson's paradox and the importance of sex-specific analysis in observational studies. *Journal of Women's Health & Gender-Based Medicine*. 2001; 10:867–872.
28. Hernán MA, Clayton D, Keiding N. The Simpson's paradox unravelled. *International Journal of Epidemiology*. 2011; 40:780–785. [PubMed: 21454324]
29. Arah O. The role of causal reasoning in understanding Simpson's paradox, Lord's paradox, and the suppression effect: Covariate selection in the analysis of observational studies. *Emerging Themes in Epidemiology*. 2008; 5:5. [PubMed: 18302750]
30. Jeon JW, Chung HY, Bae JS. Chances of Simpson's paradox. *Journal of the Korean Statistical Society*. 1987; 16:117–125.
31. Wainer H. The BK-Plot: Making Simpson's paradox clear to the masses. *Chance*. 2002; 15:60–62.
32. Baker SG, Freedman LS. A simple method for analyzing data from a randomized trial with a missing binary outcome. *BMC Medical Research Methodology*. 2003; 3:8. [PubMed: 12734019]
33. Baker SG, Kramer BS. The transitive fallacy for randomized trials: If A bests B and B bests C in separate trials, is A better than C? *BMC Medical Research Methodology*. 2002; 2:13. [PubMed: 12429069]
34. Baker SG, Kramer BS. Surrogate endpoint analysis: An exercise in extrapolation. *Journal of the National Cancer Institute*. 2013 In press.
35. Baker SG, Kramer BS. Randomized trials, generalizability, and meta-analysis: Graphical insights for binary outcomes. *BMC Medical Research Methodology*. 2003; 3:10. [PubMed: 12809566]
36. Gail MH, Wieand S, Piantadosie S. Biased estimates of treatment effect in randomized experiments with nonlinear regressions and omitted covariates. *Biometrika*. 1984; 71:431–444.
37. Greenland S, Robins JM, Pearl J. Confounding and collapsibility in causal inference. *Statistical Science*. 1999; 14:29–46.
38. Gail MH, Wacholder S, Lubin JH. Indirect corrections for confounding under multiplicative and additive risk models. *American Journal of Industrial Medicine*. 1988; 13:119–130. [PubMed: 3344751]

Figure 1.
Causal graphs where X is treatment, U is a confounder, and Y is outcome.

Figure 2.
BK-Plots involving risk difference, relative risk, and odds ratio.

Figure 3.
BK2-Plot showing bias amplification.

Figure 4.
BK2-Plot showing bias attenuation.

Fixed Availability

Time period $Z=0$

$$\begin{array}{|c|} \hline \text{pr}(T0|Z=0) \\ \hline \text{pr}(Y=1|n, T0) \text{pr}(R=n) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(T1|Z=0) \\ \hline \text{pr}(Y=1|a, T1) \text{pr}(R=a) \\ \hline \end{array}$$
Time period $Z=1$

$$\begin{array}{|c|} \hline \text{pr}(T0|Z=0) \\ \hline \text{pr}(Y=1|n, T0) \text{pr}(R=n) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(T1|Z=1) \\ \hline \text{pr}(Y=1|c, T1) \text{pr}(R=c) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(Y=1|a, T1) \text{pr}(R=a) \\ \hline \end{array}$$
Difference in $\text{pr}(Y=1|Z=z, R=c) =$ Difference in $\text{pr}(Y=1|Z=z)$

$$\text{pr}(Y=1|c, T0) \text{pr}(R=c)$$

$$\text{pr}(Y=1|c, T1) \text{pr}(R=c)$$
Divided by difference in $\text{pr}(T1|Z=z)$

$$\text{pr}(R=c)$$

Random Availability

Time period $Z=0$

$$\begin{array}{|c|} \hline \text{pr}(T0|Z=0) \\ \hline \text{pr}(Y=1|n, T0) \text{pr}(R=n) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(Y=1|c, T0) \text{pr}(R=i) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(T1|Z=0) \\ \hline \text{pr}(Y=1|c, T1) \text{pr}(R=c) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(Y=1|a, T1) \text{pr}(R=a) \\ \hline \end{array}$$
Time period $Z=1$

$$\begin{array}{|c|} \hline \text{pr}(T0|Z=1) \\ \hline \text{pr}(Y=1|n, T0) \text{pr}(R=n) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(Y=1|c, T0) \text{pr}(R=c) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(T1|Z=1) \\ \hline \text{pr}(Y=1|c, T1) \text{pr}(R=i) \\ \hline \end{array}$$

$$\begin{array}{|c|} \hline \text{pr}(Y=1|a, T1) \text{pr}(R=a) \\ \hline \end{array}$$
Difference in $\text{pr}(Y=1|Z=z, R=c) =$ Difference in $\text{pr}(Y=1|Z=z)$

$$\text{pr}(Y=1|c, T1) \text{pr}(R=c)$$

$$\text{pr}(Y=1|c, T0) \text{pr}(R=c)$$

$$\text{pr}(Y=1|c, T0) \text{pr}(R=i)$$

$$\text{pr}(Y=1|c, T1) \text{pr}(R=i)$$
Divided by difference in $\text{pr}(T1|Z=z)$

$$\text{pr}(R=c)$$

$$\text{pr}(R=i)$$
Figure 5.

PAD-Plot for paired availability design. The size of each box represents $\text{pr}(R=r)$. The size of the shaded area in the box represents $\text{pr}(y|r, \text{treatment})$.

Table 1

Hypothetical data illustrating bias when U is a confounder. The conditional RD is 0.10. The crude RD is -0.20 .

treatment	confounder									
	$U=0$					$U=1$				
	outcome					outcome				
	$Y=0$	$Y=1$	$\text{pr}(Y=1)$	$Y=0$	$Y=1$	$\text{pr}(Y=1)$	$Y=0$	$Y=1$	$\text{pr}(Y=1)$	$U=?$
$X=0$	80	20	0.20	60	240	0.80	140	260	0.65	
$X=1$	210	90	0.30	10	90	0.90	220	180	0.45	
RD			0.10			0.10			-0.20	

Table 2

Hypothetical data illustrating bias amplification. The conditional *RD* is 0.10. The crude *RD* is -0.068, which translates into a bias of magnitude 0.168. The Z-adjusted *RD* is -0.11, which translates into a larger bias of magnitude 0.21.

clinic	treatment	confounder									
		<i>U</i> =0			<i>U</i> =1				<i>U</i> =?		
		outcome			outcome				outcome		
		<i>Y</i> =0	<i>Y</i> =1	pr(<i>Y</i> =1)	<i>Y</i> =0	<i>Y</i> =1	pr(<i>Y</i> =1)	<i>Y</i> =0	<i>Y</i> =1	pr(<i>Y</i> =1)	
<i>Z</i> =0	<i>X</i> =0	240	60	0.20	20	180	0.90	260	240	0.48	
	<i>X</i> =1	239	103	0.30	0	38	1.00	239	141	0.37	
	<i>RD</i>			0.10			0.10			-0.11	
<i>Z</i> =1	<i>X</i> =0	40	10	0.20	45	405	0.90	85	415	0.83	
	<i>X</i> =1	174	74	0.30	0	372	1.00	174	446	0.72	
	<i>RD</i>			0.10			0.10			-0.11	
<i>Z</i> =?	<i>X</i> =0							345	655	0.655	
	<i>X</i> =1							413	587	0.587	
	<i>RD</i>									-0.068	

Table 3

Hypothetical data illustrating bias amplification. The conditional RD, which is the causal effect, is 0.10. The crude RD is -0.25, which translates into a bias of magnitude 0.35. The Z-adjusted RD is -0.11, which translates into a smaller bias of magnitude 0.21.

clinic	treatment	confounder							
		U=0			U=1			U=?	
		outcome			outcome			outcome	
		Y=0	Y=1	pr(Y=1)	Y=0	Y=1	pr(Y=1)	Y=0	Y=1
Z=0	X=0	240	60	0.20	20	180	0.90	260	240
	X=1	567	243	0.30	0	90	1.00	567	333
	RD			0.10			0.10		-0.11
Z=1	X=0	40	10	0.20	45	405	90	85	415
	X=1	28	12	0.30	0	60	1.00	28	72
	RD			0.10			0.10		-0.11
Z=?	X=0							345	655
	X=1							595	405
	RD								-0.25