Section Editor: **Franklin Dexter**

STATISTICAL GRAND ROUNDS

Understanding the Mechanism: Mediation Analysis in Randomized and Nonrandomized Studies

Edward J. Mascha, PhD,* Jarrod E. Dalton, PhD,*† Andrea Kurz, MD,† and Leif Saager, Dr med†

In comparative clinical studies, a common goal is to assess whether an exposure, or intervention, affects the outcome of interest. However, just as important is to understand the mechanism(s) for *how* the intervention affects outcome. For example, if preoperative anemia was shown to increase the risk of postoperative complications by 15%, it would be important to quantify how much of that effect was due to patients receiving intraoperative transfusions. Mediation analysis attempts to quantify how much, if any, of the effect of an intervention on outcome goes though prespecified mediator, or "mechanism" variable(s), that is, variables sitting on the causal pathway between exposure and outcome. Effects of an exposure on outcome can thus be divided into direct and indirect, or mediated, effects. Mediation is claimed when 2 conditions are true: the exposure affects the mediator and the mediator (adjusting for the exposure) affects the outcome. Understanding how an intervention affects outcome can validate or invalidate one's original hypothesis and also facilitate further research to modify the responsible factors, and thus improve patient outcome. We discuss the proper design and analysis of studies investigating mediation, including the importance of distinguishing mediator variables from confounding variables, the challenge of identifying potential mediators when the exposure is chronic versus acute, and the requirements for claiming mediation. Simple designs are considered, as well as those containing multiple mediators, multiple outcomes, and mixed data types. Methods are illustrated with data collected by the National Surgical Quality Improvement Project (NSQIP) and utilized in a companion paper which assessed the effects of preoperative anemic status on postoperative outcomes. (Anesth Analg 2013;117:980–94)

In clinical studies, the usual goal is to assess whether or not an intervention or exposure affects the outcome of interest. However, probing further to understand the mechanism(s) for *how* an intervention affects outcome is a vital and underpursued element of clinical research—for both randomized and nonrandomized studies. In this paper, we discuss *mediation analysis* which attempts to sort out whether and how much of the effect of an intervention goes though prespecified intermediary or *mediator* variables.

In our companion paper, Saager el al.[1] compared 119,298 patients with anemia, defined as hematocrit <36% for women and <39% for men, to the same number of propensity-matched (i.e., confounder-adjusted) nonanemia patients on a set of 9 major complications. For the main analysis, they assessed the overall or *total* effect of anemia on each outcome. Such is the standard analysis in most research studies. However, this total effect can be divided into the *direct* effect of anemia and *indirect or mediated* effects. Indirect effects are those which occur by first affecting a *mediating variable* or *mediator* which in turn causes increased (or decreased) risk of the outcome. For example, Saager et al.[1] presupposed that anemia might lead to increased wound contamination risk and thereby increased risk of mortality. The *direct* effect of anemia was estimated as the association between anemia and outcome after adjusting for the potential mediators. Since these direct effects, or effects of anemia "per se", were much smaller than the total effects, the authors concluded that the mediator variables were responsible for at least some of the total effect of anemia on outcomes. In this paper, we discuss mediation effects in more detail and demonstrate how to estimate them.

Mediation differs from confounding in the direction of causality: while mediators lie on the *causal pathway* between treatment and outcome, confounders influence both the exposure of interest and the outcome (Fig. 1). A mediating variable thus occurs temporally *after* the exposure—it is both caused by the exposure variable and is a cause of the outcome.[2,3] However, a confounding variable, by definition, temporally occurs *before* the exposure, such as past medical history or demographic data available before a surgical exposure.

When adjusting for confounding, typically in nonrandomized studies,[4–7] care must be taken to *not* include mediator variables. To the extent that the effect of an exposure on outcome goes though mediator variables, adjusting for those variables along with true confounding variables would tend to wash away the effect of interest. For example, in a study by Turan et al,[8] several variables suspected of at least partially mediating the effects of smoking on outcome, such as

From the Departments of *Quantitative Health Sciences and †Outcomes Research, Cleveland Clinic, Cleveland, Ohio.

Accepted for publication June 26, 2013.

Funding: Departmental funds.

The authors declare no conflicts of interest.

Reprints will not be available from the authors.

Address correspondence to Edward J. Mascha, PhD, Department of Quantitative Health Sciences, Cleveland Clinic, JJN-3 9500 Euclid Ave., Cleveland, OH 44195. Address e-mail to maschae@ccf.org.

DOI: 10.1213/ANE.0b013e3182a44cb9

congestive heart failure, were a priori identified and therefore were not adjusted for in the analysis of the total effect of smoking on outcome. Authors thus estimated the "overall" or "total" effect of smoking on outcome by only adjusting for the confounding variables. Biological knowledge and intuition of the proposed mechanism(s) for an exposure to affect an outcome are keys to distinguishing mediators from confounders.

Mediation can exist in both randomized and nonrandomized studies. While randomized studies typically do not include confounding variables, since with sufficient sample size baseline balance is achieved by the design, they do often involve mediators. Namely, there are often 1 or more mechanisms thought to be responsible for a hypothesized treatment effect on outcome. For example, authors in the DeLiT randomized trial[9] hypothesized that Dexamethasone administration would reduce the incidence of major complications by first reducing surgical inflammation measured by cytokines. Mediation analysis can formally assess whether a hypothesized factor actually mediates the effect of treatment on outcome.

Mediation analysis is an emerging area in statistical theory and practice,[2,10–14] and is part of the broader area of causal inference which strives to understand causal relationships in a wide variety of research settings.[15–18] The study by Saager et al.[1]—which we refer to as the "companion paper" throughout—exemplifies several of the challenges in assessing mediation. For example, special consideration must be made for multiple mediator variables, different mediator data types (i.e., binary, ordinal, and continuous), and multiple outcome variables. Binary outcomes are more challenging than continuous outcomes in mediation analysis, as we will discuss. The fact that anemia is a chronic exposure as opposed to an acute intervention adds additional challenges. Finally, making causal inference in mediation analysis typically requires making strong assumptions and having solid biological reasoning or evidence to back up the findings.

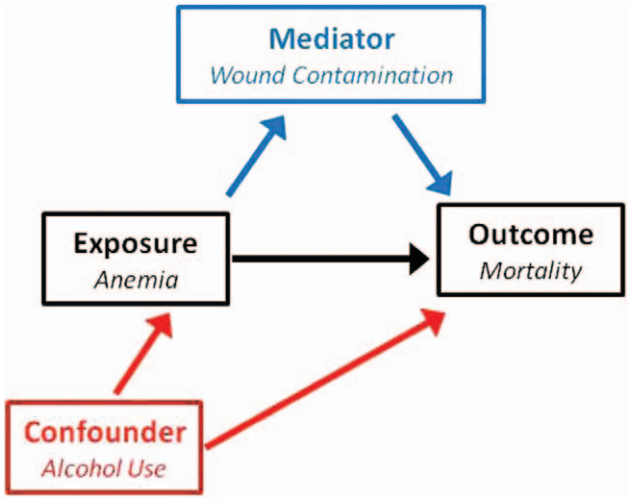

Figure 1. Mediation versus confounding. A mediator falls on the causal pathway between exposure and outcome. Wound contamination is a mediator of the effect of anemia on mortality, with BLUE arrows indicating the causal pathway exposure → mediator → outcome. However, alcohol use is a confounder of the relationship between anemia and mortality because it occurs before the exposure and influences both the exposure and the outcome, with RED arrows pointing to each.

We discuss the proper design and analysis of studies investigating mediation, using the companion paper as our primary motivating example. The remainder of this article proceeds as follows: Designing a mediation study; Effects of interest: Implementing a mediation analysis; Requirements for claiming mediation; Key assumptions in mediation analysis; Extension 1: Binary outcome with ordinal mediator; Extension 2: Binary outcome with multiple binary mediators; Sample size considerations; Discussion.

Throughout this paper, we largely discuss existing methods for designing and conducting mediation analysis. Since we do not propose new methods aside from a proposed extension to multiple mediators when the outcome is binary (see section **Extension 2**), we also do not provide proofs or analyses demonstrating the statistical properties of the mediation estimators that we discuss. The interested reader is encouraged to explore the provided references which give more details.

DESIGNING A MEDIATION STUDY

A mediation analysis can be either the primary or secondary aim of a research study. In either case, the proposed mediators should be carefully thought out and decided on before the study begins, thus requiring substantial clinical input. Biological justification and evidence as to why a variable might be a mediator of the relationship between exposure and outcome is crucial to being able to claim mediation. As we shall see throughout this paper, given the various assumptions that one must make, a statistical analysis alone is generally not enough to claim mediation.

Causal Diagramming

By definition, a proposed mediator should at least potentially lie on the causal pathway between exposure and outcome, and thus be able to mediate some of the effect of exposure on outcome (Fig. 1, top triangle). As a result, a helpful step in designing a mediation study is the development of a causal diagram, also called a directed acyclic graph,[18] which maps out the hypothesized directions of causality among the exposure of interest, potential confounders, and potential mediators. Lying on the causal pathway means that the exposure causes (or influences) the mediator, and that the mediator causes (or influences) the outcome, both at least to some degree. Therefore, all arrows on a causal pathway are in the same direction, as opposed to confounding (Fig. 1, bottom triangle) in which arrows emanate from the confounder to both exposure and outcome. Timing and known or suspected mechanism are thus key to identifying plausible mediators. In the companion paper, for example, the authors hypothesized that intraoperative wound contamination might at least to some degree and/or in some patients be the result of an anemic condition, and might also lead to wound infection as a complication, thus mediating the effect of anemia on outcome. More obviously, RBC transfusion might be caused at least in part by anemia, and might also lead to cardiovascular complications.

Acute Versus Chronic Exposures

In mediation analysis, it is helpful to distinguish exposures or interventions which are "acute" (e.g., type of anesthetic, any preoperative or intraoperative treatment) from those which are a chronic condition, habit or disease (e.g., anemia, diabetes, smoking). When the exposure is a perioperative intervention, it is easier to accurately identify suspected mediators of the treatment effect on outcome, since a true mediator would need to occur in an identifiable time interval between the intervention and the measured outcome. Acute interventions can occur in either randomized or nonrandomized studies and are also more likely to be manipulable such as in a randomized trial, making it easier to claim causal mediation.

Alternatively, when the exposure is a chronic condition, it is more difficult to accurately identify potential mediators, largely because of timing issues. We may not know exactly when such an exposure began, or when the effect of the proposed mediators began. In the companion paper, some of the investigated mediators clearly would have occurred temporally after a patient's baseline anemia diagnosis, such as intraoperative wound contamination and intraoperative red blood cell transfusion. However, for example, a patient might have acquired anemia after the occurrence of another specified mediator, such as reduced physical functioning. To the extent that the proposed mediators occur *before* an exposure began, they could not be mediators of the treatment effect on outcome. In some cases, such variables might be considered confounders. In reality a variable may act as a mediator for some patients and as a confounder for others; a mediation analysis as we describe herein should only be undertaken when it is reasonable to assume that the considered variable is a potential mediator (i.e., occurs after the exposure) for the vast majority of patients.

Another example is a study of Turan et al.[8] in which history of myocardial infarction (MI) was named a potential mediator of the effect of chronic smoking on postoperative outcomes, since smoking may cause heart disease as expressed by MI. But because details on the timing of smoking history and history of MI were not recorded in our database, we needed to assume that for the vast majority of patients MI occurred after smoking had begun.

Manipulable Intervention and Mediators

Since mediation analysis attempts to establish the *effect* of exposure on a mediator and the *effect* of a mediator on outcome, it is natural to think of mediation in a causal context rather than just association. However, claims of truly causal mediation depend on both exposures and mediators being at least theoretically manipulable, that is, modifiable, as is required in a randomized controlled trial aimed at assessing causal effects of an exposure on outcome.[19] For example, whether a patient receives an intraoperative transfusion, a proposed mediator in the companion paper, is under the control of the provider, so is clearly manipulable. Presurgical anemia might also be considered manipulable, since in many patients the condition could be corrected or improved by presurgical transfusion(s), although the contrary, inducing an anemia state, would not be done. A proposed mediator variable such as dyspnea is less likely to be considered manipulable, but this may depend on the context.

The requirement of plausible manipulation is based on the "potential outcomes" framework of causal inference in which each individual has potential responses under various levels of manipulable exposures (or similarly, mediators), with the understanding that only one of those responses is observable at any 1 time. The causal effect of interest for an individual, then, is the unobservable difference between his/her potential responses under the different levels.

In summary, cause–effect relationships in mediation analysis are made with caution, especially if assignment to either the exposure or the mediator variable(s) is nonrandomized.

EFFECTS OF INTEREST: IMPLEMENTING A MEDIATION ANALYSIS

Overview

In this main section of the paper, we describe the effects of interest for mediation analysis in detail. In doing so, we illustrate mediation analysis using data from the companion paper assessing the extent to which presurgical functional status mediates the effect of presurgical anemia on hospital length of stay (HLOS). Functional status is an ordinal variable with levels 1 (best), 2, and 3 (worst). As in the companion paper, anemia and nonanemia patients were first propensity matched on available presurgical confounding variables.

In comparative studies involving mediation, there are 3 main treatment effects of interest, as displayed in Fig. 2. The ***total effect*** is the effect of exposure on outcome ignoring any mediator(s), for example, the effect of anemia on length of stay. This is the usual treatment effect reported in studies which do not consider mediation, and is displayed in the top diagram—effect **c**. Second is the ***direct effect*** of exposure on outcome, or the effect "per se", after accounting for the mediator(s), illustrated as effect ***c′*** ("c-primed") in the lower diagram. Finally, the ***indirect or mediation effect*** of exposure on outcome is the effect of exposure on outcome through a designated mediator, as in the lower diagram (effect **a** × **b),** for example, the effect of anemia on hospital length of stay through functional status. Below we explain the effects in more detail and how to estimate them.

To estimate the effects of interest, 3 distinct regression equations are traditionally fit, as displayed in equations (1), (2), and (3) and referenced in Fig. 2. In these models, Y is the study outcome of interest; X is the intervention or exposure; M is the mediator; and E(Y) and E(M) are the mean or expected value of Y and M, respectively, given the variables on the right side of the relevant equation.

$$\mathrm{E}(Y) = \text{intercept} + cX + \text{covariates (as needed)} \quad (1)$$

$$\mathrm{E}(Y) = \text{intercept} + c'X + bM + \text{covariates (as needed)} \quad (2)$$

$$\mathrm{E}(M) = \text{intercept} + aX + \text{covariates (as needed)} \quad (3)$$

Models (1) to (3) use linear regression when the dependent variable (Y or M) is continuous, and logistic regression when

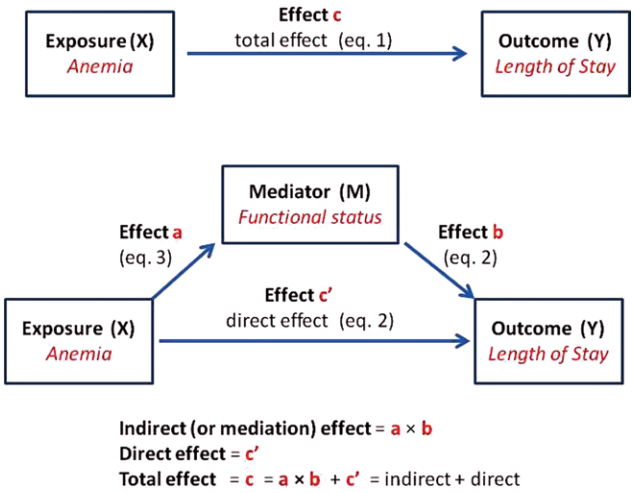

Figure 2. Mediation effects of interest. In the top diagram ***"c"*** is the total effect of exposure (*X*) on outcome (*Y*) ignoring the mediator (M). This is the usual treatment effect reported in clinical studies. In the **bottom diagram** mediation of the effect of anemia on length of stay through functional status is shown, where ***"a"*** is the effect of exposure on mediator, and "***b***" is the effect of mediator on outcome. When both effects "***a***" and "***b***" are significant, we claim mediation. The mediation effect is estimated as the product ***a*** times ***b***, or equivalently (for continuous outcome, mediator), as ***c*** (total effect) minus ***c***′. Effect ***c***′ ("**c-primed**") is the direct effect of exposure on outcome while adjusting for mediator. Eq.1, eq.2, and eq.3 refer to Equations (1)–(3) in the text.Note: Although we have added the clinical application, definitions and other details, much of the structure and notation displayed in Figure 2 were borrowed from the website of David Kenny at http://davidakenny.net/cm/mediate.htm, with permission.

binary. (Technical note: When logistic regression is used [see sections **Extension 1** and **Extension 2**], the left side of the relevant equation is a function of E(*Y*) or E(*M*), that is, log ($p/(1-p)$), where p is the probability that $Y = 1$ or $M = 1$ given *X* and all covariates.)

We assume throughout that a set of confounding variables is included on the right side (i.e., as independent variables) for each model, as needed. This is particularly important in observational studies if matching on confounding variables was either not done or has resulted in exposure groups still imbalanced on some potential confounding variables. However, in model (2), the adjustment for confounding is needed for all studies (even randomized), as discussed below.

Total Effect

The *total* effect of an exposure (e.g., anemia) on outcome (e.g., length of stay) is usually the primary interest in either a randomized or nonrandomized comparative study. It is the effect of exposure (*X*) on outcome (*Y*) while ignoring (i.e., not adjusting for) a prespecified mediator variable. The total effect is estimated by coefficient ***c*** in model (1), as seen in the top diagram in Figure 2. A total effect estimated as 1.5, for example, would indicate that the mean length of stay is 1.5 units higher for patients exposed to $X = 1$ (anemic) compared with $X = 0$ (nonanemic), and that some of that effect may be due to the ignored mediator, functional status.

For the companion paper, we defined *Y* as log-transformed HLOS, since the actual length of stay was not normally distributed (whereas the natural log-transformed data was close enough for our demonstration purposes). Total effect ***c*** was estimated as 0.21, implying that the mean (log-transformed) HLOS was estimated to be 0.21 longer (95% CI, 0.19–0.22) for patients with preoperative anemia compared with those without anemia, as in equation (1)*. By exponentiating the natural log-scaled results, we can also express ***c*** as the ratio of geometric means (95% CI) of anemia versus nonanemia patients, or 1.246 (1.237–1.254).

The *total* effect can be tested against 0 (H0: ***c***=0) using a *t* test, with the SE of ***c*** given in a typical regression output. In our current data example, it was not necessary to adjust for baseline covariates because the anemia groups were propensity matched on the available factors.

$$E(Y) = \text{intercept} + cX + \text{covariates} \quad (1)$$

$$\text{Mean log (HLOS)} = \text{intercept} + \mathbf{0.21}\ \text{anemic status (1 or 0)} \quad (1)^*$$

As explained below in the section Requirements for Claiming Mediation, a significant total effect is *not* required to claim mediation. For continuous outcomes, the total effect is the sum of the direct and indirect effects, which are explained next.

Direct Effect

The *direct* effect is the effect of exposure (*X*) on outcome (*Y*) independent of the effects of any mediating variables, and may be interpreted as the effect "per se" of the exposure on outcome. Thus, in model (2), the coefficient ***c***′ represents the direct effect of exposure on outcome after adjusting for mediator *M*. This is the direct connection between Exposure and Outcome in Figure 2, bottom diagram. For example, a ***c***′ of 0.25 would imply that *X* (anemia) is responsible for a difference of 0.25 in mean of *Y* (log-transformed HLOS), or an increase of 28% (i.e., $100 \times [1 - e^{0.25}]$) in actual mean length of stay, after removing the effect of *M* (functional status) on *Y*, the latter estimated by coefficient ***b*** in model (2). The direct effect can be tested using a simple *t* test, as with the total effect.

Again using the companion paper data, we fit equation (2) with results in (2)* below. Even though the anemia groups are propensity matched, we needed to adjust for all baseline covariables in (2) in attempts to remove confounding from the relationship between *M* and *Y* (see also **Key Assumptions In Mediation Analysis**). The *direct* effect of anemia on outcome, or ***c***′, is estimated as 0.19, which is nearly as large as the estimated total effect of 0.21.

$$Y = \text{intercept} + c'X + \mathbf{b}M + \text{covariates} \quad (2)$$

$$\text{Mean log (HLOS)} = \text{intercept} + \mathbf{0.19}\ \text{anemic status} + \mathbf{0.93}\ \text{functional status} + \text{covariates} \quad (2)^*$$

To the extent that a mediator explains most or all of the effect of *X* on *Y*, the direct effect tends to approach 0. In the companion paper, for example, all of the direct effect odds ratios for the binary outcomes were smaller than the total effect odds ratios (i.e., closer to 1.0), suggesting some mediation through the proposed mediators.

Indirect or Mediation Effect

Our main interest is in the *mediation* effect(s), quantifying whether and how much of the effect of the exposure on outcome goes through, or is mediated by, certain prespecified intermediary or mediator variables. The goal is to elucidate the mechanism(s) as to *how* the exposure affects the outcome.

The indirect or "*mediation*" effect of *X* on *Y* through mediator *M* is estimated in one of 2 ways: the difference method or the product method. These methods produce equivalent results for continuous outcomes, but as will be explained later, this is not the case for binary outcomes.

In the **difference method**, we estimate the mediation effect as the difference between the total and direct effects (i.e., $c - c'$). Using estimates of total and direct effects of 0.211 and 0.190, respectively, from our companion paper, we estimate the mediation effect as $c - c'$, or 0.211 − 0.190 = 0.021. We conclude that functional status accounts for about 0.021 log-transformed days, or about 10% of the total effect of anemia on length of stay.

In the **product method**, the mediation effect is estimated as the product $a \times b$, where a measures the strength of the effect of *X* on mediator *M* in model (3), and b is the effect of the mediator on *Y* adjusting for exposure *X* and other covariates in model (2). In Figure 2 (bottom diagram) effects a and b comprise the pathway from exposure to mediator to outcome.

In our example, we estimate effect a = 0.023 by fitting model (3) using linear regression (we included no baseline covariables since the anemia exposure groups were already propensity score matched). Anemic patients had an average of 0.023 worse functional score (on a scale of 1–3, with 1 being best) than nonanemic patients (95% CI, 0.021–0.025). We conclude that the exposure affects the mediator since the 95% confidence interval does not overlap 0.[a]

$$E(M) = \text{intercept} + aX + \text{covariates} \quad (3)$$

$$\text{Mean functional status} = \text{intercept} + \mathbf{0.023}\ \text{anemic status} \quad (3)^*$$

We have already estimated effect b as 0.93 in equation (2)* above, that is, the effect of functional status (M) on log-transformed HLOS (Y) adjusting for anemia (X). A 1-unit increase in functional status increased log-transformed HLOS days an estimated mean of 0.93 (95% CI, 0.91–0.95), or by exponentiating, a ratio of geometric mean length of stay of 2.53 (2.48–2.59).

Since effects a and b are significant, there is some evidence/suggestion of mediation (see section **Requirements for Claiming Mediation)**. We estimate the *mediation* effect as the product $a \times b$, or 0.023 × 0.93 = 0.021, with 95% CI, (0.019–0.023), statistically significant since it does not include 0. See **Appendix 1** for details on calculation of the SE (Sobel method).[20] However, since a mediation effect is typically not normally distributed, it is generally preferable to use the nonparametric bootstrap resampling method to obtain confidence intervals, as we do for later examples.[3,21]

We interpret the estimated mediation effect of 0.021 as the expected change in the outcome for a change of amount a in the mediator, where amount a is the exposure's (anemia versus nonanemia) effect on the mediator. This is an example of weak mediation because (1) the mediation effect of 0.021 is small, and (2) the anemia exposure has minimal effect on the mediator—an effect of 0.023 is only 2.3% of the distance between functional status levels. A strong mediation effect would have strong effects of both the exposure on the mediator and the mediator on the outcome.

As expected with continuous outcomes, the mediation effect estimated by the difference and product methods was the same (0.021) (although this might not hold if different variables or techniques were used to adjust for confounding in equations [1]–[3]). For binary outcomes, this equality would not generally be the case—adjustments or alternative methods are required (see sections Extension 1: Binary Outcome with Single Mediator and Extension 2: Binary Outcome with Multiple Mediators).

Proportion Mediated

An attractive summary measure is the proportion of the total effect due to a specific mediator, estimated as the mediation effect divided by the total effect of an exposure on outcome, or $a \times b/c$. This is a useful measure for individual mediators. When there are multiple mediators the ratio is not always easy to interpret, since mediation effects can go in different directions (see section Extension 2: Binary Outcome with Multiple Mediators).

For our example, given the estimated mediation effect ($a \times b$ = 0.021) and total effect (c = 0.211), the estimated proportion of the total effect of anemia due to the mediator, functional status, is 0.021/0.21, or 0.10, about 10% (95% CI, 9%–11% by bootstrap resampling).

REQUIREMENTS FOR CLAIMING MEDIATION

To claim that the effect of an exposure on outcome is at least partially mediated by a particular variable *M*, the most recent literature[2,3,19,22] generally agrees on these 3 requirements:

1. **Exposure *X* must affect mediator *M*.** That is, coefficient a in equation (3) must be found significantly different from 0.
2. **Mediator *M* must affect outcome *Y*, independent of the exposure *X*.** That is, coefficient b in equation (2) must be found significantly different from 0.
3. **Mediation effect must be significant**. Although evidence of mediation can be claimed if both effects a (the exposure affects mediator, no. 1 above) and b (mediator affects outcome, no. 2 above) are significant, it is more convincing if the mediation effect itself, the product $a \times b$, is also significantly different from 0. We and others thus make this a third requirement for claiming mediation.

In our companion paper example, the exposure anemia was significantly related to the proposed mediator,

[a]In practice one would probably *not* use linear regression when modeling functional status score since it is ordinal – more appropriate would be a cumulative logit model, for example. Here we use linear regression to keep things simple for demonstration purposes.

functional status (requirement 1). In addition, functional status was related to HLOS after removing the direct effect of anemia on length of stay (requirement 2). Finally, the product of effects ***a*** and ***b*** was significant (requirement 3). Since *all 3* required criteria were significant, we claim that some (a small portion) of the effect of anemia on HLOS goes through functional status.

The more traditional approach to claiming mediation had 2 additional requirements, [Baron and Kenny 23] namely, that the total effect is significant and that it is substantially larger than the direct effect. In **Appendix 2,** we explain why these are no longer considered required.

KEY ASSUMPTIONS IN MEDIATION ANALYSIS

1. Causal pathway

As mentioned, we assume that the mediator lies on the causal pathway between the exposure and outcome, such that the exposure causes the mediator and the mediator causes the outcome, as X → M → Y (Fig. 2, bottom). Besides the ordering, it is required that none of the arrows are bidirectional, for example, we assume that the mediator affects the outcome but not vice versa. In our example, we assume that length of stay would not affect presurgical functional status, and functional status would not affect anemia status.

2. Manipulable Exposure and Mediator

As explained above, based on a potential outcomes understanding of causal inference, both exposure and mediator should be at least theoretically manipulable as a minimal condition to be able to claim causal mediation. For a particular individual, potential values of the mediator at each value of the exposure, and potential values of the outcome at each possible value of the mediator, must be plausible to discuss the causal effects of the exposure on the mediator or the mediator on the outcome.

Based on the above, for example, could one claim causal mediation in an analysis assessing whether cholesterol mediates the relationship between a particular genotype and risk of a heart attack? Typically, baseline conditions such as demographics would not be considered manipulable, although one might envision gene therapy or other sophisticated schemes to observe potential outcomes of the mediator or outcome from the same individual both with and without the exposure (e.g., the gene).

3. No Confounding

To claim causal mediation, we assume minimal or no confounding of the relationships between exposure and mediator(s), between exposure and outcome, or between mediator(s) and outcome.

We claim to have removed most of the potential confounding of the relationships between *exposure* and *mediator* (equation [3]) and between *exposure* and outcome (equation [2]) by using a randomized exposure. In nonrandomized studies, it is important to use either propensity-matched exposure groups (as in the companion paper example) or else to adjust for available preexposure confounding variables in equations (2) and (3). To the extent this has been done well, one considers discussing the "effect" of the exposure on the mediator and on the outcome, rather than just association.

However, additional steps must be taken to adjust for confounding of the *mediator–outcome* relationship in equation (2) to discuss the effect of *mediator(s)* on outcome. Exposure groups being either randomized or propensity matched does not remove confounding from the *mediator–outcome* relationship. In some studies, the mediator variable can be randomized. A more practical solution is to adjust for all available preexposure potentially confounding variables when fitting equation (2), as done above in our reanalysis of the companion paper data, even when exposure groups are propensity matched. It is *not* appropriate to adjust for postexposure covariables, since they might be affected by the exposure and so adjusting for them would bias results. (Specifically, the mediation effect can be expressed as the effect of the mediator M on outcome Y as a result of the effect of exposure X on M [effect ***a***]. Adjusting for a variable on the causal pathway between X and M when estimating the effect of M on Y [***b*** in {2}] could remove some of the effect of X on M, thus biasing the mediation effect estimated as $\boldsymbol{a} \times \boldsymbol{b}$.)[b]

4. No Interaction Between Exposure and Mediator on Outcome

We assume that the exposure and mediator have independent effects on the outcome.[2,24] However, this is easily assessed by including the appropriate exposure–mediator interaction term(s) into model (2).

An exposure–mediator interaction means that the effect of exposure on outcome depends on the observed level of the proposed mediator, and likewise the effect of mediator on outcome varies by level of the exposure. To the extent that there is an exposure–mediator interaction, and particularly when the effects change direction (not just amount) for different levels of the other variable, the overall estimated mediation effects would not be very interpretable. This is not to say that mediation does not exist in such situations—it is just more difficult to identify and describe. The exposure–mediator interaction is a testable assumption and so should always be assessed.

In our companion paper example, the exposure–mediator interaction was found to be significant ($t = 15.3$, $P < 0.001$), implying that the effect of anemia on length of stay depends on functional status level, and the effect of functional status on outcome varies by anemia status. However, the interaction is one of degree and not of direction. For example, the anemia "direct effect" is positive for all levels of functional status (effect $\boldsymbol{c}'$ =0.20, 0.40, 0.43 for functional status 1, 2, and 3, respectively), implying that anemic patients have longer mean length of stay for any level of functional status, although less so for the best functioning patients. Likewise, the functional status effect on length of stay is positive for both anemia (***b*** = 0.58) and nonanemia (***b*** = 0.42) patients, although stronger for anemia patients.

[b]Note that since mediation is typically assessed using the product method of $\boldsymbol{a} \times \boldsymbol{b}$, it is particularly important to remove confounding from equations (2) and (3). This holds regardless of whether the outcome and mediator are continuous or binary variables.

5. Usual Model Assumptions

Finally, the usual model assumptions for linear and/or logistic regression apply. For example, we assume observations are independent and that the functional form of the model fits the data reasonably well (e.g., linear regression assumes a linear relationship). Missing data is more of an issue in mediation analysis than it usually is. Also, while some exposure–mediator correlation is required to claim mediation, too much exposure–mediator correlation can result in greatly increased SEs for the effect of the mediator on outcome (effect ***b*** in equation [2]), thus reducing power when assessing mediation.[3]

Given the above required assumptions, the reader can appreciate that it is quite difficult to confidently claim mediation, even in studies with a randomized intervention. However, careful design and analysis can lead to stronger suggestions of mediation (or lack thereof). The following sections **Extension 1** and **Extension 2** are critical for those planning or conducting mediation analyses with binary outcomes. However, these sections are not crucial to the basic concepts of mediation analysis, which have been covered above.

EXTENSION 1: BINARY OUTCOME WITH SINGLE ORDINAL MEDIATOR

We use data from the companion paper to assess whether baseline functional status (levels 1, 2, and 3, with 3 being worst) is a mediator of the effect of anemia on 30-day mortality, a binary outcome. With a binary (yes/no) outcome such as any of the 9 complication events in the companion paper, we cannot assess mediation using the simple total effect minus direct effect "difference" method (***c*** minus ***c′***) from equations (1)–(3). For a binary outcome, logistic regression is used for equations (1) and (2). Since the residual variance is fixed in logistic regression, the scale of the outcome variable is not the same across equations having different predictors, and therefore the equality between the difference method and the ***a*** × ***b*** "product" method which holds for linear regression no longer holds.[2] Methods for assessing mediation with a binary outcome, particularly how to assess and express the mediation effect, is an active area of statistical research—that is, there is still is no generally accepted method. We therefore present >1 method for assessing mediation with binary outcomes, with the qualification that the jury is still out on the best or most accurate method.

MacKinnon et al.[2] and others claim that if the outcome is binary and the mediator is continuous (or ordinal), choices are to first standardize the regression coefficients and use the ***c*** minus ***c′*** method, or else to approximate the mediation effect using the ***a*** × ***b*** "product" method, and then estimate the proportion mediated.[2] We demonstrate both methods and compare results for our data example. For both approaches, we estimate confidence intervals using bootstrap resampling[21] since it is more reliable than the traditional Sobel method, particularly for nonnormal outcomes.[2,25]

More recently, some researchers have concluded that the above-mentioned approaches are not sufficient to assess mediation effects (and particularly, to make causal inference on the mediation effects) when the outcome is binary. Imai et al.[26] (2010) showed that when using the standardization method with a binary outcome, the proportion of the total effect mediated by the mediator only accurately estimates the true proportion mediated when the direct effect is rather small relative to other effects. Thus, Imai at el.[26] (2010) and others [e.g., Albert and Nelson 12] have proposed estimating mediation effects using the "potential outcomes" framework. In this framework, mediation is defined as the expected difference in a subject's potential outcomes at levels of the mediator that would result under one value of the exposure versus the other (complicated by the fact that only 1 potential mediator value and 1 potential outcome value is observable for a subject). For a continuous outcome, assuming there is no interaction between exposure and mediator on the outcome, the potential outcome and traditional methods would give the same results. Further details on this class of methods is beyond the scope of this article, but some form of this methodology may well represent the future of mediation analysis for both continuous and noncontinuous outcomes.

We proceed with the goal of assessing whether any mediation is present in our example and also estimating the proportion of the total effect which is mediated by the chosen mediator. We use the product method (both standardized and nonstandardized), with the qualification that the estimates will only be approximate and should be interpreted with caution. We also apply the method of Imai et al.[26] (2010) as a sensitivity analysis.

Since we have a binary outcome, we use logistic regression for equations (1)^ and (2)^ below so that effects ***b***, ***c***, and ***c′*** are log-odds ratios, and ***p*** is the probability that the outcome of interest ***Y*** (here, 30-day mortality) is 1 (or yes) for a particular subject given their observed values of exposure ***X*** and any other variables on the right side of the equation (i.e., ***M*** and/or covariates).

$$\log(p/(1-p)) = \text{intercept} + \boldsymbol{c}\,X + \text{covariates} \quad (1)^{\wedge}$$

$$\log(p/(1-p)) = \text{intercept} + \mathbf{0.46}\,X + \text{covariates} \quad (1)^{\wedge *}$$

$$\log(p/(1-p)) = \text{intercept} + \boldsymbol{c'}X + \boldsymbol{b}\,M + \text{covariates} \quad (2)^{\wedge}$$

$$\log(p/(1-p)) = \text{intercept} + \mathbf{0.41}\ \text{anemia} + \mathbf{0.74}\ \textbf{functional status} + \text{covariates} \quad (2)^{\wedge *}$$

Using the traditional method, the total effect of anemia on mortality would be estimated as ***c*** = 0.46 (exponentiate to obtain odds ratio of 1.5 [95% CI, 1.4–1.7]). However, when the outcome is binary, the total effect cannot be reliably estimated from ***c*** in (1)^*, but rather (according to some authors) by the sum of the estimated direct and indirect effects, or else by first standardizing, both of which are done below. The direct effect of anemia on mortality is estimated as ***c′*** = 0.41 (exponentiate to obtain odds ratio of 1.5 [1.4–1.6]) using equation (2)^*. We further estimate the effect of anemia on the mediator as ***a*** = 0.023 (0.021–0.025), that is, the difference in mean functional status for anemic versus nonanemic, as in (3)* above. Finally, we estimate the effect of the mediator on mortality adjusting for exposure as ***b*** = 0.74 (exponentiate to obtain odds ratio 2.1 [1.9–2.3]) in (2)^*. We next

estimate the mediation (i.e., indirect) effects both with and without standardizing the effects estimated here.

Product Method—Unstandardized Effects

Since both effects ***a*** and ***b*** are significant in our example, there is some evidence of mediation. This may be a more reliable finding (given the above discussion) than the following specific estimates of the mediation effect itself. Using the ***a* × *b*** method, the mediation (or indirect) effect is estimated as ***a* × *b*** = 0.023 × 0.74 = 0.0167 (95% CI, 0.01–0.02 by bootstrap resampling), and is interpreted as the expected change in the log-odds of having the outcome for a change of 0.023 (i.e., the change due to the exposure) in the log-odds of having the mediator. The proportion of the total anemia effect explained by functional status is calculated as (***a* × *b***)/***TE***, where ***TE*** is the estimated total effect, calculated as the direct effect plus the mediation effect [as noted above, with a binary outcome, we do not rely on the estimated effect ***c*** to represent the total effect]. The estimated proportion of the total effect of anemia on mortality mediated by functional status is thus 0.0167/(0.41 + 0.0167) = 0.0167/0.4267 = 0.039, or about 4%.

Product Method—Standardized Effects

Using the standardization method detailed in **Appendix 3** to account for the fact that outcomes are not on the same scale across equations, standardized estimates of the total effect (***c***) and direct effect (***c′***) are 0.256 and 0.226, respectively, for a difference of 0.030 to estimate the mediation effect. This estimate of the mediation effect is expectedly different from the results of the unstandardized ***a* × *b*** method in the preceding paragraph (0.0167) since results are on different scales. However, as shown in **Appendix 3,** we also obtain standardized estimates of effects ***a*** and ***b***. Using the product method on those results, the standardized ***a* × *b*** mediation effect is estimated as 0.023 × 0.404 = 0.0092, which explains 4% (95% CI, 2.8–4.5) of the total effect of 0.235 (***a* × *b* + *c′*** = 0.0092 + 0.226 = 0.235), very similar to the nonstandardized ***a* × *b*** method above.

Potential Outcomes Method

Finally, we estimate the proportion of the total effect of anemia on mortality mediated by functional status using the potential outcomes method of Imai et al (applied using the "mediate" function in the R "mediation" package). This method begins with regression models (2)^ and (3) and then simulates potential outcomes under various scenarios of exposure and mediator values. The estimated potential outcomes are used to estimate the mediation effect overall and under both levels of exposure. Using this method, the percent of the total effect mediated by functional status (95% CI) was estimated as 8, (6–16), somewhat higher than what was found using the above product methods.

Finally, the interaction between the exposure (anemia status) and mediator (functional status) on outcome (mortality) was not significant (P = 0.11), fulfilling one of the requirements for claiming mediation. We conclude that there is a small percent of the anemia effect of mortality mediated by functional status.

EXTENSION 2: BINARY OUTCOME WITH MULTIPLE BINARY MEDIATORS

Anemia Study Example

In the companion paper, we assessed the association between presurgery anemic status (yes/no) and 9 primary outcomes, all of which were binary (yes/no) complication events. For each outcome, we considered 6 mediator variables, including 5 binary and 1 continuous (i.e., duration of surgery). Outcomes were analyzed individually, reporting the "total effect" and "direct effect" (i.e., adjusting for mediators) of anemia on each outcome. We concluded that there was evidence of mediation when a direct effect was much smaller than a total effect, but we did not quantitatively assess the specific mediation effects. Here we assess the mediation effects from the companion paper in more detail. For simplicity, we change duration of surgery to a binary outcome defined as "1" if above 120 minutes and "0" otherwise; we thus consider 6 binary mediators. For the sake of interest, we add a 10th outcome of "any major complication", defined as "1" if any of the 9 original outcomes were observed and "0" otherwise. (Note: Since the outcome Y is binary for this example, equations (1)^ and (2)^^ are fit using logistic regression [as done with a single binary outcome in **Extension 1**], not linear regression).

In the below analysis of the companion paper, we focus on estimating mediation effects separately for each mediator (for each outcome), as opposed to the overall proportion mediated or the total amount of mediation. Although with multiple mediators, the *overall proportion mediated* can be calculated as the sum of the individual mediation effects divided by the total effect; this is not a consistently reliable summary measure because different mediation effects can tend to cancel each other out if in different directions, that is, some positive and some negative. Even if all effects are in the same direction, the overall proportion mediated may still not be a sufficient summary measure because one would typically want to know which variables are responsible for the most mediation.

Multiple Mediator Analysis

Here we expand the methods discussed in section **Extension 1** to the case of multiple mediators. All of the qualifications expressed in that section apply here as well. When there are multiple mediators (say ***k***), the equations used to assess mediation using the traditional approach are modified as follows. Equation (2)^ above (for a single mediator) is expanded to include a term for each mediator, as in (2)^^ below, so that with ***k*** potential mediators (here, 6) there are now ***k*** estimates of ***b*** (1 for each mediator)—that is, the effect of mediator on outcome, adjusting for all other mediators in addition to the exposure ***X***. A separate equation (3) is fit for the effect of exposure on each mediator, as below, giving ***k*** estimates of parameter ***a*** (1 for each mediator). In equations (1)^ and (2)^^ below, ***p*** is the probability for a particular subject that $Y = 1$ given X, M, and covariates. We fit equations (1)^ (unchanged from above) and (2)^^ for each of the outcomes of interest.

Table 1. Specific Mediation Effects on the Mortality Outcome—Companion Paper Data

Mediator	*Effect a* exposure on mediator[a]	*Effect b* mediator on outcome[a]	*a* × *b* mediation effect[a]	Ratio (× 100) of specific mediation effect to total effect[b]
1. Open wound	0.66 (0.61 to 0.70)	0.07 (−0.11 to 0.25)	0.05 (−0.07 to 0.16)	2 (−13 to 20)
2. RBC transfusion	0.89 (0.85 to 0.93)	0.94 (0.81 to 1.05)	0.82 (0.71 to 0.94)	54 (40 to 61)
3. Dyspnea	0.11 (0.09 to 0.13)	0.54 (0.43 to 0.64)	0.06 (0.04 to 0.08)	5 (2 to 8)
4. Wound contamination	−0.07 (−0.08 to −0.05)	−0.49 (−0.59 to −0.39)	0.03 (0.02 to 0.04)	3 (2 to 4)
5. Independent functional status	−0.50 (−0.55 to −0.46)	−0.77 (−0.90 to −0.64)	0.38 (0.31 to 0.45)	26 (19 to 36)
6. Duration of surgery > 2 h	−0.04 (−0.06 to −0.03)	−0.27 (−0.41 to −0.13)	0.01 (0.00 to 0.02)	0 (0 to 1)

Confidence intervals estimated using bootstrap resampling with 500 resamples

[a]Estimated effects are on the log-odds ratio scale.

[b]Ratio calculated as 100 × ([specific mediation effect a × b]/total effect), where total effect is sum of all mediation effects (i.e., sum of a × b column) and the direct effect $\boldsymbol{c}'$ estimated from equation (2)^^.

$$\log(p/(1-p)) = \text{intercept} + cX + \text{covs [as above]} \quad (1)^{\wedge}$$

$$\log(p/(1-p)) = \text{intercept} + c'X + b_1 M_1 + b_2 M_2 + b_3 M_3 + + + b_k M_k + \text{covs} \quad (2)^{\wedge\wedge}$$

$$E(M_1) = \text{intercept } 1 + a_1 X + \text{covs} \quad (3.1)$$

$$E(M_2) = \text{intercept } 2 + a_2 X + \text{covs} \quad (3.2)$$

$$E(M_k) = \text{intercept k} + a_k X + \text{covs} \quad (3.k)$$

To estimate the ratio of each specific mediation effect to the total effect of anemia on an outcome of interest, we generalized the product method described in ***Extension 1*** to the case of multiple mediators. This ratio is not a true proportion because it can easily be negative (due to either negative specific mediation effects or negative sum of mediation effects) or greater than 1. Therefore, we refer to it as a ratio or a percent. Analogous methods have been used by others for continuous outcomes. First, for a given outcome variable (say 30-day mortality), the mediation effect for a particular mediator (say mediator 1) was estimated as the product of ***a1*** from equation (3.1) and ***b1*** from equation (2)^^, or ***a1*** × ***b1***. This was done for all 6 proposed mediators for the given outcome, resulting in 6 mediation effects. Since all outcomes and mediators are binary, the specific mediation effects are the product of 2 log-odds ratios, that is, the effect of exposure on mediator (***a***) and effect of the mediator on outcome (***b***). These were summed across mediators to estimate the total mediation effect, or total indirect effect, as $\sum_{i=1}^{6} a_i b_i$.[c] We then estimated the total effect of anemia on 30-day mortality (which we refer to as ***TE***) as the sum of the total indirect effect and the direct effect estimated from $\boldsymbol{c}'$ in equation (2)^^. Finally, we estimate the ratio of specific mediation effects to the total effect of anemia on the outcome as the specific mediation effect divided by ***TE***. As in section **Extension 1**, confidence intervals were obtained via bootstrap resampling.

We also used this product method after standardizing the estimates ***a***, ***b***, and $\boldsymbol{c}'$ in a method analogous to that shown in **Appendix 3** for a single mediator. We first divided coefficients ***a*** and ***b*** by the SD of the predicted outcome for the relevant equation and then used the product method as above. Standardized and nonstandardized results on the proportion of the total effect due to each proposed mediator were very similar (typically within 1–2 absolute percent of each other, data not shown).

Example. For the 30-day mortality outcome, Table 1 gives estimates of effects ***a*** and ***b*** and their product (i.e., the mediation effect) for each of the 6 proposed mediators. For the open wound mediator, for example, effects ***a*** and ***b*** were estimated as log-odds ratios of 0.66 (95% CI, 0.61–0.70) and 0.07 (−0.11 to 0.25), respectively, for a mediation effect (95% CI) of ***a*** × ***b*** = 0.05 (−0.07 to 0.16), nonsignificant because it contains 0. The sum of the mediation effects across mediators gave an estimate of $\sum_{i=1}^{6} a_i b_i = 1.32$. The estimate of $\boldsymbol{c}'$ (direct effect) from equation (2)^^ was a log-odds ratio of 0.28. The total effect ***TE*** was thus estimated as 1.32 (indirect effect) + 0.28 (direct effect) = 1.60. Finally, the specific/total ratio (or percent mediated) for each mediator (and 95% CI) was estimated as mediation effect divided by ***TE*** × 100 (last column of Table 1). For open wound, it was 2% (−13% to 20%). All mediation effects in Table 1 except for the open wound mediator (row 1) are significant at the 0.05 level since the 95% confidence intervals do not contain 0.

As explained in *Requirements for claiming mediation*, mediation can be claimed if both effects ***a*** and ***b*** as well as the mediation effect ***a*** × ***b*** are significantly different from the null hypothesis value (0 on raw scale, 1 on odds ratio scale). For example, RBC transfusion appears to be a clear mediator of the anemia effect on mortality since the log-odds ratio (95% CI) for anemia status on RBC transfusion (effect ***a***) is 0.89 (0.85–0.93), the effect for RBC transfusion on outcome (effect ***b***) is 0.94 (0.81–1.05), and the mediation effect (***a*** × ***b***) is 0.82 (0.71–0.94).

Mediation effects can be either positive or negative, depending on the signs of the ***a*** and ***b*** effects. When both ***a*** and ***b*** are negative, the mediation effect becomes positive. For example, anemia was associated with lower odds of independent functional status (***a*** = log-odds ratio of −0.50), and independent functional status was associated with lower odds of mortality (***b*** = log-odds ratio of −0.77), for a positive mediation effect of ***a*** × ***b*** = 0.38, Table 1.

For the 30-day mortality outcome, we also applied the potential outcomes method of Imai[26] (2010) as described above. The ratio (or percent) mediated by each of the 6 mediators was at least in the same ranking order as those reported in Table 1 for the product method, and some of them were almost identical (results not shown).

[c]Another apparent option would be to define the total indirect effect as the sum of the *absolute* values of the specific indirect effects, with the advantage that indirect effects in opposite directions would not cancel each other out when summarizing the relative contribution of a specific mediator. However, using the raw indirect effects (as we have done) allows the total indirect effect to represent the *net* mediation effect across mediators.

Table 2. Total, Direct, and Mediation Effects for Each Outcome—Companion Paper Data

Outcome complication	Naive total effect (95% CI)	Total effect (95% CI)	Direct effect (95% CI)	Ratio of specific mediation effect to total effect (× 100) (95% CI)[a]					
				open wound	RBCs	Dyspnea	Wound contam	Independent functional status	Surgery hours >2
Systemic	0.23 (0.18 to 0.28)	1.39 (1.27 to 1.52)	0.10 (0.05 to 0.15)	24 (20 to 28)	40 (37,45)	2 (1 to 3)	5 (4 to 6)	22 (18 to 25)	0 (0 to 0)
Respiratory	0.26 (0.20 to 0.31)	1.30 (1.17 to 1.42)	0.10 (0.04 to 0.16)	5 (1 to 11)	55 (50 to 61)	4 (3 to 5)	3 (2 to 4)	24 (20,28)	1 (0 to 1)
Wound Infection	0.02 (−0.01 to 0.05)	0.69 (0.62 to 0.76)	−0.03 (−0.06 to 0.00)	50 (45 to 58)	25 (18 to 30)	1 (0 to 2)	9 (7 to 12)	20 (15 to 24)	1 (1 to 1)
Urinary	0.24 (0.19 to 0.30)	1.04 (0.90 to 1.15)	0.14 (0.08 to 0.20)	7 (2 to 14)	48 (42 to 54)	2 (1 to 2)	4 (3 to 5)	26 (22 to 31)	0 (0 to 0)
CNS	0.02 (−0.11 to 0.14)	0.18 (−0.16 to 0.43)	−0.03 (−0.16 to 0.10)	[b]	[b]	[b]	[b]	[b]	[b]
Thrombotic	0.22 (0.10 to 0.31)	0.90 (0.71 to 1.08)	0.14 (0.03 to 0.24)	10 (5 to 22)	46 (36 to 59)	2 (1 to 4)	4 (2 to 5)	22 (13 to 31)	0 (0 to 1)
Cardiovascular	0.26 (0.14 to 0.36)	0.95 (0.75 to 1.17)	0.12 (0.00 to 0.22)	20 (3 to 32)	47 (36 to 62)	3 (2 to 6)	3 (2 to 4)	13 (3 to 23)	0 (0 to 1)
Return to OR	0.18 (0.13 to 0.22)	1.16 (1.05 to 1.23)	0.11 (0.07 to 0.15)	40 (36 to 44)	35 (31 to 39)	1 (0 to 2)	0 (1 to 0)	15 (11 to 18)	0 (0 to 0)
Mortality	0.46 (0.40 to 0.54)	1.60 (1.39 to 1.79)	0.28 (0.20 to 0.36)	3 (6 to 10)	50 (44 to 58)	4 (2 to 5)	2 (1 to 3)	23 (18 to 28)	0 (0 to 1)
Any	0.16 (0.13 to 0.18)	1.21 (1.15 to 1.26)	0.09 (0.06 to 0.11)	29 (26 to 32)	39 (36 to 42)	1 (1 to 2)	3 (2 to 4)	21 (18 to 23)	0 (0 to 0)

Effects reported on raw scale, so the estimated total, direct, and mediation effects are log-odds ratios (i.e., log[OR].)

Naïve total effect: *effect c* in model (1)^. Not used for our mediation assessment (see text)

Total effect: total effect of exposure on outcome calculated as direct + indirect effects, that is, direct effect (c-prime) + sum (***a*** × ***b***). Specifically, effects ***a*** and ***b*** for the mediator of interest are estimated from the relevant equation (3.X) [e.g., (3.1) for mediator 1 (open wound) and from the relevant parameter estimate from equation (2)^^.

[a]Ratio calculated as 100 × ([specific mediation effect a × b]/total effect). All confidence intervals estimated using bootstrap resampling with 500 resamples

[b]CNS had low incidence (0.5%, *N* = 923), making mediation effects adjusted for all covariates in model Y = M + unstable; therefore, reported results do not adjust for confounders of the Y = M) relationships.

Table 2 reports the estimated total and direct effect of anemia on each outcome, as well as the estimated ratio (X 100) of specific mediation effect to total effect for each mediator. The first column gives the "naïve" total effect estimated from equation (1)^, reported for information purposes only. Column 2 gives the total effect (***TE***) estimated (as explained above) as the sum of the direct effect *c′* from equation (2)^^ and the total of the mediation (or indirect) effects using the product method (from equations [2]^^ and [3]).

For example, an estimated 40% (95% CI, 37–45) of the total effect of anemia on systemic complications was mediated by RBCs, 24% (20–28) by open wound, 22% (18–25) by functional status, and trivial amounts by dyspnea, wound contamination and duration of surgery (Table 2). Receiving any RBCs appeared to be a consistent mediator of the effect of anemia, explaining between 25% and 55% of the total effect of anemia on each of the 10 outcomes (see section ***Exposure–mediator interactions*** below for an important qualification). Open wound was found to explain at least 20% of the total anemia effect in 5 of the 10 outcomes, while functional status explained at least 20% in 7 of 10 outcomes. The other suspected mediators, dyspnea, wound contamination and duration of surgery, were not found to substantially contribute to the effect of anemia on outcome (all< 10%).

Interestingly, Table 2 shows that duration of surgery> 2 hours (last column) was not an important mediator of anemia for any of the outcomes of interest. This is because for the mortality outcome (see Table 1) and most others, although duration of surgery had a nontrivial association with outcome (effect ***b***), the effect of anemia on duration of surgery (effect ***a***) was almost 0. Therefore, the mediation effect, estimated as the product of ***a*** and ***b***, was very close to 0.

Exposure–Mediator Interactions

An important part of any mediation analysis is assessing the exposure–mediator interaction(s). In the traditional mediation analysis, this is assessed by adding an interaction term into equation (2) (or its variants). From a significant interaction, we conclude that the effect of the mediator on outcome is not consistent across levels of the exposure, and vice versa, so that the "difference" and "product" estimates of the mediation effect from the traditional approach would no longer be appropriate.

Interaction *P*-values for the companion paper reanalysis are reported in Table 3A. Most interactions were nonsignificant, but for 6 of the 10 outcomes, the anemia–RBC interaction was significant at $P < 0.01$. The corresponding estimated log-odds ratios for the anemia–RBC interaction effects ranged between −0.13 and 0.42, all negative (Table 3B**).** In these models, the effect of the mediator on outcome (i.e., receiving RBCs associated with increase complications) is somewhat stronger for nonanemic than for anemic patients, although in the same direction for both. However, the anemia effect on outcome tends to be positive (associated with more complications) for those without RBC transfusion and negative (associated with less complications) for those with RBC transfusion (data now shown).

Since the anemia and RBC effects on these 6 outcomes were not consistent across levels of the other variable (RBC effect within levels of anemia, and vice versa), a

Table 3A. Exposure–Mediator Interaction Effect *P*-Values—Companion Paper Data

Outcome complication	Open wound	RBCs	Dyspnea	Wound contamination	Independent function	Surgery > 2 h
Systemic	0.64	<0.001	0.71	0.46	0.19	0.001
Respiratory	0.50	<0.001	0.99	0.25	0.35	0.004
Wound Infection	0.009	0.64	0.022	0.26	0.18	0.98
Urinary	0.28	0.16	0.89	0.10	0.25	0.16
CNS	0.65	0.83	0.35	0.03	0.02	0.48
Thrombotic	0.36	0.02	0.17	0.005	0.76	0.88
Cardiovascular	0.49	0.003	0.41	0.72	0.31	0.21
Return to OR	0.02	<0.001	0.59	0.52	0.52	0.58
Mortality	0.10	0.005	0.054	0.54	0.33	0.38
Any	0.77	0.002	0.41	0.77	0.91	0.06

P-values assessing the interaction between each mediator (columns) and outcome (rows) separately. Corresponding interaction effects given in Table 3B.

Table 3B. Exposure–Mediator Interaction Effect Estimates—Companion Paper Data

Outcome complication	Open wound	RBCs	Dyspnea	Wound contamination	Independent function	Surgery > 2 h
Systemic	0.05 (0.10)	−0.26 (0.07)	0.02 (0.07)	−0.04 (0.06)	0.12 (0.09)	−0.19 (0.06)
Respiratory	0.08 (0.12)	−0.37 (0.07)	0.001 (0.06)	0.07 (0.06)	0.08 (0.09)	−0.07 (0.06)
Wound Infection	−0.20 (0.08)	−0.03 (0.06)	−0.12 (0.05)	0.05 (0.04)	0.10 (0.08)	0.001 (0.04)
Urinary	−0.13 (0.12)	−0.11 (0.08)	−0.01 (0.07)	0.09 (0.06)	0.11 (0.09)	−0.08 (0.06)
CNS	0.14 (0.32)	0.04 (0.19)	0.16 (0.17)	−0.29 (0.14)	0.52 (0.22)	−0.10 (0.14)
Thrombotic	0.20 (.22)	−0.27 (0.12)	−0.16 (0.11)	−0.26 (0.09)	0.05 (0.16)	0.01 (0.10)
Cardiovascular	−0.14 (0.20)	−0.42 (0.14)	−0.11 (0.13)	0.04 (0.11)	−0.19 (0.19)	−0.16 (0.12)
Return to OR	0.17 (0.08)	−0.23 (0.06)	−0.03 (0.05)	−0.03 (0.04)	−0.05 (0.08)	−0.02 (0.04)
Mortality	−0.29 (0.17)	−0.34 (0.12)	−0.20 (0.10)	−0.06 (0.10)	0.12 (0.13)	−0.09 (0.10)
Any	−0.05 (0.03)	−0.13 (0.04)	−0.03 (0.03)	−0.007 (0.03)	−0.006 (0.05)	−0.05 (0.03)

mediation analysis as described above is not cleanly interpretable for this proposed mediator. Instead, the relationship may be best described as an interaction between anemia and RBC transfusion on outcome. However, the potential outcomes method of assessing mediation introduced above is more flexible and can be used to estimate mediation effects for each level of the exposure (beyond the scope of this paper).

Sample Size Considerations

Since our minimal definition of mediation requires an association between the exposure and the mediator as well as between the mediator and the outcome, sample size can be calculated to assure adequate power to detect both of these alternative hypotheses. One would then choose the larger sample size for these 2 hypotheses. Hypothesis-specific power must be adequate to account the fact that both hypotheses must be significant. For example, assuming independent tests, an overall power of P is achieved by using a power of $\sqrt{P}$ for each test (i.e., use 0.949 to maintain 0.90 overall power). Since both tests are required to be significant (i.e., this is an intersection-union joint hypothesis test), no Bonferroni correction is needed to the significance criterion for each.

Power for the mediator–outcome relationship in equation (2) is reduced to the extent that exposure and mediator are correlated. The effective sample size for linear regression will be $N^* = N(1 - r_{xm})$,[2] where N is the required sample size ignoring the correlation, and r_{xm} is the correlation between X and M. Sample size for the mediator–outcome relationship is thus $N/(1 - r_{xm})$,[2] where $1/(1 - r_{xm})^2$ is the variance inflation factor in linear regression to adjust for loss of precision due to adjustment for exposure X. Similar adjustments can be made for binary outcomes and other types.[27]

We present an example using a continuous mediator and outcome. Suppose investigators believe that an exposure (new drug versus placebo) decreases the mean of an intraoperative mediator, say, a particular cytokine, by 0.4 SDs or more. With power of 0.949 and α of 0.05, a total of 324 patients are needed for the exposure–mediator relationship (i.e., effect $\boldsymbol{a}$). In addition, the investigators expect the mediator to be correlated with the continuous outcome and would consider correlations of 0.25 or more to be clinically relevant. Thus, with power of 0.949 and α of 0.05, a total of 200 patients are initially calculated for the mediator–outcome relationship (i.e., effect $\boldsymbol{b}$). However, after incorporating an expected correlation of 0.5 between exposure and mediator, a sample size of $200/(1 - 0.5^2) = 267$ is required for effect $\boldsymbol{b}$. The study is planned for the larger of the 2 sample sizes, or N = 324.

More complex approaches involve directly estimating the mediation effect size of interest based on the product $\boldsymbol{a} \times \boldsymbol{b}$ or else the ratio of mediation effect to total effect ($\boldsymbol{a} \times \boldsymbol{b}$/total effect) and then calculating the sample size to detect it with the desired power. Standardized effect sizes such as $\boldsymbol{ab}/(\sigma_x\, \sigma_y)$ are also attractive since they describe the mediation effect on the SD scales.[28]

DISCUSSION

While comparative research typically focuses on whether or not a particular intervention affects outcome, mediation

analysis goes further and asks how, or by what mechanism. Since most researchers do hypothesize 1 or more mechanisms when designing a study, mediation analysis has strong implications for perioperative medicine—from reanalyzing previous studies in which mediation analysis was not considered to designing new studies with a primary or secondary goal of understanding the mechanism(s) of the intervention of interest. Mediation analysis can also be useful for planning future studies in which an identified mediator is manipulated to maximize the effect of an intervention. For example, if in an observational study lower cytokine levels are found to explain much of the effect of a particular steroid on postoperative outcomes, a prospective study might try to manipulate cytokine levels by additional means to further improve outcomes after steroid use. The availability of large databases with rich sets of baseline covariables creates an environment for these analyses to be feasible. Finally, in perioperative medicine, there is often >1 suspected mechanism, and mediation analysis facilitates estimation of the relative contribution of each factor to the total effect of an intervention.

In this paper, we applied the traditional 3-equation approach to mediation analysis of Mackinnon[2] and Barron and Kenny[23] to continuous outcomes and mediators as well as for binary outcomes and mediators. Using data from a companion paper studying the relationship between preoperative hematocrit and outcomes[1] we expanded the 3-equation approach to multiple mediators and a binary outcome. For continuous outcomes and mediators, methods for assessing mediation are fairly straightforward and clear. As we explained, there is still no agreement in the literature on the best way to assess mediation effects when the outcome and/or mediator is binary. We therefore used several competing methods and obtained generally consistent results. Even so, for the binary outcomes, our estimates of mediation effects and proportion mediated should be considered only rough approximations.

We analyzed a set of 10 binary outcomes one at a time. Alternatively, for example, path analysis via structural equation models can be used to assess more complex relationships between exposures, mediators, confounders, and outcomes,[2] typically using LISREL,[29] Mplus,[30] or EQS[31] covariance structure programs. In these models, mediators may be temporally separated, and 1 mediator may be assumed to cause another. Multiple dependent variables, such as in the companion paper, can be included simultaneously in a comprehensive system of equations which accounts for the covariances among all variables. Design and interpretation of such analyses rely on substantial clinical and biological input and should not be treated as a "black box".

As we introduced, mediation can be most clearly defined using the potential outcomes framework, which assumes underlying potential responses to each level of treatment and mediator.[13,14,16,19] Mediation is then the average difference between the potential outcomes that would result for a patient under the value of the mediator if treatment had been received versus under the value of the mediator if treatment had not been received. The difficulty, of course, is that we only observe 1 exposure condition on a patient, and only 1 mediator value, yet we want to make causal inference on the mean effects for both exposure on mediator and mediator on outcome (thus, the fundamental problem of causal inference[15] is magnified).

For any method of assessing mediation, to approach causal inference, we need to remove confounding from each of the exposure–mediator, exposure–outcome, and mediator–outcome relationships. In the present reanalysis of the companion paper data, for example, not only were the anemia and nonanemia exposure groups propensity-matched on a host of baseline confounding variables (as in the original study), but also the same variables were again adjusted for when assessing the mediator–outcome relationships. Although not directly testable, sensitivity analyses can assess how strong an unmeasured confounder would need to be to have affected the conclusions.[32,33]

Direct and indirect effects in mediation analysis can be either "controlled" or "natural", both of which may be of interest, depending on the situation.[11] A controlled direct effect compares the exposure groups on outcome at a fixed level of the mediator, which is the same for all patients. A natural direct effect compares the exposure groups on outcome at the level of the mediator that would be observed for a particular subject at the given value of the exposure, thus allowing the value of the mediator to be subject specific. A similar distinction can be made for the indirect, or mediation, effects. For a continuous outcome variable, the controlled and natural effects are the same, as long as there is no exposure–mediator interaction. For binary outcomes, 2-stage regression or other approaches are needed to estimate natural direct and indirect effects.[12,19]

Identification of potential mediators is particularly challenging when the exposure of interest is a chronic condition as opposed to an acute or intraoperative intervention. For chronic exposures, it may not be clear whether the exposure or the proposed mediator typically occurs first in a patient. Further, to claim mediation in the causal sense, both the exposure and the mediator variables need to be at least theoretically manipulable, or modifiable. This is a questionable assumption for exposures which are chronic conditions, such as diabetes, but reasonable for factors such as smoking status and intraoperative interventions.

The proportion of the total effect due to a specific mediator is an attractive summary measure in mediation analysis. However, since individual mediation effects can be either positive or negative, this "proportion mediated" can also be negative. Therefore, it is better termed a ratio—of individual mediation effect(s) to the total effect. Since positive and negative mediation effects can offset each other, it is also possible that the total effect can be smaller than a particular specific effect so that the specific-to-total ratio can be larger than 1.0. Another option might be to summarize the total mediation effect as the sum of the absolute values of the specific mediation effects, as opposed to the sum of the raw values. A drawback of that approach, though, is that the total mediation effect would no longer represent the net mediation effect. A current area of research in mediation analysis is to identify the

most useful and reliable summary measures for mediation analysis, particularly when there are multiple mediators.

Inconsistent mediation occurs when at least one mediated effect has a different sign than the direct effect (or another mediation effect). For example, an exposure might directly increase values of the outcome variable (i.e., *positive direct effect*). The same exposure might decrease values of a mediator (negative effect ***a***), and higher values of the mediator might increase the outcome variable after adjusting for the exposure (positive effect ***b***), for a *negative mediation effect* (i.e., product of a positive and negative is negative). In such a case, the direct and mediation effects might tend to cancel each other out, resulting in a small (or even 0) total effect (i.e., sum of direct and mediation effects).[10]

Alternatively, *reverse* mediation refers to situations when the outcome variable might cause the mediator, instead of vice versa. Ideally, a researcher would be fairly confident about the direction of potential causality between a mediator and outcome (i.e., mediator causing outcome), including the temporal sequence of events. However, such information is not always available, making conclusions more tenuous.

Finally, sometimes a fourth variable may affect the traditional exposure–mediator–outcome relationship, in the form of either "moderated mediation" or "mediated moderation".[2,23,34,35] Moderated mediation occurs when there is no overall interaction effect between the exposure and the fourth variable, but mediation occurs particularly within levels of fourth "moderating" or "interacting" variable. However, mediated moderation occurs when there is an overall interaction effect of exposure and a fourth variable on outcome, and, for example, people of different combinations of exposure and the interacting variable have different levels of a mediator, thus explaining a difference in the outcome.

In conclusion, mediation analysis is a valuable tool for identifying the mechanisms for why an intervention may or may not affect an outcome of interest and the relative strength of those mechanisms. Careful design and analysis of more studies to assess mediation and mechanism will be a practical step in furthering the march of science in perioperative medicine.

APPENDIX 1. SOBEL METHOD FOR CALCULATING SE OF MEDIATION EFFECT

Using the multivariate delta method, the Sobel SE of the mediation effect[20] for a single mediator can be expressed as

$$se_m = m\sqrt{\frac{t_a^{\,2} + t_b^{\,2}}{t_a t_b}}, \quad \text{here estimated as} \qquad (A1)$$

$$.021\sqrt{\frac{(.023/.001)^2 + (.93/.011)^2}{23.1 \times 85.2}} = 0.00095$$

where $t_a = a/se_a$, $t_b = b/se_b$, ***a*** is effect of exposure on mediator, ***b*** is effect of mediator on outcome adjusting for exposure, m = mediation effect = a × b, and are the SEs for effects ***a*** and ***b***, respectively. Estimates of ***a*** (0.023) and ***b*** (0.093) and their SEs (0.001, 0.011, respectively) from regression models (2) and (3) are plugged into equation (A1) to estimate the mediation SE. This method can be easily extended to multiple mediators with a continuous outcome.

APPENDIX 2. TRADITIONAL REQUIREMENTS FOR CLAIMING MEDIATION

We stated in the section Requirements for Claiming Mediation that contemporary guidelines for claiming mediation include:

1. Intervention affects the mediator (effect ***a*** in (1) is significant)
2. Mediator affects the outcome adjusting for the effect of intervention on outcome (effect ***b*** in (2) is significant)
3. The mediation effect is significant (for example, $\boldsymbol{a} \times \boldsymbol{b} \neq 0$)

The more traditional approach to claiming mediation had 2 additional requirements, [Baron and Kenny 23] namely, that the total effect is significant and that it is larger than the direct effect. Here we explain each and discuss why they are no longer considered generally required:

4. **Total effect of *X* on *Y* must be nonzero (i.e., *c* in (1) ≠ 0).** On the one hand, it makes sense that we cannot discuss mediation of the effect of *X* on *Y* without having a significant marginal effect of *X* on *Y*. However, multiple mediation effects, either observed or nonobserved, may cancel each other out and lead to an overall effect of X on *Y* which is 0. Therefore, many authors (including ourselves) do not require a significant total effect. However, with a single mediator, it would generally follow that the total effect would be significant if both ***a*** and ***b*** were significant.
5. **The direct effect of *X* on *Y* (controlling for *M*) must be smaller than the total effect of *X* on *Y*** (i.e., *c*′ in [2] must be smaller than ***c*** in [1]). Some had gone further to insist that the direct effect had to be nonsignificant, thus implying a complete mediation effect (i.e., mediator explains all of effect of *X* on *Y*).[23] In a single mediation model with effects ***a*** and ***b*** in the same direction, it makes sense to expect the direct effect to be at least smaller than the total effect. However, especially in models with multiple mediators, which might have effects in different directions, this requirement is relaxed.

APPENDIX 3: STANDARDIZED ESTIMATES OF MEDIATION EFFECTS

We use the companion paper data to assess whether baseline functional status mediates the effect of anemia on 30-day mortality. With a binary outcome, we cannot assess mediation using the simple total effect minus direct effect method (***c*** minus ***c***′) since the outcome variables are on different scales for the various equations. Choices are to approximate the mediation effect using the $\boldsymbol{a} \times \boldsymbol{b}$ method or first standardize the regression coefficients and then use either the ***c*** minus ***c***′ method or the $\boldsymbol{a} \times \boldsymbol{b}$ method. The latter method is preferred and most generalizable. Below we compare results using all 3 methods.

Variables are as follows: Y = mortality (1 = Yes, 0 = No); X = anemic status (1 = Yes, 0=No); M = Functional status (1 = independent, 2 = somewhat independent, 3 = dependent).

We first fit equations (1) to (3) with the following estimates

Total effect ***c*** = 0.46 (odds ratio 1.5 (95% CI 1.4–1.7)

Direct effect of anemia on mortality as c' = 0.41 (odds ratio 1.5 (1.4–1.6)
Effect of anemia on mediator (functional status), a = 0.023 (0.021–0.025)
Effect of mediator on mortality, adjusting for mediator, b = 0.74 (odds ratio 2.1 (1.9–2.3)

1. Nonstandardized product (a × b) method

In the product method, the mediation (or indirect) effect is estimated as $a \times b$ = 0.023 × 0.74 = 0.0167 (95% CI, 0.01–0.02). Proportion of the total effect explained by functional status is $(a \times b)/TE$, where TE is the estimated total effect, calculated as the direct effect plus the mediation effect (i.e., $c' + a \times b$). The proportion of the total effect of anemia on mortality mediated by functional status is estimated as 0.0167/(0.41 + 0.0167) = 0.0167/0.4267 = 0.039, or about 4%.

2. Standardized coefficient method

We standardize [36] coefficients c (from equation [1], and c′ and b (from equation [2]) by dividing by the variance of Y in their respective equations (Steps 1–3) and then assess mediation effects (Step 4).

Step 1. Estimated variance of Y for equation (1) is $\hat{\sigma}_Y^2(1) = \hat{c}^2\hat{\sigma}_X^2 + \pi^2/3$ where $\hat{\sigma}_X^2$ is the estimated variance of X and π=3.14159.
Plugging in the companion paper estimates, we have

$$\hat{\sigma}_Y(1) = \sqrt{\hat{\sigma}_Y^2(1)} = \sqrt{0.46^2 \times 0.093^2 + \pi^2/3} = 1.813$$

Step 2. Estimated variance of Y for equation (2) is: $\hat{\sigma}_Y^2(2) = \hat{b}^2\hat{\sigma}_M^2 + \hat{c}'^2\sigma_X^2 + 2\hat{b}\hat{c}'\hat{\rho}_{XM} + \pi^2/3$ where $\hat{\sigma}_X^2$and $\hat{\sigma}_M^2$ are estimated variances of X and M, $\hat{\rho}_{XM}$is the X-M correlation and π=3.14159
Plugging in the companion paper estimates, we have:

$$\hat{\sigma}_Y(2) = \sqrt{\hat{\sigma}_Y^2(2)} = \sqrt{0.74^2 \times 0.24^2 + 0.41^2 \times 0.093^2 + 2 \times 0.74 \times 0.41 \times 0.0472 + \pi^2/3} = 0.2262$$

Step 3. Use above estimates to standardize c (total effect), c' (direct effect), b and their standard errors, as:

Standardized c = 0.4638 / 1.813 = 0.2558
Standardized SE of c = 0.0450 / 1.813 = 0.0248
Standardized c' = 0.4122 / 1.822 = 0.2262
Standardized SE of c' = 0.0462 / 1.822 = 0.0254
Standardized b = 0.7361 / 1.822 = 0.4040
Standardized SE of b = 0.0521 / 1.822 = 0.0286

Effect a from equation (3) does not need to be standardized since linear regression.

Step 4. Estimate standardized meditation effects.

A. Difference method:
   Mediation effect $c - c'$ = 0.2558 − 0.2262 = 0.0296
   Proportion mediated = 0.0296/0.2558 = 0.12, or about 12%.
B. Product method:
   Mediation effect (i.e., indirect effect) = $a \times b$ = 0.0227 × 0.4040 = 0.00917
   1) Total effect using indirect + direct = $a \times b + c'$ = 0.00917 + 0.2262 = 0.2354Proportion mediated = $a \times b$/total effect = 0.00917/0.2354 = 0.041, or about 4%
      We obtained a bootstrap confidence interval of 4% (2.8%–4.5%).

The standardized $a \times b$ mediation effect is thus estimated as 0.023 × 0.404 = 0.0092, which explains 4% (95% CI, 2.8–4.5) of the total effect, very similar to the nonstandardized $a \times b$ method above.

Reasonable results can also be obtained by using the product method with standardized c as estimate of total effect, as below:

   2) Total effect using standardized c = 0.2558.Proportion mediated = $a \times b/c$ = 0.00917/0.2258= 0.026, about 3%

DISCLOSURES

Name: Edward J. Mascha, PhD.
Contribution: This author conceived the idea, designed the study, researched the statistical methods, analyzed the data, and drafted the manuscript; senior author.
Attestation: Edward Mascha approved the final manuscript.
Name: Jarrod E. Dalton, PhD.
Contribution: This author helped design the study and edit the manuscript.
Attestation: Jarrod Dalton approved the final manuscript.
Name: Andrea Kurz, MD.
Contribution: This author helped design the study.
Attestation: Andrea Kurz approved the final manuscript.
Name: Leif Saager, MD.
Contribution: This author helped design the study and edit the manuscript.
Attestation: Leif Saager approved the final manuscript.
This manuscript was handled by: Franklin Dexter, MD, PhD.

ACKNOWLEDGMENTS

The authors wish to heartily thank the reviewers for their thoughtful insights and suggestions which have substantially improved the quality of this paper.

REFERENCES

1. Saager Leif , Turan A, Reynolds LF, Dalton JE, Turner PL, Kurz A. The Association Between Preoperative Anemia and 30-Day Mortality and Morbidity in Noncardiac Surgical Patients. Anesth Analg 2013; 909–15
2. MacKinnon DP. Introduction to Statistical Mediation Analysis. New York, NY: Psychology Press, 2008
3. Shrout PE, Bolger N. Mediation in Experimental and Nonexperimental Studies. Psychol Methods 2002;7:422–45
4. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. Biometrika 1983;70:41–55
5. Imai K, Dyk DAV. Causal Inference with General Treatment Regimes: Generalizing the Propensity Score. J Am Stat Assoc 2004;99:854–66
6. Ming K, Rosenbaum PR. Substantial gains in bias reduction from matching with a variable number of controls. Biometrics 2000;56:118–24
7. Schwartz S, Li F, Reiter JP. Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables. Stat Med 2012;31:949–62

8. Turan A, Mascha EJ, Roberman D, Turner PL, You J, Kurz A, Sessler DI, Saager L. Smoking and perioperative outcomes. Anesthesiology 2011;114:837–46
9. Abdelmalak B, Maheshwari A, Mascha E, Srivastava S, Marks T, Tang WW, Kurz A, Sessler DI. Design and Organization of the Dexamethasone, Light Anesthesia and Tight Glucose Control (DeLiT) Trial: a factorial trial evaluating the effects of corticosteroids, glucose control, and depth-of-anesthesia on perioperative inflammation and morbidity from major non-cardiac surgery. BMC Anesthesiol 2010;10:11
10. David P. MacKinnon, Amanda J. Fairchild, Matthew S. Fritz. Mediation Analysis. Annu Rev Psychol 2007;58
11. VanderWeele TJ. Mediation and mechanism. Eur J Epidemiol 2009;24:217–24
12. Albert JM, Nelson S. Generalized causal mediation analysis. Biometrics 2011;67:1028–38
13. Gallop R SD, Lin J, Elliott M, Joffe M, Ten Have T. Mediation analysis with principal stratification. Stat Med 2009;8:1108–30
14. Imai K, Keele L, Tingley D, Yamamoto T. Unpacking the Black Box of Causality: Learning about Causal Mechanisms from Experimental and Observational Studies. Am Polit Sci Rev 2011;105
15. Rubin DB. Estimating causal effects of treatments in randomized and nonrandomized studies. J Educational Psychol 1974;66:688–701
16. Rubin DB. Direct and indirect causal effects via potential outcomes. Scand J Stat 2004;31:161–70
17. Robins JM, Hernán MA, Brumback B. Marginal structural models and causal inference in epidemiology. Epidemiology 2000;11:550–60
18. Pearl J. Causal diagrams for empirical research Biometrika 1995;82:669–88
19. Albert JM. Mediation analysis via potential outcomes models. Stat Med 2008;27:1282–304
20. Sobel ME. Direct and Indirect Effects in Linear Structural Equation Models. Sociol Methods Res 1987;16:155–76
21. Efron B TR. An Introduction to the Bootstrap. London: Chapman and Hall, 1993
22. Collins L, Graham J, Flaherty B. An alternative framework for defining mediation. Multivariate Behav Res 1998;33:295–312
23. Baron RM, Kenny DA. The moderator-mediator variable distinction in social psychological research: conceptual, strategic, and statistical considerations. J Pers Soc Psychol 1986;51:1173–82
24. Vanderweele TJ, Vansteelandt S. Odds ratios for mediation analysis for a dichotomous outcome. Am J Epidemiol 2010;172:1339–48
25. Preacher KJ, Hayes AF. Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. Behav Res Methods 2008;40:879–91
26. Imai K, Keele L, Tingley D. A general approach to causal mediation analysis. Psychol Methods 2010;15:309–34
27. Vittinghoff E, Sen S, McCulloch CE. Sample size calculations for evaluating mediation. Stat Med 2009;28:541–57
28. Preacher KJ, Kelley K. Effect size measures for mediation models: quantitative strategies for communicating indirect effects. Psychol Methods 2011;16:93–115
29. Joreskog KG, Van Thillo M. LISREL: A General Computer Program for Estimating a Linear Structural Equation System Involving Multiple Indicators of Unmeasured Variables (RB-72-56), 1972
30. Muthén B, Muthén L. Mplus User's Guide. 6th ed. Los Angeles, CA: Muthén & Muthén. Available at: www.statmodel.com. Los Angeles, CA, 2010. Accessed January 6, 2013
31. Bentler PM. EQS Structural Equations Program Manual. Encino, CA: Multivariate Software, 1995
32. Lin DY, Psaty BM, Kronmal RA. Assessing the sensitivity of regression results to unmeasured confounders in observational studies. Biometrics 1998;54:948–63
33. Brumback BA, Hernán MA, Haneuse SJ, Robins JM. Sensitivity analyses for unmeasured confounding assuming a marginal structural model for repeated measures. Stat Med 2004;23:749–67
34. Muller D, Judd CM, Yzerbyt VY. When moderation is mediated and mediation is moderated. J Pers Soc Psychol 2005;89:852–63
35. Morgan-Lopez AA, MacKinnon DP. Demonstration and evaluation of a method for assessing mediated moderation. Behav Res Methods 2006;38:77–87
36. Winship C, Mare RD. Structural Equations and Path Analysis for Discrete Data. The Am J Sociol 1983;89:54–110