Section Editor: **Thomas R. Vetter**

SPECIAL ARTICLE

Bias, Confounding, and Interaction: Lions and Tigers, and Bears, Oh My!

Thomas R. Vetter, MD, MPH,* and Edward J. Mascha, PhD†

Epidemiologists seek to make a valid inference about the causal effect between an exposure and a disease in a specific population, using representative sample data from a specific population. Clinical researchers likewise seek to make a valid inference about the association between an intervention and outcome(s) in a specific population, based upon their randomly collected, representative sample data. Both do so by using the available data about the sample variable to make a valid estimate about its corresponding or underlying, but unknown population parameter. Random error in an experiment can be due to the natural, periodic fluctuation or variation in the accuracy or precision of virtually any data sampling technique or health measurement tool or scale. In a clinical research study, random error can be due to not only innate human variability but also purely chance. Systematic error in an experiment arises from an innate flaw in the data sampling technique or measurement instrument. In the clinical research setting, systematic error is more commonly referred to as systematic bias. The most commonly encountered types of bias in anesthesia, perioperative, critical care, and pain medicine research include recall bias, observational bias (Hawthorne effect), attrition bias, misclassification or informational bias, and selection bias. A confounding variable is a factor associated with both the exposure of interest and the outcome of interest. A confounding variable (confounding factor or confounder) is a variable that correlates (positively or negatively) with both the exposure and outcome. Confounding is typically not an issue in a randomized trial because the randomized groups are sufficiently balanced on all potential confounding variables, both observed and nonobserved. However, confounding can be a major problem with any observational (nonrandomized) study. Ignoring confounding in an observational study will often result in a "distorted" or incorrect estimate of the association or treatment effect. Interaction among variables, also known as effect modification, exists when the effect of 1 explanatory variable on the outcome depends on the particular level or value of another explanatory variable. Bias and confounding are common potential explanations for statistically significant associations between exposure and outcome when the true relationship is noncausal. Understanding interactions is vital to proper interpretation of treatment effects. These complex concepts should be consistently and appropriately considered whenever one is not only designing but also analyzing and interpreting data from a randomized trial or observational study. (Anesth Analg 2017;125:1042–8)

Lions and tigers, and bears, oh my!— Dorothy, the Tin Man, and the Scarecrow

L. Frank Baum (1900), *The Wonderful Wizard of Oz*

In this basic statistical tutorial, we discuss the complex yet important, plus often misunderstood, topics of: (1) random error (chance) versus systematic error (bias); (2) common types of study bias; (3) confounding; and (4) interaction.

RANDOM ERROR VERSUS SYSTEMATIC ERROR (SYSTEMATIC BIAS)

Epidemiologists seek to make a valid inference about the causal effect between an exposure and a disease in a specific population, using representative sample data from a specific population.[1,2] Clinical researchers likewise seek to make a valid inference about the association between an intervention and outcome(s) in a specific population, based upon their randomly collected, representative sample data.[3] Both do so by using the available data about the sample variable to make a valid estimate about its corresponding or underlying, but unknown population parameter.[4–6]

The relationship between the observed estimate, the population parameter, the random error, and the systematic error can be represented as a simple equation (Figure 1).[7]

Random Error

Random error in an experiment can be due to the natural, periodic fluctuation or variation in the accuracy or precision of virtually any data sampling technique or health measurement tool or scale.[8] Random error is equally likely to distort study measurements in either a positive or negative direction.[9]

For example, a conventional manual sphygmomanometer reports the systolic blood pressure to be ±20 mm Hg, from moment to moment, due to respiration, emotion, exercise, meals, tobacco or alcohol use, body temperature, bladder distension, pain, and circadian rhythm.[10]

In a clinical research study, random error can be due to not only innate human variability but also purely chance.[9,11]

From the *Department of Surgery and Perioperative Care, Dell Medical School at the University of Texas at Austin, Austin, Texas; and †Departments of Quantitative Health Sciences and Outcomes Research, Cleveland Clinic, Cleveland, Ohio.

Accepted for publication June 5, 2017.

Funding: None.

The authors declare no conflicts of interest.

Reprints will not be available from the authors.

Address correspondence to Thomas R. Vetter, MD, MPH, Department of Surgery and Perioperative Care, Dell Medical School at the University of Texas at Austin, Dell Pediatric Research Institute, Suite 1.114, 1400 Barbara Jordan Blvd, Austin, TX 78723. Address e-mail to thomas.vetter@austin.utexas.edu.

DOI: 10.1213/ANE.0000000000002332

Figure 1. The relationship between the observed estimate, the population parameter, the random error, and the systematic error.[7]

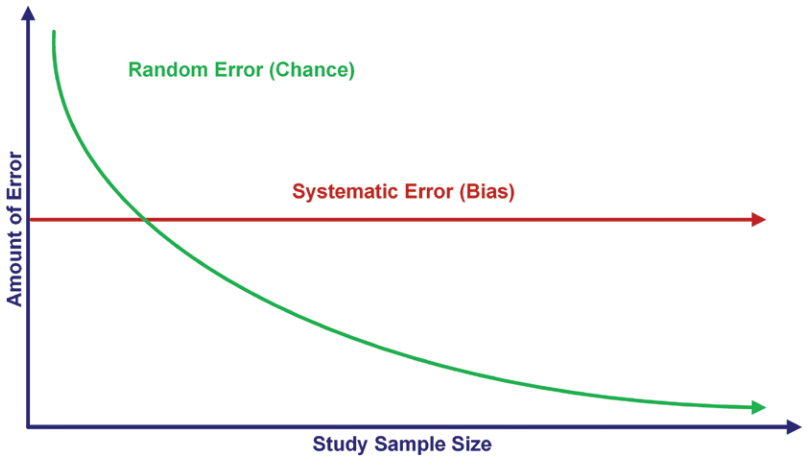

Figure 2. The relationship between random error, systematic error, and sample size.[16]

Random error is nothing more than variability in the sample data that cannot be readily explained.[12] However, the amount of such random error can be mathematically estimated and adjusted for statistically.[12] The methods for doing so will be the topics of future statistical tutorials.

Systematic Error (Systematic Bias)

Systematic error in an experiment arises from an innate flaw in the data sampling technique or measurement instrument.[8] Systematic error is nonrandom variation that distorts the research study findings in 1 direction.[9]

For example, a conventional manual sphygmomanometer consistently reports the systolic blood pressure to be 10–20 mm Hg greater than its true value when a cuff size is used that is too small relative to the patient's arm circumference.[13]

In the clinical research setting, systematic error is more commonly referred to as systematic bias. This systematic error or bias results in an incorrect (invalid) estimate of the measure of association or treatment effect.[14] Such systematic bias thus undermines the internal validity of the study (ie, whether it actually and accurately measured what it set out to).[14,15] If its presence is recognized, the amount of some forms of systematic error or systematic bias can be mathematically estimated and adjusted for statistically. However, unlike in conventional epidemiological studies, this is not as commonly done in randomized controlled trials.

Effect of Sample Size

As the study sample size increases, the effect of random error or chance typically decreases (Figure 2).[16] Thus, a larger sample size typically results in a more precise estimate of the primary outcome or relationship of interest.[9] In contrast, because its adverse effect is essentially fixed, systematic error or bias typically does not change with an increasing sample size (Figure 2).[9,16]

COMMON TYPES OF STUDY BIAS

There are 3 main mechanisms through which bias is introduced into health care–related research: (1) factors that relate to the exposure of patients to treatments in the study population; (2) factors that influence inclusion of patients in the study; and (3) factors that affect the assessment and measurement of outcomes.[17] A myriad of over 70 potential types and sources of bias exist in conducting human subjects research.[18] However, we have elected to focus here on a handful of the most commonly encountered types of bias in anesthesia, perioperative, critical care, and pain medicine research (Table).

Recall Bias

Recall bias is a form of information bias. Recall bias occurs when 1 study group or subgroup has a differential recall of exposures or events prior to the onset of a disease.[19] Differential recall between study groups is especially a concern with a retrospective case–control study.[20,21] Compared to controls without the disease, cases with the disease are more likely to provide an extensive and complete report of their true exposure to a hypothesized risk factor, thereby biasing upwards the estimate of its effect.[20,21]

A number of factors can influence study participant recall. Cases tend to search their memories to identify what might have caused their disease; healthy controls have no such motivation.[19] Past exposures may be more meaningful and hence apparent to cases, possibly because of their greater awareness of potential risk factors for their condition or because of repeated physician interviews.[22] Conversely, controls may have had less contact with health care providers and be less sensitized to questions about previous exposures.[22]

In an effort to minimize recall bias in a case–control study: (1) investigators should blind the data gatherers to the case or control status of participants, or if not possible, at least blind them to the main study hypothesis; (2) data

Table. Some of the Most Commonly Encountered Types of Bias in Anesthesia, Perioperative, Critical Care, and Pain Medicine Research

Type of Bias	Definition
Recall bias	Recall bias occurs when 1 study group or subgroup has a differential recall of exposures or events before the onset of a disease
Observational bias	Hawthorne effect is a specific form of observational bias in which the mere awareness of being under observation can alter the way in which a person (patient or provider) behaves
Attrition bias	Attrition bias exists if the characteristics of study participants who are lost to follow-up differ between the randomized treatment groups or observational study cohorts
Misclassification bias	A process that produces a systematic error in the information about the exposure and/or disease (outcome) results in nondifferential or differential misclassification bias
Selection bias	Selection bias is a distortion of the study findings that results from the factors that determine study participation, specifically, the procedure or the way in which study subjects are selected

gatherers need to be thoroughly trained to elicit exposure in a similar manner from cases and controls; and (3) data gathers should use the same memory aids to facilitate and balance recall between cases and controls.[21]

However, recall bias is not unique to case–control studies. An increasing number of health economic studies rely on patient survey-based, self-reported data to obtain information on health care utilization, out-of-pocket expenses, health behaviors, and health status.[23] The length of the recall period in self-reported health care questions varies among surveys (eg, 1, 3, 6, or 12 months), and this variation can affect the results of such studies.[23]

Observational Bias

First described by Landsberger[24] in 1961, the Hawthorne effect is a specific form of observational bias in which the mere awareness of being under observation can alter the way in which a person behaves.[25,26] The Hawthorne effect is a form of reactivity in which study subjects may improve or modify their behavior, which is being experimentally measured, in response to their knowing that they are being observed, not in response to a specific experimental intervention.[25] For example, if they know that their behavior is being observed, patients may increase their level of engagement and hence improve compliance with preoperative medication instructions on the day of surgery.[27] The Hawthorne effect likely also plays a role in all blood utilization programs seeking to optimize transfusion practice, if clinicians know that their blood ordering practices are being monitored and tracked.[28]

Such audit and feedback is widely used in quality improvement initiatives as a strategy to improve professional practice and clinical outcomes.[29] Like the employees in the 1920s and 1930s at the eponymous Western Electric Hawthorne Works factory outside Chicago,[24] modern-day health care providers may also, at least temporarily, change their behavior (eg, hand hygiene protocol compliance) in response to their clinical performance being overtly monitored.[30] Alternative approaches like a proxy variable (eg, measuring instead the amount hand hygiene product used)[31] or unrecognized observers (mystery or secret "shoppers")[32] can lessen the Hawthorne effect in clinical performance improvement settings.

Observational bias is generally considered less likely in a randomized study when both groups are being observed, in which case, the Hawthorne effect within the groups should be equal.[33] The Hawthorne effect is a greater concern in nonrandomized or quasiexperimental studies, or when only 1 study group is aware that it is being observed.[26]

Attrition Bias

In a randomized clinical trial (RCT) or an observational cohort study, some participants invariably drop out of the study for multiple reasons.[34] When such attrition occurs and data are missing at random, the power of the study is weakened, but this is typically not a major problem.[35] Conversely, such attrition can introduce systematic bias if the characteristics of study participants who are lost to follow-up differ between the randomized treatment groups or observational study cohorts.[36]

This differential loss to follow-up is important if the resulting differing characteristic is correlated with the primary outcome measure(s) of the study.[36] Likewise, when data are missing because of aspects of treatment or the disease, major bias can arise.[35] Specifically, patients with missing outcome observations are more likely to be patients with poor outcomes.[35]

However, there is no definitive amount of attrition or loss to follow-up above which attrition bias is an acknowledged problem.[36] Schulz and Grimes,[37] and others, have historically suggested a simple "5-and-20 rule of thumb," with <5% attrition or loss to follow-up probably resulting in little bias, >20% loss potentially posing serious threats to validity, and in-between levels leading to an intermediate problem.

Although this 5-and-20 attrition or loss to follow-up rule may have overall utility, it is less valid and applicable in studies with infrequent or rare outcomes.[37] So-called fragility of findings may therein exist, in which substantial changes in *P* values occur with small changes in the number of patients experiencing an event in the treatment group or exposed cohort.[38,39]

Misclassification or Informational Bias

Misclassification bias is also referred to as an informational bias. The mechanism that produces a systematic error in the information about the exposure and/or disease (outcome) can result in either nondifferential or differential misclassification bias.[16,18,40] Differential misclassification bias means that this bias is more prevalent in 1 group than the other, which is problematic. The reader is referred to more in-depth discussions of the key distinctions between nondifferential versus differential misclassification bias and their related analytical adjustments.[41–44]

Informational bias is of particular concern in medical research that uses medical information systems (electronic medical records) and administrative claims databases to

determine whether patients have preexisting medical conditions or have undergone previous procedures.[45,46] For example, potentially confounding variables like preexisting medical conditions can be under-reported or incorrectly reported in a hospital database or administrative claims database and this can occur disproportionately between the exposure groups of interest (eg, patients receiving 2 different types of anesthesia or analgesia). As a result of such misclassification, the estimated relationship between the exposure and outcome of interest can be distorted.

Selection Bias and Randomization

It is possible that the way researchers collect their study sample is not truly random, but instead subjects are preferentially sampled to have particular characteristics.[6] Selection bias is a distortion of the study findings that results from the factors that determine study participation, specifically, the procedure or way in which study subjects are selected.[41,47]

With selection bias, the relationship between the exposure (or treatment) and the disease (or outcome) is fundamentally different for those individuals who participate or are enrolled in the study versus for all those who should have been eligible for the study, including those who did not participate.[41]

Selection bias can even occur in a RCT, if the recruiters and observers can guess the next group assignment with >50% probability.[48] "When future allocations can be predicted, which is the case when masking [blinding] is absent or imperfect and restricted [block] randomization is used (ie, just about always), one can funnel healthier patients to one group and sicker patients to the other group."[49]

Simple randomization ("complete" or "unrestricted" randomization), in which every study participant has an equal chance of group assignment, regardless of the previous participants' allocations (analogous to repeated coin-tossing), is the simplest and most effective method to prevent selection bias in an RCT.[48,50,51]

However, simple randomization is infrequently used because researchers (and journal editors and reviewers) typically prefer a randomization method that generates a balanced number of patients assigned to each treatment group.[48,52,53] Restricted randomization (eg, permuted block randomization) is thus instead applied, which results in similar number of patients being assigned to each treatment group—and thus minimizes loss of statistical power due to sample size imbalance—as well as better balances known and unknown confounders.[48,51–53]

Block randomization (either fixed or randomly sized) is also helpful to achieve balance within sites of a multicenter randomized trial and to achieve balance within predetermined strata such as age or gender groups.

Nevertheless, we agree with others that simple randomization should be used more frequently than fixed-block randomization.[48,50,52] For non–double-blinded RCTs with more than 200 total subjects, researchers should use simple randomization and accept moderate disparities in group sizes.[50,52] For non–double-blinded RCTs with a total sample size of <200, researchers can apply block randomization but should randomly vary the block sizes and include larger block sizes than 10.[51]

The reader is referred to a more advanced discussion of selection bias, allocation concealment, and randomization design in clinical trials,[54,55] in addition to a concise, practical guide on randomization and allocation concealment for clinical researchers.[56,57]

CONFOUNDING

"A simple definition of confounding is the confusion of effects."[16] A confounding variable (confounding factor or confounder) is a variable that correlates (positively or negatively) with both the exposure and outcome.[14] Confounding can be a major problem with any observational (nonrandomized) study.[14,16,58,59] Ignoring confounding in an observational study will often result in a distorted or incorrect estimate of the association or treatment effect.[14,16,58,59] Epidemiologists often define a confounding variable as 1 that will alter the estimated treatment effect by 15% or more when not accounted for mathematically.

For example, in a retrospective database study assessing the relationship between general anesthesia versus neuraxial or regional anesthesia on postoperative outcomes, researchers need to take into account if the general anesthesia patients have more or less comorbidity at baseline than those receiving neuraxial or regional anesthesia. Not doing so would confound or "muddle" the actual treatment effect of the anesthetic technique.[60,61]

Researchers likewise assessing whether patients receiving an intraoperative blood transfusion is associated with greater postoperative complications compared to those patients not receiving a blood transfusion need to adjust for the fact that those receiving blood are sicker at baseline by adjusting for as many baseline risk factors (comorbidities) as possible.[62] However, when the exposure groups are as different as in this example, it may not be possible to adequately adjust for confounding.

When defining confounding, it is important to consider the temporal relationship between the purported confounding variable and the exposure variable. A confounding variable must occur or be measured before the exposure variable (or exposure period).[58,59,63] A contributing factor that occurs after the exposure of interest, even though associated with the exposure and the outcome, would not be a confounding variable. Such a variable might instead be a mediator variable.

A mediator variable is located along the causal pathway between the exposure and the outcome; it occurs a result of the exposure and is a contributing cause of the outcome.[63–65] Researchers need to be careful to not adjust for mediators when assessing an exposure versus outcome relationship, since doing so will tend to "washout" the treatment effect of interest.[66]

Confounding is typically not an issue in a randomized trial because the randomized groups are sufficiently balanced on all potential confounding variables, both observed and nonobserved. Therefore, even though many baseline variables such as age, body mass index, and baseline comorbidities may be even strongly related to the outcome of interest in a clinical trial, these variables are not confounding variables because they are not associated with the exposure of interest—only the outcome. In smaller clinical

trials, some nontrivial imbalance in baseline variables may occur, and such variables can be adjusted for in the analyses to solve the problem.[67–69]

The Relationship Between Selection Bias and Confounding

Selection bias is a common source of confounding.[70] Selection bias may lead to confounding, when 1 or more of the predictor variables that determine or predispose assignment to the intervention also directly affects the outcome.[71] This uncorrected association of a predictor variable with both the intervention and the outcome can distort the treatment effect of interest and thus result in a type I error, in which the outcome of the intervention (treatment effect) is falsely attributed to the intervention rather than to the confounding variable.[71] Alternatively, selection bias and the resulting confounding can result in a type II error, in which the study incorrectly concludes that there is no treatment effect from the intervention itself.[71] Importantly, such distortion can occur even though neither of the 2 error types occur.

A common and important type of confounding in clinical research is confounding by indication, which occurs when the clinical indication for selecting a particular treatment or intervention (eg, the severity of the illness) also affects the outcome.[72] Confounding by indication is sometimes is referred to interchangeably as selection bias.[47,70]

Controlling for Confounding

Researchers can routinely implement study design procedures to prevent confounding (eg, randomization, study eligibility restriction, and/or a priori participant matching), but most confounding is removed by statistical procedures in the subsequent data analysis (eg, multivariable regression models, propensity score methods, and/or instrumental variables) for both clinical trials and more typically observational studies.[68,69,71–77] This will be the topic of a future statistical tutorial.

INTERACTION (EFFECT MODIFICATION)

Interaction among variables, also known as effect modification, exists when the effect of 1 explanatory variable on the outcome depends on the particular level or value of another explanatory variable.[14,78,79]

For example, researchers might find that the effect of anesthetic A versus anesthetic B on the outcome of interest is stronger (or even in opposite directions) for men than for women, indicating an interaction between effect of anesthetic choice and gender on the outcome.

In other words, an interaction exists when the incidence of the disease or outcome in the presence of 2 or more risk factors or exposures differs from the incidence rate observed or expected to result from their individual effects. If interaction between risk factors or exposures is present, the factors are not independent in causing a specific outcome.[79–82]

When the joint effect of 2 or more explanatory variables is discernibly larger or smaller than the "sum of the parts" (or individual effects), there is an interaction among the explanatory variables.[78,83] The effect can be greater than would be expected (positive interaction or synergism) or less than would be expected (negative interaction or antagonism).[78,83] From a more rigorous, epidemiological perspective, synergism is fundamentally defined by and thus requires the presence of a positive interaction among variables.[78,83] When reporting on the presence of interactions in their manuscript, the authors need to decide whether or not to report the treatment effect separately for each level (or subgroup) of the interacting variable. The decision on how to proceed can be made at least partly by knowing whether the interaction is quantitative or qualitative.[79] "A key principle for interpretation of subgroup results is that quantitative interactions (differences in degree) are much more likely than qualitative interactions (differences in kind)."[84]

A quantitative interaction (noncrossover interaction) exists when the treatment effect on the outcome is in the same direction (positive or negative) for both levels (or subgroups) of an interacting variable, but the size of the treatment effect differs significantly between levels (or subgroups).[79] "Quantitative interactions are to be expected, but may not be important clinically."[85] In such cases, the interaction effect thus may be interesting, but it may be valid to ignore the interaction and to report the observed effect of treatment on the outcome as the primary result.[83,84]

A qualitative interaction (crossover interaction) occurs when the treatment effects are in opposite directions for both levels (or subgroups) of an interacting variable.[79] In the presence of a qualitative interaction, the primary result for the treatment effect on outcome can be reported for each level (subgroup) of the interacting variable.[83,86]

CONCLUSIONS

Bias and confounding are common potential explanations for statistically significant associations between exposure and outcome when the true relationship is noncausal. Understanding interactions is vital to proper interpretation of treatment effects.[14] These complex concepts should be consistently and appropriately considered whenever one is not only designing but also analyzing and interpreting data from a randomized trial or observational study.[14]

DISCLOSURES

Name: Thomas R. Vetter, MD, MPH.
Contribution: This author helped write and revise the manuscript.
Name: Edward J. Mascha, PhD.
Contribution: This author helped write and revise the manuscript.
This manuscript was handled by: Jean-Francois Pittet, MD.

REFERENCES

1. Rothman KJ, Greenland S, Poole C, Lash TL. Causation and causal inference. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. Revised 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012:5–31.
2. Rothman KJ. What is causation? In: *Epidemiology: An Introduction*. 2nd ed. Oxford, UK: Oxford University Press; 2012:23–37.
3. Glasser SP. Introduction to clinical research concepts, essential characteristics of clinical research, overview of clinical research study designs. In: Glasser SP, ed. *Essentials of Clinical Research*. 2nd ed. Cham, Switzerland: Springer; 2014:11–32.
4. Salkind NJ. Statistics or sadistics? It's up to you. In: *Statistics for People Who (Think They) Hate Statistics*. 6th ed. Thousand Oaks, CA: Sage Publications; 2016:5–18.

5. Urdan TC. Introduction to social science research principles and terminology. In: *Statistics in Plain English*. 4th ed. New York, NY: Routledge, Taylor & Francis Group; 2017:1–12.
6. Motulsky H. From sample to population. In: *Intuitive Biostatistics: A Nonmathematical Guide to Statistical Thinking*. New York, NY: Oxford University Press; 2014:22–28.
7. Gerstman BB. Error in epidemiologic research. In: *Epidemiology Kept Simple: An Introduction to Traditional and Modern Epidemiology*. 3rd ed. Chichester, UK: John Wiley & Sons; 2013:201–221.
8. Bowling A. The principles of research. In: *Research Methods in Health: Investigating Health and Health Services*. 4th ed. Maidenhead, UK: Open University Press; 2014:146–188.
9. Hulley SB, Newman TB, Cummings SR. Getting started: The anatomy and physiology of clinical research. In: *Designing Clinical Research*. 4th ed. Philadelphia, PA: Wolters Kluwer Health/Lippincott Williams & Wilkins; 2013:2–13.
10. Beevers G, Lip GY, O'Brien E. ABC of hypertension. Blood pressure measurement. Part I-sphygmomanometry: factors common to all techniques. *BMJ*. 2001;322:981–985.
11. Keus F, Wetterslev J, Gluud C, van Laarhoven CJ. Evidence at a glance: error matrix approach for overviewing available evidence. *BMC Med Res Methodol*. 2010;10:90.
12. Rothman KJ. Random error and the role of statistics. In: *Epidemiology: An Introduction*. 2nd ed. Oxford, UK: Oxford University Press; 2012:148–163.
13. Ogedegbe G, Pickering T. Principles and techniques of blood pressure measurement. *Cardiol Clin*. 2010;28:571–586.
14. Glasser SP. Bias, confounding, and effect modification (interaction). In: Glasser SP, ed. *Essentials of Clinical Research*. 2nd ed. Cham, Switzerland: Springer; 2014:362–373.
15. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359:248–252.
16. Rothman KJ. Dealing with biases. In: *Epidemiology: An Introduction*. 2nd ed. Oxford, UK: Oxford University Press; 2012:124–147.
17. Gerhard T. Bias: considerations for research practice. *Am J Health Syst Pharm*. 2008;65:2159–2168.
18. Delgado-Rodríguez M, Llorca J. Bias. *J Epidemiol Community Health*. 2004;58:635–641.
19. Grimes DA, Schulz KF. Bias and causal associations in observational research. *Lancet*. 2002;359:248–252.
20. Neugebauer R, Ng S. Differential recall as a source of bias in epidemiologic research. *J Clin Epidemiol*. 1990;43:1337–1341.
21. Schulz KF, Grimes DA. Case-control studies: research in reverse. *Lancet*. 2002;359:431–434.
22. Coughlin SS. Recall bias in epidemiologic studies. *J Clin Epidemiol*. 1990;43:87–91.
23. Kjellsson G, Clarke P, Gerdtham UG. Forgetting to remember or remembering to forget: a study of the recall period length in health care survey questions. *J Health Econ*. 2014;35:34–46.
24. Landsberger HA. *Hawthorne Revisited: Management and the Worker, Its Critics and Developments in Human Relations in Industry*. Ithaca, NY: Cornell University Press; 1961.
25. De Amici D, Klersy C, Ramajoli F, Brustia L, Politi P. Impact of the Hawthorne effect in a longitudinal clinical study: the case of anesthesia. *Control Clin Trials*. 2000;21:103–114.
26. McCambridge J, Witton J, Elbourne DR. Systematic review of the Hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol*. 2014; 67:267–277.
27. Vetter TR, Downing ME, Vanlandingham SC, Noles KM, Boudreaux AM. Predictors of patient medication compliance on the day of surgery and the effects of providing patients with standardized yet simplified medication instructions. *Anesthesiology*. 2014;121:29–35.
28. Savage W. Implementing a blood utilization program to optimize transfusion practice. *Hematology Am Soc Hematol Educ Program*. 2015;2015:444–447.
29. Ivers N, Jamtvedt G, Flottorp S, et al. Audit and feedback: effects on professional practice and healthcare outcomes. *Cochrane Database Syst Rev.* 2012:Cd000259.
30. Kohli E, Ptak J, Smith R, Taylor E, Talbot EA, Kirkland KB. Variability in the Hawthorne effect with regard to hand hygiene performance in high- and low-performing inpatient care units. *Infect Control Hosp Epidemiol*. 2009;30:222–225.
31. Gould DJ, Creedon S, Jeanes A, Drey NS, Chudleigh J, Moralejo D. Impact of observing hand hygiene in practice and research: a methodological reconsideration. *J Hosp Infect*. 2017;95:169–174.
32. Fortenberry JL, McGoldrick PJ. Internal marketing: a pathway for healthcare facilities to improve the patient experience. *Int J Health Plann Manage*. 2016;9:28–33.
33. McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P. The Hawthorne effect: a randomised, controlled trial. *BMC Med Res Methodol*. 2007;7:30.
34. Thiese MS. Observational and interventional study design types; an overview. *Biochem Med (Zagreb)*. 2014;24:199–210.
35. Stanley K. Evaluation of randomized controlled trials. *Circulation*. 2007;115:1819–1822.
36. Dumville JC, Torgerson DJ, Hewitt CE. Reporting attrition in randomised controlled trials. *BMJ*. 2006;332:969–971.
37. Schulz KF, Grimes DA. Sample size slippages in randomised trials: exclusions and the lost and wayward. *Lancet*. 2002;359:781–785.
38. Sessler DI, Devereaux PJ. Emerging trends in clinical trial design. *Anesth Analg*. 2013;116:258–261.
39. Devereaux PJ, Chan MT, Eisenach J, Schricker T, Sessler DI. The need for large clinical studies in perioperative medicine. *Anesthesiology*. 2012;116:1169–1175.
40. Alexander L, Lopes B, Ricchetti-Masterson K, Yeatts K. Sources of systematic error or bias: information bias. *Epidemiol Res Inf Center (ERIC) Notebook*. 2015;2:1–5. Available at: https://sph.unc.edu/files/2015/07/nciph_ERIC14.pdf. Accessed May 21, 2017.
41. Rothman KJ, Greenland S, Lash TL. Validity in epidemiologic studies. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. Revised 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012:128–147.
42. Greenland S, Lash TL. Bias analysis. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. Revised 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012:345–380.
43. Lash TL, Fox MP, Fink AK. *Applying Quantitative Bias Analysis to Epidemiologic Data*. London, UK: Springer; 2009.
44. Kleinbaum DG, Sullivan KM, Barker ND. Information bias. In: *ActivEpi Companion Textbook: A Supplement for Use with the ActivEpi CD-ROM*. New York, NY: Springer; 2013.
45. Funk MJ, Landi SN. Misclassification in administrative claims data: quantifying the impact on treatment effect estimates. *Curr Epidemiol Rep*. 2014;1:175–185.
46. Duan R, Cao M, Wu Y, et al. An empirical study for impacts of measurement errors on EHR based association studies. *AMIA Annu Symp Proc*. 2016;2016:1764–1773.
47. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol*. 1999;149:981–983.
48. Kahan BC, Rehal S, Cro S. Risk of selection bias in randomised trials. *Trials*. 2015;16:405.
49. Berger VW. Testing for baseline balance: can we finally get it right? *J Clin Epidemiol*. 2010;63:939–940.
50. Schulz KF, Grimes DA. Unequal group sizes in randomised trials: guarding against guessing. *Lancet*. 2002;359:966–970.
51. Schulz KF, Grimes DA. Generation of allocation sequences in randomised trials: chance, not choice. *Lancet*. 2002;359:515–519.
52. Hewitt CE, Torgerson DJ. Is restricted randomisation necessary? *BMJ*. 2006;332:1506–1508.
53. Kahan BC, Morris TP. Reporting and analysis of trials using stratified randomisation in leading medical journals: review and reanalysis. *BMJ*. 2012;345:e5840.
54. Zhao W. Selection bias, allocation concealment and randomization design in clinical trials. *Contemp Clin Trials*. 2013;36:263–265.
55. Rosenberger WF, Lachin JM. *Randomization in Clinical Trials: Theory and Practice*. 2nd ed. Hoboken, NJ: John Wiley & Sons, Inc; 2016.
56. Scales DC, Adhikari NK. Maintaining allocation concealment: following your SNOSE. *J Crit Care*. 2005;20:191–193.
57. Doig GS, Simpson F. Randomization and allocation concealment: a practical guide for researchers. *J Crit Care*. 2005;20:187–191.
58. Fitzmaurice G. Confused by confounding? *Nutrition*. 2003;19:189–191.

59. Brookhart MA, Stürmer T, Glynn RJ, Rassen J, Schneeweiss S. Confounding control in healthcare database research: challenges and potential approaches. *Med Care*. 2010;48:S114–S120.
60. Perlas A, Chan VW, Beattie S. Anesthesia technique and mortality after total hip or knee arthroplasty: a retrospective, propensity score-matched cohort study. *Anesthesiology*. 2016;125:724–731.
61. Helwani MA, Avidan MS, Ben Abdallah A, et al. Effects of regional versus general anesthesia on outcomes after total hip arthroplasty: a retrospective propensity-matched cohort study. *J Bone Joint Surg Am*. 2015;97:186–193.
62. Koch CG, Li L, Duncan AI, et al. Morbidity and mortality risk associated with red blood cell and blood-component transfusion in isolated coronary artery bypass grafting. *Crit Care Med*. 2006;34:1608–1616.
63. Bauman AE, Sallis JF, Dzewaltowski DA, Owen N. Toward a better understanding of the influences on physical activity: the role of determinants, correlates, causal variables, mediators, moderators, and confounders. *Am J Prev Med*. 2002;23:5–14.
64. Richiardi L, Bellocco R, Zugna D. Mediation analysis in epidemiology: methods, interpretation and bias. *Int J Epidemiol*. 2013;42:1511–1519.
65. VanderWeele TJ. Mediation analysis: a practitioner's guide. *Annu Rev Public Health*. 2016;37:17–32.
66. Mascha EJ, Dalton JE, Kurz A, Saager L. Statistical grand rounds: understanding the mechanism: mediation analysis in randomized and nonrandomized studies. *Anesth Analg*. 2013;117:980–994.
67. Altman DG. Comparability of randomised groups. *Statistician*. 1985:125–136.
68. Fitzmaurice G. Adjusting for confounding. *Nutrition*. 2004;20:594–596.
69. Fitzmaurice G. Confounding: regression adjustment. *Nutrition*. 2006;22:581–583.
70. Maciejewski ML, Diehr P, Smith MA, Hebert P. Common methodological terms in health services research and their synonyms [correction of symptoms]. *Med Care*. 2002;40:477–484.
71. Starks H, Diehr P, Curtis JR. The challenge of selection bias and confounding in palliative care research. *J Palliat Med*. 2009;12:181–187.
72. Kyriacou DN, Lewis RJ. Confounding by indication in clinical research. *JAMA*. 2016;316:1818–1819.
73. Penrod JD, Goldstein NE, Deb P. When and how to use instrumental variables in palliative care research. *J Palliat Med*. 2009;12:471–474.
74. Tolles J, Meurer WJ. Logistic regression: relating patient characteristics to outcomes. *JAMA*. 2016;316:533–534.
75. Haukoos JS, Lewis RJ. The propensity score. *JAMA*. 2015;314:1637–1638.
76. Fitzmaurice G. Confounding: propensity score adjustment. *Nutrition*. 2006;22:1214–1216.
77. Rothman KJ. Using regression models in epidemiologic analysis. In: *Epidemiology: An Introduction*. 2nd ed. Oxford, UK: Oxford University Press; 2012:211–234.
78. Fitzmaurice G. The meaning and interpretation of interaction. *Nutrition*. 2000;16:313–314.
79. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods*. 2014;3:33–72.
80. Pearce N, Greenland S. Confounding and interaction. In: Ahrens W, Pigeot I, eds. *Handbook of Epidemiology*. Berlin, Heidelberg, Germany: Springer Berlin Heidelberg; 2005:371–397.
81. Greenland S, Lash TL, Rothman KJ. Concepts of interaction. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2012:71–86.
82. Rothman KJ. Measuring interactions. In: *Epidemiology: An Introduction*. 2nd ed. Oxford, UK: University Press; 2012:198–210.
83. Fitzmaurice G. How to explain an interaction. *Nutrition*. 2001;17:170–171.
84. Yusuf S, Wittes J, Probstfield J, Tyroler HA. Analysis and interpretation of treatment effects in subgroups of patients in randomized clinical trials. *JAMA*. 1991;266:93–98.
85. Gail M, Simon R. Testing for qualitative interactions between treatment effects and patient subsets. *Biometrics*. 1985;41:361–372.
86. Sun X, Ioannidis JP, Agoritsas T, Alba AC, Guyatt G. How to use a subgroup analysis: users' guide to the medical literature. *JAMA*. 2014;311:405–411.