

Difficulties in Demonstrating Superiority of an Antibiotic for Multidrug-Resistant Bacteria in Nonrandomized Studies

Kristen A. Stafford,¹ Mallory Boutin,¹ Scott R. Evans,² and Anthony D. Harris^{1,3}

¹Department of Epidemiology and Public Health, University of Maryland School of Medicine, Baltimore; ²Department of Biostatistics and Center for Biostatistics in AIDS Research, Harvard School of Public Health, Boston, Massachusetts; and ³VA Maryland Health Care System, Baltimore

The discovery and development of new antimicrobials is critically important, especially as multidrug-resistant bacteria continue to emerge. Little has been written about the epidemiological issues in nonrandomized trials aiming to evaluate the superiority of one antibiotic over another. In this manuscript, we outline some of the methodological difficulties in demonstrating superiority and discuss potential approaches to these problems. Many of the difficulties arise due to confounding by indication, which we define and explain. Epidemiological methods including restriction, matching, stratification, multivariable regression, propensity scores, and instrumental variables are discussed.

Keywords. antibiotics; epidemiology; nonrandomized studies; research design; superiority.

There is a critical need for the development of new antibiotics, particularly for antibiotics effective against antibiotic-resistant bacteria. Much has been written about the costs and difficulties of new antibiotic development [1].

Epidemiological issues relative to demonstrating superiority in randomized trials (eg, phase 3) have been discussed, especially as most of these antibiotic trials are noninferiority trials [2]. However, the epidemiological issues associated with nonrandomized, for example, retrospective database trials aiming to demonstrate superiority of one antibiotic over another antibiotic are less well documented [3].

Herein we briefly describe the epidemiological challenges of randomized superiority trials. We then describe epidemiological issues in studies aiming to demonstrate the superiority of antibiotics in nonrandomized retrospective database trials. Issues discussed include confounding

by indication and the epidemiological methods (restriction, matching, stratification, multivariable regression, propensity scores and instrumental variables) aimed at controlling for confounding by indication.

Throughout the manuscript, we use an example of a company developing a novel antibiotic targeted for methicillin-resistant *Staphylococcus aureus* (MRSA) treatment to illustrate the epidemiological concepts. We will outline the study design and analysis issues that arise when trying to demonstrate superiority of the new antibiotic to vancomycin in randomized trials and nonrandomized retrospective database studies. Although this example is used throughout the article, the area of application is more general and can be applied to other antibiotic-resistant bacteria. Additional examples would be a company developing a novel antibiotic targeted for extremely drug-resistant, gram-negative bacteria susceptible only to colistin or the use of a novel antibiotic for the treatment of *Clostridium difficile* either in the initial infection or for recurrent *C. difficile* infection.

EPIDEMIOLOGICAL ISSUES IN RANDOMIZED TRIALS

Generally, randomized trials are superior to nonrandomized observational studies as randomization is

Received 12 February 2014; accepted 18 June 2014; electronically published 30 June 2014.

Correspondence: Anthony Harris, MD, MPH, Division of Genomic Epidemiology and Clinical Outcomes, Department of Epidemiology and Public Health, University of Maryland School of Medicine, 685 W Baltimore St, MSTF 330, Baltimore, MD 21201 (aharris@epi.umaryland.edu).

Clinical Infectious Diseases 2014;59(8):1142–7

© The Author 2014. Published by Oxford University Press on behalf of the Infectious Diseases Society of America. All rights reserved. For Permissions, please e-mail: journals.permissions@oup.com.

DOI: 10.1093/cid/ciu486

the foundation for statistical inference. However, even randomized antibiotic trials have a number of complicated issues. The first difficulty in demonstrating superiority of an antibiotic lies in the outcome definition. Although there are distinct, specific outcomes required by the US Food and Drug Administration (FDA) for each type of infection (eg, catheter-related bloodstream infection), almost all of the antibiotic randomized controlled trials (RCTs) for antibiotics boil down to a binary outcome: infection clearance (yes/no) [4, 5]. Due to the high overall clearance rates for existing antibiotics, especially against the most common susceptible organisms, there is little room for improvement to demonstrate superiority of a novel antibiotic using this outcome even if the antibiotic has superiority in other clinically relevant outcomes [2]. Although the new antibiotic in question may have fewer side effects, increased tolerability, or a simpler dosing regimen, or may be superior in some other way, these are not part of the current FDA-regulated primary outcome definitions. This is why some have argued for more common use of composite outcomes in superiority trials for antibiotics [2].

An additional challenge is enrollment of the most informative patient population. To demonstrate the superiority of a new drug over existing therapies when treating resistant organisms, the ideal patient population consists of individuals infected with antibiotic-resistant organisms. Despite the fact that rates of these incredibly difficult-to-treat pathogens are growing, it is difficult to find and enroll a sufficient number of these patients to populate a comparative trial [1]. This is especially challenging as federal regulators require RCTs to be limited by infection site/type [1].

Often, a physician will immediately begin empiric therapy before the laboratory tests reveal an antibiotic-resistant bacteria and will prescribe >1 empiric antibiotic to cover the full range of the most likely organisms. This practice of prescribing multiple antibiotics empirically prior to culture results, along with the potential for simultaneous infection with multiple organisms, makes it even more difficult to determine if treatment outcomes are a result of the new antibiotic or if they are a result of concomitant antibiotic use [6]. These problems have led authors to suggest alternate designs, including the nested superiority-noninferiority trial based on culture results [2]. Finally, new antibiotics are often studied for clinical indications for which they are seldom used or needed, leaving significant gaps in the literature for situations in which clinicians often rely on new antibiotics.

In light of the difficulties associated with RCTs focused on demonstrating superiority, many antibiotics are now brought to market using clinical trials that aim to show that the agent in question is no worse than existing therapies by an acceptable margin (ie, noninferiority trials) [7]. There are numerous issues in noninferiority trials that have been well outlined by other authors [8–11].

EPIDEMIOLOGICAL ISSUES IN NONRANDOMIZED RETROSPECTIVE DATABASE STUDIES

Confounding by Indication

In clinical practice, clinicians prescribe antibiotic treatment based on the diagnostic factors related to the disease being treated (eg, culture results, white blood cell count, creatinine), as well as the prognostic factors of a particular patient (eg, severity of illness, comorbid conditions) [12]. These factors alter physician prescribing for each clinical situation. The specific clinical situation and prescribing patterns are often dependent on known factors as well as unknown or unmeasured factors (ie, factors that are not recorded in the medical record or clinical database). This type of prescribing is a standard part of good medical practice, and is something patients want from their providers. Although this is a good thing for patients, it also presents significant challenges for researchers attempting to determine whether one antibiotic is better than another when retrospectively using data from clinical databases.

When the factors that influence the clinician to choose a particular drug are also independently associated with the outcome under study, a failure to control for these factors can lead to a confounding of the true association between the agent prescribed and the outcome [13]. This type of confounding is defined by epidemiologists as confounding by indication [14]. Controlling for severity of illness, type of diagnosis, and patient comorbid conditions is difficult to do, and controlling for other *unmeasured* confounders is almost impossible. Figure 1 demonstrates the mechanism of confounding by indication. The inability to control for all of the known factors in the figure and/or the unmeasured factors can lead to confounding by indication.

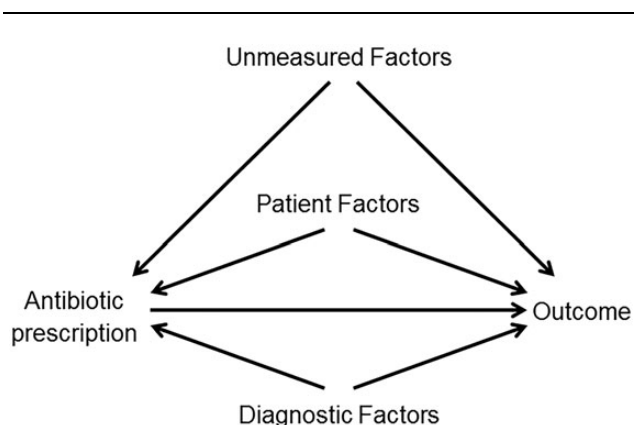


Figure 1. Causal diagram demonstrating the mechanism of confounding by indication in observational studies. The unmeasured factors, patient factors, and diagnostic factors represent the potential confounders in this diagram.

Consider our example: a company is developing a novel antibiotic targeted for MRSA treatment. The first challenge one would face in attempting to evaluate superiority in a retrospective clinical database is that the new drug is not being used randomly for every patient with MRSA; in other words, not every patient with a MRSA infection or possible MRSA infection would randomly receive the new antibiotic vs the existing antibiotic (vancomycin). Instead, clinicians are taking into account all that they know about the diagnostic factors (eg, culture results, creatinine), and prognostic factors (eg, severity of illness and comorbid conditions of patient) when choosing whether to prescribe the new anti-MRSA antibiotic or vancomycin. In most cases, clinicians will be more likely to use older antibiotics for healthier patients and reserve the new antibiotic for patients with more aggressive infections or in patients with significant comorbid conditions. For example, a clinician may be reluctant to prescribe vancomycin to treat MRSA in a patient with elevated creatinine or a patient with severe sepsis and potential subsequent renal failure due to sepsis. Additionally, for severe MRSA infections, a clinician may be concerned about the potential suboptimal efficacy of vancomycin and thus use a new antibiotic. Therefore, when a researcher tries to compare outcomes from patients who received the new antibiotic with those from patients who received vancomycin, it could appear that the new antibiotic is strongly associated with a poor outcome such as kidney disease when, in truth, patients with underlying kidney disease were simply more likely to receive the new drug. Because the new antibiotic is often used in the sicker patient population, the new antibiotic may appear to be equivalent to or inferior to vancomycin in terms of patient outcomes such as mortality, morbidity including clinical cure, and hospital length of stay because the sicker patients are receiving the new antibiotic. However, this finding would be confounded with the difference in clinical scenarios and patients for whom the new antibiotic was prescribed vs the distinct clinical scenarios and patients for whom vancomycin was prescribed. The potential superiority of the new antibiotic may be masked by confounding by indication. As electronic medical records advance, and as software for data retrieval from free text fields improve, particular emphasis should be placed on the ability to collect and retrieve variables useful for this type of research.

Methods to Control for Confounding by Indication

It is advisable to use strategies that try to make the 2 groups more similar before comparing them and more similar to what would have happened in a randomized trial—that is, try to make it similar to the situation that would arise if a patient with an MRSA infection was randomized to receive either vancomycin or the new antibiotic. Several methods are used in observational epidemiologic research that help to reduce confounding when attempting to compare one drug to another;

we discuss these below. However, these methods may not fully resolve the issue of confounding by indication when attempting to demonstrate superiority of a new antibiotic.

Restriction

The principle of restriction is to try to make the groups being compared more homogeneous with respect to measured factors, thereby removing the possibility of confounding by the measure to which you restricted [15]. In our MRSA example, an example of restriction would be to only analyze patients with a creatinine level of <1.0 upon starting antibiotics; that is, one would compare the new antibiotic with patients who received vancomycin only among patients with a creatinine level of <1.0 . Although restriction is a strong tool for factors that are measured, it is unable to control for unmeasured factors [16]. An additional limitation of restriction is that it can dramatically reduce the sample size of the study. One thing investigators may want to consider is whether sample size limitations can be addressed through multicenter collaborations.

Matching

Matching is a process by which the distributions of important measured factors are actively balanced between the 2 comparison groups. In our MRSA example, one could match on a severity of illness score prior to receipt of the antibiotic; this would attempt to allow a comparison of the new antibiotic to vancomycin among patients with a similar severity of illness. In our example, as demonstrated in Figure 1, there may be a large number of confounding variables that are important and it becomes difficult to match on several variables. Additionally, the problem of unmeasured factors/unmeasured confounding remains.

Stratification

Similar to restriction, stratification defines subgroups of patients based on factors that have been measured but without discarding entire strata [12]. For example, you could analyze the association between the new antibiotic and renal insufficiency based on categories of creatinine at the time of prescription (eg, compare the new antibiotic to the old antibiotic in the stratum of creatinine of <1 , 1.0 – 2.0 , etc). A weighted measure of effect can then be calculated allowing one to compare the new antibiotic to the old antibiotic while, in our example, controlling for creatinine baseline level. Assuming that all confounding factors were measured, stratified estimates can be unbiased. However, assuming that *all* relevant factors are known and measured is a very strong assumption [13]. Although stratification can be a reasonable option when there are sufficient patients for subgroup analyses, strata with small numbers of subjects can cause challenges, such as reduced power to detect a clinically relevant difference. Rothwell describes in detail the design, analysis, and interpretation of subgroups [17]. With respect to our specific

question of showing superiority of a new antibiotic in the presence of confounding by indication, stratification will not fully control for confounding due to unmeasured confounding factors that may influence both antibiotic selection (exposure group) and outcome [16].

MULTIVARIABLE REGRESSION

Regression analysis allows for an easier approach to a stratified analysis [12], particularly when there are multiple confounding factors that you want or need to include in your model to estimate a treatment effect. There are several different types of multivariable regression. Most commonly seen in the literature are regression methods which assume that a parametric distribution fits the data [12]. There are also semiparametric and nonparametric regression methods that can relax the parametric assumptions associated with the models described above and are described elsewhere [18–20]. A basic rule of thumb when creating multivariable regression models is that you should have at least 10 events for each covariate you include in your model [21, 22]. Multivariable regression methods are commonly used to answer questions from databases, and are fairly easy to perform given the availability of statistical software and computing power. However, they do not alone address the issue of unmeasured confounding. Unless all of the factors that are associated with antibiotic selection and are also predictors of the outcome are measured, documented, and accounted for in your regression model, your desired measure of association between the new antibiotic and your outcome will be confounded. Multivariable studies using data that fail to account for the prognostic differences between patients at the time of prescription may not yield valid results [23].

PROPNESITY SCORES

A propensity score is a patient's probability of being exposed to a particular treatment as a function of all of the patient factors that were measured before treatment began that are potentially associated with outcomes [23, 24]. In our case, a propensity score would be the probability of receiving the new antibiotic instead of vancomycin based on all of the pretreatment factors measured in a clinical database. Propensity scores are essentially a summary score of the measured confounders that are created by modeling the probability of exposure as the outcome in a regression model [20]. Propensity scores have become a popular method to try and address residual confounding that remains when there is incomplete or imperfect adjustment of factors in the model and unmeasured confounding [13, 25]; however, they still present significant limitations in the presence of confounding by indication.

Propensity scores are a tool to balance the measured factors across exposure (eg, antibiotic) classification. They can be

particularly useful when you have a large number of factors you need to control for, but very few events in your sample. By using the factors to develop the propensity score and using that summary score in your regression, you can gain statistical efficiency that you would lose by using each confounding factor individually in a model, for example, via multivariable regression. It is important to check that covariate balance has been achieved after propensity scores have been incorporated, to ensure control for potential confounding and the resulting validity of the results.

You can also use propensity scores to assess the amount of unmeasured confounding that may be present in your treatment effect estimate by comparing the overlap of propensity scores across antibiotic groups. If there is not a lot of overlap in the propensity scores between your 2 groups, then it is likely that there are still factors that are unbalanced—for example, your groups are very different in ways other than the treatment they received and, thus, treatment effect estimates may be biased by unmeasured confounding.

Propensity scores can be used in several ways to make treatment groups more balanced with respect to confounders. Some of the ways propensity scores are used include matching on propensity scores, subclassification of scores, weighting, or using the propensity score in regression [24, 26]. These propensity score methods are sufficient to remove bias in your treatment effect estimate based on measured factors; however, they assume that treatment assignment is ignorable [24]. In the case of confounding by indication, we cannot assume that the measured factors that went into the development of the propensity score capture all of the relevant factors that could confound our association. Although propensity scores were not developed to address unmeasured confounding, they are, unfortunately, increasingly used in this capacity [19, 24, 27]. A useful commentary is available on the principles of modeling propensity scores in medical research and discusses propensity scores in the context of actual practice [28]. Although propensity scores can be a useful tool, they are not sufficient to fully adjust for confounding by indication even when documented prognostic factors associated with treatment assignment are included, and often yield similar results to multivariable regression in such instances [16, 26, 29]. In our example, it is unlikely that all of the diagnostic and prognostic factors that went into a clinician's choice of new antibiotic vs vancomycin will be recorded in the database. Therefore, using propensity scores may still yield biased results and not get us any closer to true superiority evaluation, thus limiting the value of propensity scores.

INSTRUMENTAL VARIABLES

Instrumental variables methodology has been around since the 1920s and is an important tool in the field of econometrics, but has only recently gained popularity in epidemiologic studies [30]. Instrumental variables are designed to produce an

Table 1. Summary of Strengths and Weaknesses of Methods to Address Confounding by Indication^a

Method	Strengths	Weaknesses
Restriction	Easy to do, easy to understand	Can limit power
Matching	Easy to do, easy to understand	Difficult to match on a large number of variables
Stratification	More powerful than restriction and matching	Difficult to deal with strata with small numbers. Can introduce residual confounding if subjects with very different values of a factor are grouped into a strata (eg, quartiles)
Multivariable regression	More powerful than restriction and matching	Model building can be complicated Can introduce residual confounding if subjects with very different values of a factor are grouped into a strata (eg, quartiles)
Propensity scores	Can be used to assess amount of unmeasured confounding that exists	Not designed specifically to deal with unmeasured confounding
Instrumental variables	One of the newer, more powerful methods	Difficulty in finding the optimal instrumental variable

^a All methods cannot control for unmeasured confounding.

unbiased estimate of causal effects in observational studies by grouping patients by their likelihood of exposure to a particular treatment, thus simulating the random assignment aspect of an RCT [12]. An instrumental variable is a factor that is associated with treatment, unrelated to measured and unmeasured patient factors, and unrelated to the outcomes under study except through its association with the treatment [31]. In an RCT, the instrumental variable is the original random assignment under intention-to-treat analysis [31–33]. In observational studies, finding an instrumental variable that meets all of the validity criteria necessary would be highly advantageous. However, many of the assumptions underlying a valid instrumental variable are not empirically verifiable [31]; there really is not a way to check and see if it meets the assumptions. This is particularly important to consider if the exposure of interest is time variable (as antibiotic administration can often be), in which case standard instrumental variable methods are not well equipped to control for the unmeasured confounding [31].

In pharmacoepidemiology, provider prescription preference has been investigated as a potential instrumental variable using observational data. An example of provider prescription preference is the use of a provider's last prescription as an

instrumental variable. Although this instrumental variable appears promising in some cases, it may still not address the confounding by indication we need to control to evaluate superiority of a new antibiotic. For example, we cannot come up with an instrumental variable that is associated with antibiotic prescribing but is not associated with the outcome or other confounding variables. Provider prescription preference is likely to be associated with patient prognostic factors that are also predictive of outcomes, whether or not they are measured, and thus violates the definition of an instrumental variable. Using an imperfect instrumental variable can bias the estimate in unpredictable directions (either toward or away from superiority) [31]. The lack of the ability to test the assumption that the instrumental variable is only associated with the outcome through the exposure to a particular drug means that we will not know if we have actually estimated the true treatment effect. Although this should not discourage the investigation of the use of instrumental variables, even instrumental variables do not fully overcome confounding by indication.

CONCLUSIONS

The development and discovery of new antimicrobials is critical, especially as multidrug-resistant bacteria continue to emerge. However, companies launching new antibiotics face difficult challenges in demonstrating superiority. In this manuscript, we have outlined some of these methodological difficulties and discussed methods to address these problems. Table 1 gives a summary of the methods to address confounding by indication. Many of the remedies are, unfortunately, only partial solutions, and the method researchers will need to use will depend on a variety of factors related to one's question, the study design, and the available data. When performing studies to evaluate superiority, collaboration between investigators, epidemiologists, and statisticians is critical.

Notes

Disclaimer. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Financial support. This work was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health (award number UM1AI104681; grant number K24AI079040).

Potential conflicts of interest. All authors: No reported conflicts. All authors have submitted the ICMJE Form for Disclosure of Potential Conflicts of Interest. Conflicts that the editors consider relevant to the content of the manuscript have been disclosed.

References

1. Boucher HW, Talbot GH, Bradley JS, et al. Bad bugs, no drugs: no ESKAPE! An update from the Infectious Diseases Society of America. *Clin Infect Dis* 2009; 48:1–12.
2. Infectious Diseases Society of America. White paper: recommendations on the conduct of superiority and organism-specific clinical trials of

- antibacterial agents for the treatment of infections caused by drug-resistant bacterial pathogens. *Clin Infect Dis* **2012**; 55:1031–46.
3. National Institutes of Health. FAQ: ClinicalTrials.gov—clinical trial phases. Available at: <http://www.nlm.nih.gov/services/ctphases.html>. Accessed 7 January 2014.
 4. Food and Drug Administration. Guidance for industry: complicated urinary tract infection—developing drugs for treatment (draft guidance), **2012**. Available at: <http://www.fda.gov/downloads/Drugs/Guidances/ucm070981.pdf>. Accessed 3 July 2014.
 5. Food and Drug Administration. Guidance for industry: catheter related bloodstream infections—developing antimicrobial drugs for treatment (draft guidance), **1999**. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/ucm070972.pdf>. Accessed 3 July 2014.
 6. Food and Drug Administration. Guidance for industry: antibacterial therapies for patients with unmet medical need for the treatment of serious bacterial diseases (draft guidance), **2013**. Available at: <http://www.fda.gov/downloads/Drugs/GuidanceComplianceRegulatoryInformation/Guidances/UCM359184.pdf>. Accessed 3 July 2014.
 7. Food and Drug Administration. Guidance for industry: non-inferiority clinical trials (draft guidance), **2010**. Available at: <http://www.fda.gov/downloads/Drugs/Guidances/UCM202140.pdf>. Accessed 3 July 2014.
 8. Kaul S, Diamond GA. Good enough: a primer on the analysis and interpretation of noninferiority trials. *Ann Intern Med* **2006**; 145:62–9.
 9. Powers JH, Cooper CK, Lin D, Ross DB. Sample size and the ethics of non-inferiority trials. *Lancet* **2005**; 366:24–5.
 10. Evans S. Noninferiority clinical trials. *Chance* **2009**; 22:53–8.
 11. Snapinn SM. Noninferiority trials. *Curr Control Trials Cardiovasc Med* **2000**; 1:19–21.
 12. Schneeweiss S. Developments in post-marketing comparative effectiveness research. *Clin Pharmacol Ther* **2007**; 82:143–56.
 13. Rothman KJ, Greenland S, Lash TL. *Modern epidemiology*. Philadelphia, PA: Lippincott Williams & Wilkins, **2008**.
 14. Walker AM. Confounding by indication. *Epidemiology* **1996**; 7:335–6.
 15. Greenland S, Morgenstern H. Confounding in health research. *Annu Rev Public Health* **2001**; 22:189–212.
 16. Bosco JLF, Silliman RA, Thwin SS, et al. A most stubborn bias: no adjustment method fully resolves confounding by indication in observational studies. *J Clin Epidemiol* **2010**; 63:64–74.
 17. Rothwell PM. Treating individuals 2. Subgroup analysis in randomised controlled trials: importance, indications, and interpretation. *Lancet* **2005**; 365:176–86.
 18. Fan J. Design-adaptive nonparametric regression. *J Am Statist Assoc* **1992**; 87:998–1004.
 19. Hosmer DW, Lemeshow H, May S. *Applied survival analysis: regression modeling of time-to-event data*. Hoboken, NJ: John Wiley & Sons, Inc, **2008**.
 20. Robins JM, Hernan MA, Brumback B. Marginal structural models and causal inference in epidemiology. *Epidemiology* **2000**; 11:550–60.
 21. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. *J Clin Epidemiol* **1996**; 49:1373–9.
 22. van Belle G. *Statistical rules of thumb*. New York, NY: Wiley, **2002**.
 23. Rubin DB. Propensity score methods. *Am J Ophthalmol* **2010**; 149: 7–9.
 24. Rosenbaum PR, Rubin DB. The central role of the propensity score in observational studies for causal effects. *Biometrika* **1983**; 70: 41–55.
 25. Sturmer T, Joshi M, Glynn RJ, Avorn J, Rothman KJ, Schneeweiss S. A review of the application of propensity score methods yielded increasing use, advantages in specific settings, but not substantially different estimates compared with conventional multivariable methods. *J Clin Epidemiol* **2006**; 59:437–47.
 26. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of logistic regression versus propensity score when the number of events is low and there are multiple confounders. *Am J Epidemiol* **2003**; 158:280–7.
 27. Poses RM, Smith WR, McClish DK, Anthony M. Controlling for confounding by indication for treatment: are administrative data equivalent to clinical data? *Med Care* **1995**; 33:AS36–46.
 28. Austin PC. An introduction to propensity score methods for reducing the effects of confounding in observational studies. *Multivariate Behav Res* **2011**; 46:399–424.
 29. Austin PC, Grootendorst P, Normand SL, Anderson GM. Conditioning on the propensity score can result in biased estimation of common measures of treatment effect: a Monte Carlo study. *Stat Med* **2007**; 26: 754–68.
 30. Ray WA. Evaluating medication effects outside of clinical trials: new-user designs. *Am J Epidemiol* **2003**; 158:915–20.
 31. Hernan MA, Robins JM. Instruments for causal inference: an epidemiologist's dream? *Epidemiology* **2006**; 17:360–72.
 32. Greenland S. An introduction to instrumental variables for epidemiologists. *Int J Epidemiol* **2000**; 29:722–9.
 33. Dohoo I, Martin W, Stryhn H. *Methods in epidemiologic research*. Prince Edward Island, Canada: VER, Inc, **2012**.