A Theoretical Exploration of Therapeutic Monomania as a Physician-Based Instrumental Variable

Brian J. Potter MDCM SM FRCPC,[1,2]

Colin Dormuth MD MA SM ScD,[3]

&

Jacques Le Lorier MD PhD.[2]

1. Cardiology Service, Department of Medicine, Centre Hospitalier de l'Université de Montréal (CHUM), Montréal, QC, Canada.
2. Pharmacoepidemiology and Pharmacoeconomics Unit, Carrefour d'innovation et évaluation en santé (CIÉS), Centre de Recherche du CHUM (CRCHUM), Montréal, QC, Canada.
3. Department of Anesthesiology, Pharmacology, and Therapeutics, Faculty of Medicine, University of British Columbia, Vancouver, BC, Canada.

Author of Correspondence:

Jacques Le Lorier MD PhD
Pharmacoepidemiology and Pharmacoeconomics Unit,
Centre de Recherche du CHUM (CRCHUM),
800 rue St-Denis
Montréal, QC, Canada.
H2X 0A9
jacques.le.lorier@umontreal.ca

Running Title: Therapeutic Monomania as an Instrumental Variable

(49 characters with spaces max 50))

ABSTRACT

In attributing causality, randomised controlled trials (RCTs) are considered the gold standard for their ability to address both known and unknown confounders. When RCTs are not possible, however, researchers must attempt to account for confounding by other means. First described by Wright in 1926, instrumental variable analysis remains the only well-validated method directly addressing unknown confounders in the analysis phase of a study, but identifying a suitable instrumental variable can be challenging. Herein we review the criteria for an optimal instrumental variable and present arguments in support of individual physician and physician group prescribing preferences as one such optimal instrument variable, with suggestions for how to approach specific scenarios that may be encountered.

"But, in his narrow-flowing monomania,... That was living agent, now became the living instrument."

-- Herman Melville, *Moby-Dick*

In attributing causality, randomised controlled trials (RCTs) are considered the gold standard for their ability to address both known and unknown confounders. When RCTs are not possible, however – be it for ethical, practical, or financial reasons – researchers must attempt to account for confounding by other means. In the design phase, it has been suggested that case-only study designs may better address unknown confounders.[1] Yet, in the analysis phase, nearly all methods, including propensity score matching and marginal structural models, address only known confounders directly. While, it has been suggested that high-dimension propensity score models can indirectly account for at least some unmeasured confounding[2] and the "missing cause" approach and sensitivity analysis show promise,[3,4] the only well-validated method that directly addresses unknown confounders outside of RCTs is instrumental variable (IV) analysis.

Herein, we set forth arguments in support of IV analysis and discuss considerations specific to appropriately using the prescription preferences of individual and groups of providers as an IV, including potential pitfalls. We also describe the theoretical benefit of using PS matching prior to IV analysis to ensure that the equipoise population is clearly ascertained and within which the IV assumptions discuss below are most likely to hold. It is important to note that supporting evidence for these hypotheses is currently lacking, but is the subject of ongoing research. The

recommendations made therefore represent the considered opinion of the authors regarding what might constitute best practices.

The Case for Instrumental Variable Analysis

Indeed, the specter of a dreaded unknown confounder is the Achilles heel of observational research. This is particularly true in pharmacoepidemiology where there is the potential for unmeasured confounding by indication. Confounding by indication (and its corollary, confounding by contraindication) refers to any way in which patients receiving a therapy or intervention might be inherently different from those not receiving the intervention by virtue of having been prescribed a treatment *for a reason*.[5] As opposed to RCTs, in observational data, physicians have selected treatments based on everything they know about a given patient. As everything known to treating physicians about their patients is unlikely to be captured in a database, particularly administrative ones, the potential for unknowable (and therefore potentially unaccounted for) confounding persists.

The concept of IVs was originally described in a letter by Philip G. Wright to his son in 1926 and first published in a 1928 economic treatise, but it wasn't until 1945 that Olav Reiersøl gave the method its current name,[6] and not for another 50 years before it became more widely used in health research, owing in large part to the difficulty in identifying suitable instrumental variables.[7]

The key step in IV analysis is the identification of an optimal *instrument*. In the context of pharmacoepidemiology, an instrument is an observable variable that, by virtue of being correlated with the prescribed drug but not with the outcome, except perhaps through its effect on treatment selection, allows one to use it as a *proxy* for randomization.[7-9] More specifically, the instrument does not itself belong in a regression model as an explanatory variable, but, as it is correlated with

the exposure and the *endogenous error (but not the outcome)* – due to omitted information or measurement error – its inclusion can reduce the error of the regression estimate. An instrument is therefore not a confounder, as a confounder is a variable that directly affects both the intervention and outcome. To be a good instrument, a variable must satisfy three criteria: (1) It must predict the actual treatment received (Strength of the Instrument); (2) it cannot influence the outcome other that through the treatment received (Exclusion Restriction); and (3) it cannot be influenced by any factor that also influences the outcome (Independence Assumption).[7-9] These concepts are described schematically in Figure 1. Ideally, then, an IV would also be unrelated to both known and unknown patient characteristics that might influence treatment.[10]

Prescribing Preference Estimates as an Instrumental Variable

As early as 2003, authors have posited that provider prescribing preferences are, at least in some cases, a stronger predictor of the prescribed drug than the patient's characteristics.{Brookhart, 2006 #10;Brookhart, 2007 #9;Schneeweiss, 2005 #30;Solomon, 2003 #29} They therefore proposed past prescription behavior as a method of quantifying each care provider's prescription preference (Prescribing Preference Estimate, PPE). Since, several studies have evaluated the performance of specific methods prescription preference estimation. However, results have varied. Henessey et al. reported that the PPE based on the previous prescription is a stronger IV than the one based on alternative sets of prior prescriptions,[11] whereas Ionescu-Ittu et al found the opposite,[12] reporting that the proportion of patients that are prescribed a given drug by a specific physician produces a stronger IV with less variance than relying solely on the last prescription. Reasons for these discrepancies include that physician preference may change over time,[13] that instruments may have variable performance depending on the number of patients seen by a

provider,[8] and that the ideal PPE may depend on the treatment studied,[14] or the timeframe of the effect of interest.[15] The clinical setting may also have an impact on PPE performance as an IV.

Notwithstanding, Davies et al found that past prescription behaviour was less correlated with patient characteristics than current prescriptions and thus should also be less related to unknown confounders, suggesting that PPE may well be, in many instances, an ideal IV. It would therefore be of interest to establish the ideal clinical research conditions that would support using PPE as an IV.

Only comparisons between two therapies or two classes of therapies should be undertaken. Ideally, the market share of each therapy (i.e. rate of prescription) should be reasonably similar between the two therapies and there should not be differential reimbursement restrictions on the therapies studied. In the case of pharmaceuticals, the patent status (generic or brand name) of the two therapies should also be the same and stable over the period of study. (Favoring comparisons of generic medications has the advantage of minimizing the influence of industry marketing initiatives on physician practice patterns. However, those same industry influences may give rise to monomania for branded therapies.) Like other statistical instruments, PPE will have more statistical power for a given instrument strength when the disease to be treated and outcome are common, when patients are seen by physicians who see a large number of patients, and when the outcome of interest can be assessed over a relatively short period. In addition, in line with general pharmacoepidemiological principles, patients would ideally be "treatment naïve" for the same or a similar condition. Practically, this implies a reasonable time exclusion based on the known duration of treatment effects and the natural history of the illness. Finally, appropriate restrictions of the population under study based on clinical knowledge are essential to ensuring that patients unlikely to be treated with one or either therapy are excluded.

Individual Providers: The Preference-Monomania Spectrum

Physicians, like anyone, are creatures of habit. These habits might change over time[13] depending on the evolution of the scientific literature, clinical practice guidelines, or pharmaceutical industry marketing practices. However, particularly for specific patient populations and over shorter time periods, individual prescribing preferences may be highly stable. We have termed such high levels of prescription stability *therapeutic monomania*. In a recent informal survey of a convenience sample of Québec general practitioners (JLL), 82% reported always using the same drug when they saw new patients with the following uncomplicated conditions: lower urinary tract infections in women, upper and lower respiratory tract infections, first therapy in essential hypertension, first therapy in light to moderate asthma, and first therapy in type II diabetes.

In contrast to a physician who would always prescribe the same first line therapy for all patients with diabetes, for example, we would describe a physician who prescribes the same first line therapy 82% of the time as having a strong *preference,* or imperfect monomania. It also stands to reason that, when available, strict monomania is likely to be a stronger IV than preference. However, the stronger the preference (i.e. the closer to monomania), the better PPE is likely to perform as an IV.

To this point, PPEs of general practitioners might be better IVs than those of specialists who may more inclined to individualize therapy. In contrast to general practitioners, who might prefer the efficiency of "routine" first-line therapies for certain diseases, specialists are more likely to see patients who have either failed previous therapy, are difficult to control, or are otherwise more complex or at higher risk of events.

Moreover, it is possible that not all general practitioner practices are suitable for use as IVs. General practitioners, too, sometimes orient their practices to certain diseases. Clustering of high-risk cases would violate the third assumption of IV methods.[8] Furthermore, as patients' lack of knowledge of a doctor's prescribing preference is a prerequisite for using PPEs as an IV, the possibility of "doctor shopping" would also violate the third assumption.[8] It may therefore be necessary to restrict analyses to relatively homogenous groups of physicians.[16] We therefore suggest that busy walk-in clinics or emergency rooms are likely the ideal scenario for using individual physician PPEs as IVs.

As the goal of IV analysis is to identify a variable that will most closely approximate the coin toss of randomization, consider a patient with symptoms of sinusitis who presents to an emergency outpatient clinic for evaluation and treatment. By the nature of the clinic, the patient has no assigned physician. He takes a number and waits. Physicians on duty may well have pre-established preferences for first-line treatment of sinusitis, but the patient is unaware and unable to choose by whom he is seen. His number is eventually called and he is assigned an examination room with a physician. It is the luck of the draw. This "natural" situation is perhaps as close to randomization as we will ever get in an observational setting. In this scenario, PPEs are likely to be strong and valid IVs.

It is important to bear in mind that the optimal method for establishing prescriber preference when using PPEs as an IV is a matter of debate and may change from one study to another for the reasons outlined herein. We suggest that using a relatively rigorous definition of monomania, such as uniformity of the two previous prescriptions with the index prescription, is a reasonable starting point for defining the PPE. It is, however, critical to assess different PPEs in a sensitivity analysis

of the IV definition in all studies, such as varying the number of identical prescriptions necessary for assigning monomania (i.e. 4 vs. 3 vs. 2 prior prescriptions).

Groups of Providers: Strength in Numbers

Analogous to the situation with monomaniac prescribers, groups of providers may have similar prescribing behaviour within their practice group, but that may differ from the equally homogenous practices of other groups. Indeed, formal practice standardization can be an efficient way of ensuring consistent quality care. However, informal practice standardization may also occur within practice groups, leading to divergent practice "cultures" between groups or institutions, particularly in areas of clinical equipoise. Financial considerations and contractual obligations may also lead to homogenous treatment for certain conditions at different institutions. (This applies particularly to medical devices and other expensive therapies where it might not be economically efficient to offer two seemingly equivalent competitive therapies.) Indeed, Ionescu-Ittu et al found that hospital-based PPE IVs performed better than individual-physician based IVs.[12] However, similar concerns to those with individual providers with regards to case-mix, specialization, and patients actively seeking out treatment at specific centers for perceived or real differences in care deserve careful consideration when selecting physician groups for study.

For argument's sake, let us consider two hospitals (Scenario 1). Hospital A consistently prescribes Treatment X at discharge to all patients presenting with a certain disease. Hospital B, on the other hand, consistently prescribes Treatment Y for the same disease. Each group of prescribers could be considered therapeutic monomaniacs. On the assumption that the patients treated at both hospitals are part of the same population, the situation could be considered analogous to cluster randomization based on the hospital where a given patient seeks treatment. In

this scenario, one can appreciate how IV methodology could also be employed to account for unmeasured confounders of the relationship between Treatment X versus Y and outcomes.

However, while therapeutic monomania in individual prescribers might be difficult to establish, perfectly monomaniac treatment cultures in groups of providers may be even rarer depending on the treatment or pathology studied. Consider Scenario 2, in which Hospital B is replaced by Hospital C where *most* (90%), but not all, patients with the same disease are managed with Treatment Y (Scenario 2).

The reasons for receiving Treatment X at Hospital C may be varied, including a minority of "rogue" practitioners who prefer Treatment X despite the prevailing local culture, a minority of patients deemed unsuitable for Treatment Y by all practitioners, or the spurious influence of pharmaceutical industry marketing at various points in time. It may not be discernible within a dataset which influence lead to this minority treatment population at Hospital C. As such, it may be difficult to determine how best to statistically address this nearly, but imperfectly, monomaniac hospital.

With individual prescribers, one may simply exclude practitioners that do not fit whatever definition of "monomania" one sets. However, the number of practice groups from which one may choose is necessarily a smaller pool than that of individual prescribers and it might not be possible to analyze only perfectly monomaniac hospitals. Moreover, it is possible that perfectly monomaniac hospitals also differ from imperfectly monomaniac centers in unmeasured ways.

One solution to Scenario 2 is to simply ignore the imperfection and apply an instrumental variable methodology as for Scenario 1. Retaining all patients at Hospital C would be analogous to an intention-to-treat analysis of an RCT with imperfect compliance. This solution might lead to a dilution of the treatment effect or, conversely, if the 10% of patients at Hospital C receiving

Treatment X are substantially different, lead to a spurious exaggeration of a treatment effect. The risk of substantial bias appears small to us when there is not more than 10% practice divergence.

Another simple solution is to "transform" Hospital C into a center with a perfectly monomaniac practice, as in Hospital B, by excluding the minority treatment population from the analysis and applying instrumental variable methods. This is analogous to a per-protocol analysis of a prospective trial and may yield a better estimate of the real-world treatment effect. In a per-protocol analysis, only patients receiving the randomly assigned treatment are analyzed and those deviating from the assigned treatment are excluded from the analysis. Exclusion of these patients in an IV setting may well be appropriate if the minority population was substantially different from the majority population. However, there is nonetheless a risk of introducing bias if there is an equivalent unidentified subset of patients in Hospital A that we cannot so easily exclude. Again, we believe that this risk is low if the divergent population is relatively small (≤10%).

However, both solutions (either ignoring or excluding the minority population at Hospital C) might introduce significant bias if the size of the divergent population is moderate or large (Scenarios 3 & 4). Either excluding or ignoring the minority population is unlikely to be an optimal solution in these scenarios. However, foregoing instrumental variable methods opens the door to confounding.

We propose that combining propensity score (PS) and instrumental variable methodologies is a promising solution if used correctly. It is critical to view IV and PS methods as separate and complementary solutions in this scenario to be applied sequentially or in parallel, as it has been shown that embedding an IV in the PS model can lead to increased bias and regressor inconsistency.[17,18] While Myers et al suggest that the introduced bias is minor in comparison to the correction for confounding afforded by instrumental variable adjustment,[19] we propose that a two-

step method of PS matching followed by IV adjustment if appropriate that avoids this potential problem.

PSs are generated by creating a regression model for treatment assignment as opposed to outcome. Matching patients based on their likelihood of treatment with X or Y is then possible.[20] An important principle for propensity matching, however, is that of "common support". That is, for matching to be effective (i.e. both identifying the population of clinical equipoise while preserving sufficient statistical power), there must be a substantial number of patients who received either Treatment X or Y who would have had a statistical likelihood of receiving the other treatment. Conversely, patients with a zero likelihood of receiving Treatment X should not be compared to patients who indeed received treatment X, as they are fundamentally a different population, and similarly for treatment Y.

From this point on, we define the population of interest as the population for whom there is evidence of clinical equipoise in the data set. That is, the *equipoise population*. On the assumption that observed practice (prescription of Treatment X or Treatment Y) represents medically reasonable practice, the exclusion of sub-populations without common support (i.e. the non-overlapping tails of treatment propensity with a zero-likelihood of receiving the other treatment) would leave only the subpopulation for whom clinical equipoise could be established based on the available data. This concept is important in that it suggest that PS matching, in addition to clinically-guided exclusions and restrictions, can help one hone in on the population where true clinical equipoise exists. In Scenario 4 above, if the 40% of patients at Hospital E would have received Treatment X regardless of where they were treated, there is no clinical equipoise in this subpopulation and their exclusion is appropriate. However, without PS methodology, it is difficult to know who at Hospital A should also be excluded for not being part of the population of interest.

While not the typical use of PS methodology, if the two hospital populations are of the same clinical population, common support should exist between them. We could therefore identify the population of patients at Hospital A who could have received Treatment Y had they instead presented to Hospital E and excluding those patients who would have been managed with Treatment X at both hospitals. In this way, we would be appropriately "transforming" both Hospitals A and E into two monomaniac centers with equipoise populations fit for comparison with IV methodology.

Because residual unmeasured confounding is likely even after propensity matching, we recommend then using IV methodology for analysis of treatment in the transformed (matched) monomaniac center populations, provided that near-perfect monomania (≤10% treatment divergence) was obtained. However, IV analysis is likely not appropriate if substantial divergence (i.e. low levels of prescriber preference <80%) persists.

Consider Scenario 5A, in which we compare two non-monomaniac hospitals. It is conceivable that the minority population in Hospital F (20% receiving Treatment Y) would have also all received treatment Y had they presented to Hospital E instead. These patients would therefore not represent the equipoise population and would be excluded by PS matching because of their 100% likelihood of receiving Treatment Y at either center. Similarly, if the same were true of the minority population at Hospital E, they would also appropriately be excluded by propensity matching, generating two "monomaniac" centers with regards to the equipoise population. Propensity score matching can therefore be an effective tool for both restricting the analysis to only the population of interest and for providing a robust instrumental variable to address residual unmeasured confounding.

Alternatively, consider now that only a portion of the minority populations at both centers would have received the same treatment regardless of the hospital to which they presented (Scenario 5B). Those patients would be excluded, as above. Similarly, a portion of the majority populations at either center might also be excluded for the same reason. The remaining majority and minority populations at either center all have a non-zero likelihood of treatment with either therapy and therefore represent the equipoise population of interest.

However, in this scenario, the hospitals are no longer perfectly monomaniac and the appropriateness of using IV methods at this stage is debatable. If the minority population remains substantial at either center after PS transformation, standard multivariable regression (MVR) analysis is probably the most appropriate next step. However, IV analysis might still be appropriate if the residual minority population is small in both centers (≤10-20%), so long as the limitations of ITT or PP type analyses are acknowledged. It is also perhaps indicated to perform sensitivity analyses comparing standard MVR, ITT IV, and PP IV methodologies.

Finally, we strongly recommend using PS methods as a check in all analyses, even when comparing two seemingly monomaniac hospitals, as the appearance of therapeutic monomania may in fact be a by-product of the unique patient populations treated at different centers. For example, if practitioners at Hospital A had instead seen the patient population of Hospital B (Scenario 1), they might have also have prescribed exclusively Treatment Y. PS matching would in such a case be expected to identify a lack of common support between two disparate populations that are therefore inappropriate for comparison and prompting the consideration of other centers for analysis.

Recommendations

Our recommendations for addressing both perfectly and imperfectly monomaniac hospitals are outlined in Table 1 according to the proportion of patients not treated with the primary therapy of the institution. PS matching prior to IV methods may ensure bias minimization in all cases. It should be noted in Scenarios VI and VIII that physician-based IV methods might not be appropriate depending on the residual divergent population in Hospital B. Again, these recommendations represent the opinion of the authors and not all recommendations have been empirically validated. Additionally, comparing PS and IV methods in isolation to one another may also serve as a form of sensitivity analysis for the magnitude of effect of unmeasured confounding.

Conclusions

In conclusion, instrumental variable analysis can be a powerful tool for addressing unmeasured confounding in pharmacoepidemiogic studies and physician-based instrumental variable exploiting strong prescription preferences are particularly promising. Perfect therapeutic monomania (extreme preference, 100%) is perhaps the ultimate IV for both individual and groups of providers. Near-perfect monomania (≥90%) is likely a reasonably strong IVs and low PPEs (<80%) are likely poor IVs. When strict and robust monomania definitions are possible for individual prescribers, they should be preferred. We strongly recommend PS matching prior to IV analysis when studying groups of providers to (1) ensure that there is common support and (2) to attempt to "transform" centers with relatively weak treatment preferences into centers with strong preference (≥80-90%) for the equipoise population of interest. We do not recommend physician-based IV analysis for centers that have lower than 80% treatment preference after PS matching.

Acknowledgements

This publication was supported financially by the Canadian Network for Observational Drug Effect Studies (CNODES), a collaborating centre of the Drug Safety and Effectiveness Network (DSEN), funded by the Canadian Institutes of Health Research (CIHR, Grant #DSE – 146021).

Ethics: NA

References

1. Uddin MJ, Groenwold RH, Ali MS, et al. Methods to control for unmeasured confounding in pharmacoepidemiology: an overview. *Int J Clin Pharm.* 2016;38(3):714-723.
2. Guertin JR, Rahme E, LeLorier J. Performance of the high-dimensional propensity score in adjusting for unmeasured confounders. *Eur J Clin Pharmacol.* 2016;72(12):1497-1505.
3. Abrahamowicz M, Bjerre LM, Beauchamp ME, LeLorier J, Burne R. The missing cause approach to unmeasured confounding in pharmacoepidemiology. *Stat Med.* 2016;35(7):1001-1016.
4. Schwartz S, Li F, Reiter JP. Sensitivity analysis for unmeasured confounding in principal stratification settings with binary variables. *Stat Med.* 2012;31(10):949-962.
5. Salas M, Hofman A, Stricker BH. Confounding by indication: an example of variation in the use of epidemiologic terminology. *Am J Epidemiol.* 1999;149(11):981-983.
6. Stock JH, Trebbi F. Retrospectives: Who invented instrumental variable regression? *Journal of Economic Perspectives.* 2003;17(3):177-194.
7. Brookhart MA, Wang PS, Solomon DH, Schneeweiss S. Instrumental variable analysis of secondary pharmacoepidemiologic data. *Epidemiology.* 2006;17(4):373-374.
8. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables I: instrumental variables exploit natural variation in nonexperimental data to estimate causal relationships. *J Clin Epidemiol.* 2009;62(12):1226-1232.
9. Brookhart MA, Rassen JA, Schneeweiss S. Instrumental variable methods in comparative safety and effectiveness research. *Pharmacoepidemiol Drug Saf.* 2010;19(6):537-554.
10. Ali MS, Uddin MJ, Groenwold RH, et al. Quantitative falsification of instrumental variables assumption using balance measures. *Epidemiology.* 2014;25(5):770-772.
11. Hennessy S, Leonard CE, Palumbo CM, Shi X, Ten Have TR. Instantaneous preference was a stronger instrumental variable than 3- and 6-month prescribing preference for NSAIDs. *J Clin Epidemiol.* 2008;61(12):1285-1288.
12. Ionescu-Ittu R, Abrahamowicz M, Pilote L. Treatment effect estimates varied depending on the definition of the provider prescribing preference-based instrumental variables. *J Clin Epidemiol.* 2012;65(2):155-162.
13. Abrahamowicz M, Beauchamp ME, Ionescu-Ittu R, Delaney JA, Pilote L. Reducing the variance of the prescribing preference-based instrumental variable estimates of the treatment effect. *Am J Epidemiol.* 2011;174(4):494-502.
14. Ionescu-Ittu R, Delaney JA, Abrahamowicz M. Bias-variance trade-off in pharmacoepidemiological studies using physician-preference-based instrumental variables: a simulation study. *Pharmacoepidemiol Drug Saf.* 2009;18(7):562-571.
15. Davies NM, Gunnell D, Thomas KH, Metcalfe C, Windmeijer F, Martin RM. Physicians' prescribing preferences were a potential instrument for patients' actual prescriptions of antidepressants. *J Clin Epidemiol.* 2013;66(12):1386-1396.
16. Rassen JA, Brookhart MA, Glynn RJ, Mittleman MA, Schneeweiss S. Instrumental variables II: instrumental variable application-in 25 variations, the physician prescribing

preference generally was strong and reduced covariate imbalance. *J Clin Epidemiol.* 2009;62(12):1233-1241.
17. Bhattacharya J, Vogt WB. Do instrumental variables belong in propensity scores? (NBER Technical Working Paper no. 343). *Cambridge, MA: National Bureau of Economic Research.* 2007.
18. Wooldridge J. Should instrumental variables bes used as matching variables? *East Lansing, MI: Michigan State Unviersity.* 2009.
19. Myers JA, Rassen JA, Gagne JJ, et al. Effects of adjusting for instrumental variables on bias and precision of effect estimates. *Am J Epidemiol.* 2011;174(11):1213-1222.
20. Seeger JD, Williams PL, Walker AM. An application of propensity score matching using claims data. *Pharmacoepidemiol Drug Saf.* 2005;14(7):465-476.

Figure 1. Causal diagram (directed acyclic graph) illustrating the difference between an optimal instrumental variable and a confounder.

The instrumental variable (IV) is related only to the treatment received and only in that way to the outcome. It does not influence the outcome directly or indirectly by any other means. C_0 is a classic confounder that influences both the treatment received and the outcome. Variable "L" is a poor choice for an instrumental variable owing to an indirect relationship with the outcome via confounder C_1.

Scenario 1.

Scenario 2.

Scenarios 3.

Scenario 4.

Scenario 5A.

Scenario 5B.

Table 1. Predicted Impact of Methodology on the Risk of Bias due to Unmeasured Confounding.

Scenario*	% Divergent Treatment Population Hospital A	% Divergent Treatment Population Hospital B	Proposed Method	Bias Risk
I	0%	0%	IV	Minimized†
II	0%	0%	PS + IV	Minimized
III	0%	10%	ITT or PP IV	Low
IV	0%	20%	PS + ITT or PP IV	Minimized
V	0%	20%	ITT or PP IV	Moderate
VI	10-20%	10-20%	PS ± ITT or PP IV	Minimized
VII	0%	40%	PS ± ITT or PP IV	Minimized††
VIII	0%	40%	ITT or PP IV	High

*Scenarios do not correspond to the scenarios referenced in the text.

†The risk of bias is minimized only assuming both hospital populations represent the equipoise population of interest. We recommend propensity score matching for this reason in all cases.

†† We do not recommend using IV methods if the residual divergent treatment population after PS matching is >20% and caution should be used if the population is 11-20% divergent with acknowledgement of the shortcomings of either an ITT or PP instrumental variable analysis.