METHOD ARTICLE

The ACCE method: an approach for obtaining quantitative or qualitative estimates of residual confounding [version 1; peer review: 2 approved]

Eric G. Smith^{1,2}

¹Psychiatrist, The Center for Organizational and Implementation Research (CHOIR) and the Mental Health Service Line of the Department of Veterans Affairs, Edith Nourse Rogers Memorial Medical Center, Bedford, MA 01730, USA

²Department of Psychiatry, University of Massachusetts Medical School, Worcester, MA 01655, USA

V1 First published: 11 Aug 2014, 3:187
<https://doi.org/10.12688/f1000research.4801.1>
Latest published: 29 Apr 2015, 3:187
<https://doi.org/10.12688/f1000research.4801.2>

Abstract

Background: Nonrandomized studies typically cannot account for confounding from unmeasured factors.

Method: A method is presented that exploits the recently-identified phenomenon of “confounding amplification” to produce, in principle, a quantitative estimate of total residual confounding resulting from both measured and unmeasured factors. Two nested propensity score models are constructed that differ only in the deliberate introduction of an additional variable(s) that substantially predicts treatment exposure. Residual confounding is then estimated by dividing the change in treatment effect estimate between models by the degree of confounding amplification estimated to occur, adjusting for any association between the additional variable(s) and outcome.

Results: A hypothetical example is provided to illustrate how the method produces a quantitative estimate of residual confounding if the method's requirements and assumptions are met. Previously published data is used to illustrate that, whether or not the method routinely provides precise quantitative estimates of residual confounding, the method appears to produce a valuable qualitative estimate of the likely direction and general size of residual confounding.

Limitations: Uncertainties exist, including identifying the best approaches for: 1) predicting the amount of confounding amplification, 2) minimizing changes between the nested models unrelated to confounding amplification, 3) assessing the association of the introduced variable(s) with outcome, and 4) deriving confidence intervals for the method's estimates (although bootstrapping is one plausible approach).

Open Peer Review

Approval Status

	1	2
version 2 (revision) 29 Apr 2015		
version 1 11 Aug 2014	view	view

1. **Mark Lunt**, University of Manchester, Manchester, UK

2. **Gregory Matthews**, Loyola University Chicago, Chicago, USA

Any reports and responses or comments on the article can be found at the end of the article.

Conclusions: To this author's knowledge, it has not been previously suggested that the phenomenon of confounding amplification, if such amplification is as predictable as suggested by a recent simulation, provides a logical basis for estimating total residual confounding. The method's basic approach is straightforward. The method's routine usefulness, however, has not yet been established, nor has the method been fully validated. Rapid further investigation of this novel method is clearly indicated, given the potential value of its quantitative or qualitative output.

Corresponding author: Eric G. Smith (Eric.Smith5@va.gov)

Competing interests: No competing interests were disclosed.

Grant information: This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and Development (HSR&D). Specifically, this work was supported by a VA HSRD&D Career Development Award (09-216) and by support from the Center for Healthcare Organization and Implementation Research. The views expressed in this article are those of the author and do not necessarily reflect the position or policy of the Department of Veterans Affairs or the United States Government.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2014 Smith EG. This is an open access article distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Data associated with the article are available under the terms of the [Creative Commons Zero "No rights reserved" data waiver](#) (CC0 1.0 Public domain dedication).

How to cite this article: Smith EG. **The ACCE method: an approach for obtaining quantitative or qualitative estimates of residual confounding [version 1; peer review: 2 approved]** F1000Research 2014, 3:187 <https://doi.org/10.12688/f1000research.4801.1>

First published: 11 Aug 2014, 3:187 <https://doi.org/10.12688/f1000research.4801.1>

Introduction

Confounding is a central challenge for virtually all nonrandomized studies. Recent research¹⁻⁴ has revealed that propensity score methods may actually increase, or “amplify”, the residual confounding remaining after their application. In general, this recently recognized property of propensity score methods has been viewed as a limitation or complication to the use of propensity scores, for understandable reasons. More recently, however, a study has indicated that the degree of confounding amplification (also termed “bias amplification”⁴) occurring between propensity score models appears to be *quantitatively predictable* (at least in simulation)⁵. Not yet recognized, to my knowledge, is the *extremely* valuable corollary that results: the predictability of confounding amplification should, in principle, permit extrapolation back to an unamplified value of the total residual confounding originally present prior to amplification. (Throughout this manuscript “confounding” refers to baseline confounding. Confounding occurring after treatment initiation from differential discontinuation of the intervention in the treatment group and comparison group is not addressed, but some consideration is given to post-initiation confounding and a possible related approach to addressing to its estimation is briefly discussed in [Appendix 1.3.b](#)). In this manuscript and the associated appendices, I describe the general framework and detailed specifics of a new method designed to use amplified confounding to estimate total residual confounding and an unconfounded treatment effect estimate.

The basic logic of this method is straightforward, but its performance in practice has yet to be confirmed. Testing of this method on both simulated and real-world data is clearly needed. Under specific circumstances, this method may theoretically provide a quantitative estimate of total residual confounding, including from unmeasured factors. Whether and how often this is attainable in practice remains to be determined. This manuscript also illustrates, however, that even when this method is not able to provide a precise quantitative estimate of residual confounding, it may provide a very helpful qualitative estimate of the likely direction and general size of residual confounding. This manuscript is intended to provide detailed information to the research community to facilitate the rapid evaluation of the practical feasibility of this proposed approach.

Method

Step 1 – Create nested propensity score models and generate treatment effect estimates

The “Amplified Confounding-based Confounding Estimation (ACCE) Method” depends on the use of two propensity score models, one (“Model 1”) nested in the other (“Model 2”) so that Model 2 contains all the Model 1 covariates plus an additional variable or variables. Importantly, these added variable(s) should be sufficiently associated with treatment exposure to produce discernible confounding amplification. That is, the variables introduced to the model should further predict treatment exposure sufficiently to substantively *increase* differences between the treatment groups in the prevalences of those confounding factors that are not present in the either model.

Step 2 – Estimate both the proportional amplification of confounding and the quantitative change in the treatment effect estimate between Model 1 and Model 2

In principle, the original confounding existing prior to amplification can be estimated by extrapolation backwards if the proportional amount of confounding amplification and quantitative change in the treatment effect occurring between two propensity score models can be estimated with precision. For example, 2-fold confounding amplification that changed the observed treatment effect odds ratio (OR) from 1.10 in Model 1 to 1.21 in Model 2 (a difference in coefficients of 0.09531) would imply that residual confounding initially existed in Model 1 at such a magnitude as to entirely explain the initial, Model 1 treatment effect ($\beta = 0.09531$, or approximately OR = 1.10). (Please see [Endnote A](#), provided at the end of the manuscript, for more detail). That is, doubling the residual confounding doubled the observed treatment effect estimate, implying all the original treatment effect estimate was due to residual confounding. Attention is needed during the method’s implementation, however, to ensure that changes between the two models distinct from confounding amplification are minimized to the extent feasible ([Appendix 1](#)).

The method requires an ability to estimate the proportional amount of confounding amplification occurring between two propensity score models. Two very different approaches suggest themselves. One approach would be to estimate amplification from existing or future simulation research based on particular metrics of exposure prediction. An example of this approach is research published⁵ using the linear measure of exposure prediction, R^2 . This work demonstrated that, for propensity score stratification or matching approaches, a linear relationship exists between *unexplained* variance in exposure and confounding amplification across the range of $R^2 = 0.04$ to 0.56. This simulation study⁵, using a propensity score based on a linear probability model, also made the important demonstration that different unmeasured confounders appear to be amplified to a highly similar degree. Whether this is true in real-world datasets, or is simply a byproduct of this simulation, clearly merits further investigation. (Further discussion is provided in [Appendix 2.2](#)). Additional research is clearly needed to determine if a similarly predictable relationship exists for other metrics of exposure prediction (such as those proposed for logistic regression^{6,7}), and whether apparent nonlinearities between the prediction of exposure and confounding amplification at more extreme ranges of prediction⁵ can be addressed quantitatively.

A second approach would be to adopt an “internal marker” strategy: deliberately withholding a measured covariate from both models to allow the increase in its imbalance between treatment groups in Model 2 to serve as an approximate indicator of the proportional confounding amplification that has occurred. It is possible, however, that the “internal marker” strategy might consistently yield at least a slight degree of underestimate of the amount of confounding amplification ([Appendix 2.1](#)).

A key assumption of this ACCE method is that residual confounding attributable to different confounders is uniformly or relatively

uniformly amplified in Model 2 compared to Model 1. This important characteristic has been observed in the initial simulation that this method draws upon⁵, but some possibility still exists that the quantitative predictability of amplification that was observed may be merely a consequence of the particular conditions of this simulation (Appendix 2.2).

Step 3 – Adjust for the association between the introduced variable and outcome

The addition of a variable(s) to Model 2 will almost always alter the amount of residual confounding present compared to Model 1, independent of its effect producing confounding amplification (i.e., true instrumental variables are rare). A challenge arises in that it is not the total residual confounding in Model 1 (the quantity being sought) that is amplified in Model 2, only the fraction of that residual confounding that remains *after* introduction of the introduced variable. Because of this, adjustments are needed that reflect the confounding attributable to the introduced variable. However, to estimate total residual confounding through this method, such adjustments must occur to two quantities: 1) the change in the treatment effect estimate between Model 1 and Model 2, and 2) the Model 1 treatment effect estimate.

To make these adjustments, I propose obtaining coefficients for the introduced variable from regression models of the outcome that include all other propensity score covariates. (Please see Endnote B for more detail). This regression coefficient for the introduced variable may be biased by partially reflecting the associations with outcome of those unmeasured confounders that are correlated with the introduced variable. However, the adjustment that is needed at this step of the Method needs to reflect both the change resulting from both the improved balance in the introduced variable *and* from the less extensive changes in the balance of correlated variables that result. To the extent that correlations between the introduced variable and unmeasured confounders produce biases in the introduced variable-outcome association that are similar in size to the amount of increased balance occurring in these covariates, an adjustment that partially reflects the unmeasured covariates could actually be advantageous. The degree of similarity in how correlation affects the introduced variable-outcome association compared to how such correlation affects the balance between treatment groups for the correlated variables is currently uncertain. This is an area worthy of further research.

Once this introduced variable regression coefficient(s) is estimated, the Bross equation⁸ is used to estimate the confounding attributable to the introduced variable(s) and its correlates in both Model 1 and Model 2. (The Bross equation⁸, which recently has been used by Schneeweiss and colleagues in their high-dimensional propensity score algorithm⁹, quantifies the amount of confounding attributable to a confounder by combining the strength of the association between the covariate and outcome and the imbalance in the covariate between the treatment groups. Please see the demonstration of its use in Appendix Table 1). The amount of such confounding in Model 1 is then subtracted from the amount in Model 2 to produce an estimate of the amount of the change in the treatment effect estimate between Model 1 and Model 2 that is attributable to increased balance in the introduced variable(s)

and its correlates. This estimate then is subtracted from the overall treatment effect estimate change from Model 1 to Model 2 to produce the quantity being amplified (the residual confounding in Model 1 separate from the introduced variable). (Please see Endnote C for more detail).

The degree to which this step functions successfully to separate the effect of confounding amplification from any change in the treatment effect estimate attributable directly to the improved balance in the introduced variable has yet to be determined, especially if the introduced variable is correlated with other uncontrolled confounders. However, the theoretical potential to perform the proposed adjustment suggests that this method possibly might provide a quantitatively or qualitatively accurate estimate of an unconfounded treatment effect in circumstances in which instrumental variable analysis may not be possible. At a minimum, the method may prove to provide a relatively accurate estimate of an unconfounded treatment effect in the special case in which the introduced variable is suspected to be largely uncorrelated with important unmeasured confounders. Stated in other words, unlike instrumental variable analysis, it is possible that associations between the exposure-predicting introduced variables and outcome simply complicate, but do not preclude, the use of the method. Further research, however, is clearly needed to determine whether this is the case.

Step 4 – Calculate the unconfounded treatment effect estimate

The final step involves two substeps. First, divide the result from Step 3 (the change in the treatment effect estimate from Model 1 to Model 2, adjusted to remove the change produced by increased balance in the introduced variable(s) and its correlates) by the amount of confounding amplification. This calculation derives by extrapolation an estimate of the total residual confounding in Model 1 except for the confounding attributable to the yet-to-be-introduced variable(s). Finally, subtract both that extrapolated estimate of residual confounding and the confounding attributable to the yet-to-be-introduced variable from the Model 1 treatment effect estimate. (Please see Endnote C for more detail). The result is, in general principle, an estimate of the unconfounded treatment effect.

The accuracy of this estimate, however, is not yet established. The largest uncertainty in this estimate, as discussed above, likely involves the accuracy of the adjustments proposed in Steps 3 and 4 in the context of unmeasured confounders correlated with the introduced variable. In addition, the consistent predictability of confounding amplification needs to be further established. The degree to which other differences between the models can be sufficiently minimized to prevent them from biasing the quantitative estimate of confounding amplification also deserves investigation.

Other research needs include: 1) determining whether random variability particularly reduces the method's usefulness in smaller samples; 2) developing a methodology, such as bootstrapping, to estimate the variance for the final effect estimates; and 3) investigating whether multiple variables can be introduced together if needed to produce sufficient amplification. Nevertheless, the potential significance of a method that may produce estimates of total residual confounding and unconfounded treatment effects from

nonrandomized studies should spur research into the method's feasibility.

Results

Hypothetical example

Consider an example in which the (confounded) Model 1 treatment effect estimate equals $OR = 1.265$ (with an R^2 of 0.25), the (confounded) Model 2 treatment effect estimate equals $OR = 1.2985$ (with an R^2 of 0.50), the introduced variable has an association of approximately $OR = 1.05$ with outcome, an 80% prevalence in the treated group and 20% prevalence in the comparison group in Model 1, and a 52% prevalence the treated group and 48% prevalence in the comparison group in Model 2. (This example assumes a linear propensity score model but a logistic regression outcome model. Please see [Endnote E](#) for more detail). What is observed is an increase in the treatment effect estimate *away from the null* in Model 2. This change away from the null occurs despite tight control in Model 2 (but not Model 1) of a variable (the “introduced variable”) that is not only highly predictive of exposure but is also, to some degree, a confounder that would have been expected to have biased the treatment-outcome association at least modestly away from the null in Model 1. This suggests, in the absence of confounding amplification, that the tight control of this covariate in Model 2 would ordinarily result in a less biased treatment effect estimate moving *towards, not away from*, the null. Furthermore, given the mere 1.5-fold amplification of confounding that would be expected to result (0.75 remaining variance unexplained in Model 1 versus 0.50 variance remaining unexplained in Model 2, or $0.75/0.50 = 1.5$), the fact that this modest confounding amplification is sufficient to move the treatment effect estimate away from the null despite tight control of a confounder with an $OR = 1.05$ implies that a substantial proportion of the Model 1 effect estimate is attributable to confounding (biasing away from the null). Specifically, these findings would imply that more than half of the original, sizeable “treatment effect” estimate ($OR = 1.265$) was attributable to residual confounding, and would suggest a genuine unconfounded treatment effect estimate of only $OR = 1.10$. (Please see [Supplementary Table 1](#) for complete calculations).

Thus, despite the fact that the treatment effect estimates for Model 1 and Model 2 are both confounded, knowledge of the amount of expected confounding amplification allows the comparison of the effect estimates of models (with appropriate adjustments) to yield an estimate of an unconfounded treatment effect.

Application to published data

The study of Patrick *et al.*¹⁰ provides sufficient detail to fortuitously provide a partial opportunity to test some aspects of the ACCE methodology on real-world data. Obviously, this study was not constructed to illustrate the ACCE Method; therefore it is being used *post hoc* to explore the potential of the method. As a result, the data provided include several additional uncertainties beyond those that would accompany a deliberate implementation of the ACCE Method. However, by permitting an examination of the performance of even a partial version of the ACCE Method, this study illustrates the potential value this method may have as a probe indicating whether substantial residual confounding is likely (and its likely direction), even in circumstances in which a firm quantitative estimate of residual confounding is not able to be derived.

Patrick *et al.*¹⁰ derived a substantial number of propensity scores during their analyses of the association between statins and both all-cause mortality and hip fracture outcomes. Of note, two of the propensity scores used (for both outcomes) included an important pair in which one propensity score was nested within a slightly larger propensity score identical to the original propensity score except for the addition of a single covariate (glaucoma diagnosis). Glaucoma diagnosis was considered to be a potential instrumental variable in these analyses. First, glaucoma diagnosis was associated extremely strongly with treatment exposure (since the treatment group compared to statins for both analyses consisted of users of medications for glaucoma). Patients with a glaucoma diagnosis had an odds ratio for statin exposure of 0.07 (that is, patients with glaucoma diagnosis had approximately a 14× greater odds of being in the comparison treatment group than the statin treatment group). Second, it is plausible (although not provable) that glaucoma diagnosis lacks a substantial association with the outcomes of all-cause mortality and hip fracture, and thus may be functioning as an instrumental variable or near-instrumental variable. (Although not termed an “instrumental variable” originally¹⁰, such a term was used for glaucoma diagnosis in these analyses in a subsequent manuscript describing these findings¹¹).

Patrick *et al.*¹⁰ reported both effect estimates and a measure of prediction (the *c* statistic) for the original (“Model 1”) model and after adding the “introduced variable” (i.e., glaucoma diagnosis) (“Model 2”). This permits an examination of the valuable qualitative findings that might result even when the ACCE Method is unable to produce a precise quantitative estimate of residual confounding. In this somewhat artificial case, the partial version of the ACCE Method that can be implemented is unlikely to produce precise quantitative estimates of residual confounding for several reasons, including the fact that the relationship of the *c* statistic to confounding amplification has yet to be explored, unlike the relationship between R^2 and confounding amplification. In addition, the partial version of the method that can be implemented does not include the possible checks of model similarity in confounding control, patient sample, and intervention delivered (e.g., dose) described in [Appendix 1](#). Of particular importance, this partial, illustrative version of the method does not include any adjustment to account for the association of the introduced variable (glaucoma diagnosis) with outcome (using information estimated from a full multivariate regression containing the other propensity score covariates). This lack of adjustment somewhat limits this example, since even a small association with outcome of a covariate with such an imbalance in prevalence between the treatment groups may contribute substantively to overall confounding. In fact, the manuscript notes that the minimally-adjusted hazard ratio (HR) for glaucoma diagnosis (adjusted for age, age², and sex) is >1.175 or $<1/1.175$ for both outcomes. (The actual age and-sex-adjusted HR observed is $HR \approx 0.85$ for both outcomes [Amanda Patrick, Personal Communication]). What is lacking, however, is the glaucoma diagnosis HR adjusted for *all* the covariates in the propensity score model, rather than just age and sex. (This adjustment would involve including a total of 143 covariates for the mortality analysis and 120 covariates for the hip fracture analysis¹⁰). This fully-adjusted HR would provide information about whether or not the age and sex-adjusted glaucoma diagnosis HR might be related to aspects of care-seeking, care access, health attitudes, or other factors that might be also represented by other

covariates (leaving a much lesser or close-to-null association for glaucoma diagnosis in the actual analysis). Most importantly, this fully-adjusted association would provide the quantity needed to help calculate the estimate of the unconfounded treatment effect estimate for Model 1 (Steps 3 and 4 of the method).

Interpretation of the published results using a highly partial version of the ACCE Method

Despite the limitation of not having a fully-adjusted regression coefficient for the glaucoma diagnosis-outcome association, as well as the other substantial limitations mentioned above, application of even this highly partial version of the ACCE Method appears to provide useful qualitative estimates of residual confounding for these two analyses (all-cause mortality and hip fracture).

Table 1A shows that in the all-cause mortality analyses, addition of the introduced variable (glaucoma diagnosis) moves the treatment effect estimate away from the null by a modest amount. This implies

that the total residual confounding (including residual confounding from unmeasured factors) likely biases, but only very modestly, towards observing a larger effect size for statins than is genuinely present. This result is consistent with the effect estimate derived from available randomized data. In contrast, Table 1B shows that in the hip fracture analyses, addition of the same introduced variable changes the observed treatment effect HR from 0.76 to 0.69. This is a much more sizeable change in the treatment effect estimate, implying a larger quantity of underlying residual confounding biasing the estimate away from the null. If glaucoma diagnosis is in fact a near-instrumental variable, the results would imply that the unconfounded hip fracture treatment effect estimate is considerably closer to null, the approximate value that the authors expect to be the genuine treatment effect based on randomized data¹².

Even if glaucoma diagnosis is not functioning as a near-instrumental variable, as long as the full multivariate regression coefficients for glaucoma diagnosis are even somewhat similar between the models,

Table 1. Application of the qualitative version of the ACCE Method to published data (Patrick et al., 2011).

A. Nested Models differing by single variable with observed strong association with exposure and expected minimal association with outcome.

Exposure-Predicting Introduced Variable	Intervention	Outcome	Original Model ("Model 1")	Larger Model (by a single covariate) ("Model 2")	C statistics	
					Model 1	Model 2
Glaucoma Diagnosis	Receipt of statin (vs. receipt of antiglaucoma medication)	All-cause mortality	PS Model restricted to variables with a +/- 20% association with the outcome	PS Model restricted to variables with a +/- 20% association with outcome PLUS Introduced variable (glaucoma diagnosis)	0.82	0.90
Results						
Model 1 Hazard Ratio (Central Estimate)	Expected Genuine Treatment Effect (Based on Randomized Trial Meta-analyses)	Model 2 Hazard Ratio (Central estimate)	Residual Confounding Suggested by Model 2 Treatment Effect Estimate compared to Model 1 Treatment Effect Estimate			
0.84	HR ≈ 0.85 or less (i.e., closer to the null) ^a	0.82 ^b	Modest ^c , in the direction away from the null (i.e., towards a more protective apparent effect than likely genuinely exists) That is, the ACCE Method suggests the likely direction and general size of residual confounding (and thus that the genuine treatment effect is likely closer to the null than the initial Model 1 estimate), even in the absence of a precise quantitative estimate of residual confounding.			

^a Reference 10, Table 2 and Discussion.

^b Reference 10, Results section text (4th paragraph).

^c For residual confounding not to be modest (relative to treatment effect estimate) either 1) the introduced variable would have to have a substantial association with *increased* mortality risk. (This seems rather unlikely, since the age and sex-adjusted HR is in the protective direction [HR = 0.85; M. Patrick, personal communication], but cannot be rigorously excluded), or 2) the amplification would have to be distinctly minor (e.g., approximately 1.25x). It is assumed here that amplification from the c statistic occurs in similar fashion as with R² in the simulation of Reference 5; that is, that the change in the remaining unexplained variance of exposure predicts amplification. This has not been established for the c statistic (and it is generally appreciated that the c statistic is not a very desirable metric for comparisons between models). Nevertheless, while we do not know the amplification precisely, amplification would appear to have to be much less than that observed by Reference 5 in similar ranges of exposure prediction using R² for the partial ACCE method applied here to predict a large amount of residual confounding in this analysis. Furthermore, whatever the amplification is, it is likely to be highly similar between Table 1A and Table 1B. Thus, the conclusion concerning the *relative* amount of unmeasured confounding in the all-cause mortality compared to the hip fracture analyses given in Table 1B is likely to be valid (as long as the fully-adjusted glaucoma diagnosis association does not differ markedly for the two outcomes).

Table 1. Application of the qualitative version of the ACCE Method to published data (Patrick et al., 2011) (continued).

B. Nested Models differing by single variable with observed strong association with exposure and expected minimal association with outcome.

Exposure-Predicting Introduced Variable	Intervention	Outcome	Original Model (“Model 1”)	Larger Model (by a single covariate) (“Model 2”)	C statistics	
					Model 1	Model 2
Glaucoma Diagnosis	Receipt of Statin (vs. receipt of antiglaucoma medication)	All-cause Mortality	PS Model restricted to variables with a +/- 20% association with the outcome	PS Model restricted to variables with a +/- 20% association with outcome PLUS Introduced variable (Glaucoma Diagnosis)	0.81	0.89
Results						
Model 1 Hazard Ratio (Central Estimate)	Expected Genuine Treatment Effect (Based on Randomized Trial Meta-analyses)	Model 2 Hazard Ratio (Central Estimate)	Residual Confounding Suggested by Model 2 Treatment Effect Estimate compared to Model 1 Treatment Effect Estimate			
0.76	HR $\approx$ 1.03 ^a	0.69 ^b	More substantial than for all-cause mortality, in the direction away from the null (i.e., towards a more protective apparent effect than likely genuinely exists) ^c . That is, the ACCE Method suggests the likely direction and general size of residual confounding (and thus that the genuine treatment effect is likely to be <i>closer to the null</i> than the initial Model 1 estimate), even in the absence of a precise quantitative estimate of residual confounding.			

^a Reference 10, Table 2 and Discussion.

^b Reference 10, Results section text (4th paragraph).

^c Given that the all-cause mortality and hip fracture analyses have propensity score c statistics suggesting highly similar predictions of exposure, seemingly the only likely plausible scenario by which the all-cause mortality analysis could be more confounded than the hip fracture analysis is if the introduced variable of glaucoma diagnosis has a substantially stronger protective association with outcome after control for the other propensity score covariates than the association between glaucoma diagnosis and all-cause mortality. Since these results are not available (i.e., results from an extensive multivariate regression), such a possibility cannot be rigorously excluded. Some difference might even be plausible given that considerably less is known about the predictors of hip fracture (and what is known may be less represented in healthcare databases) than for all-cause mortality. It can be inferred, however, that the magnitude of this difference would need to be substantial for the ACCE Method to suggest that the hip fracture analysis is less confounded than the all-cause mortality analysis. In an actual implementation of the ACCE method, highly-adjusted multivariate regression of the introduced-variable-outcome association would be conducted involving all or many (if the number of outcomes did not permit all the covariates to be simultaneously included) of the propensity score covariates.

these two analyses considered together suggest the presence of considerably more residual confounding in the hip fracture analysis than the all-cause mortality analysis. (Please see [Endnote F](#) for more details). This is a conclusion independently suggested by the randomized trial meta-analyses^{12,13} cited by the authors. That is, based on the differences between the propensity score findings compared to the randomized trial meta-analyses (i.e., the hip fracture HR differed much more from previous randomized findings than the all-cause mortality HR), the general supposition would be that the hip fracture analysis is likely to be considerably more confounded than the all-cause mortality analysis. The ACCE Method, even when applied in a very partial and qualitative form, suggests the same conclusion. In this fashion, the ACCE Method may prove useful for estimating at least the likely general size and direction of residual confounding in the many circumstances where substan-

tial randomized trial data is not available to guide one's interpretation. This capacity of the method to provide even a qualitative estimate of residual confounding may constitute an important analytic advance.

Discussion

This paper presents a relatively straightforward four-step method exploiting the phenomenon of confounding amplification to potentially provide quantitative estimates of total confounding and unconfounded treatment effects. To my knowledge, it has not been previously recognized that the phenomenon of confounding amplification, if predictable (as suggested by recent simulation⁵), provides a potential mechanism to estimate total residual confounding. The fundamental approach of deliberately introducing amplified confounding into an analysis to evaluate the total residual confounding existing prior to amplification appears to possess both

clear logic and considerable promise. The method hinges on part on whether the recently observed predictability of confounding amplification is found to be a generally observed phenomenon; in addition, at this stage it is unclear whether the method will need particularly large sample sizes to be routinely useful in providing quantitative estimates. Nevertheless, although aspects of the method's implementation and precise accuracy are not yet fully resolved, further research is clearly indicated given the potential value of a new approach that may advance efforts to remove confounding from nonrandomized treatment effect estimates.

Furthermore, even if subsequent research determines that the estimates from this approach typically are sufficiently imprecise as to limit the quantitative usefulness of the method, this general approach may have considerable value as a semi-quantitative or qualitative "probe" of whether a substantial amount of residual confounding likely exists. It is hoped that the description of the method provided here is sufficient to permit the larger research community to immediately begin participating in the validation and refinement of this novel approach.

Considerations for validation and further research

The ACCE method is fundamentally a conceptually simple approach, but one that may require some care in its implementation (e.g., in the need to structure the two models so as to minimize other changes that might influence the treatment effect estimate while obtaining sufficient confounding amplification). The value of this method will depend on how often in practice it provides a useful quantitative or qualitative estimate of residual confounding. Answering this question will involve more detailed and precise examination of both simulated and real-world data, and almost certainly will involve the contributions of multiple research teams.

Useful avenues for validation research likely include: 1) the predictability of the relationship between a particular metric measuring prediction of exposure and confounding amplification and/or the potential substitutability of an "internal marker" as an alternative approach; 2) approaches to, or circumstances that would, ensure other changes between the models (in patient sample, intervention received, and the degree of control achieved for measured, included confounders) are minimized; 3) confirming that multivariate regressions provide an accurate measure of the change in confounding resulting from balancing of the introduced variable in Model 2 (and thus permits adjustment for the direct and indirect contributions of the introduced variable(s) to confounding in Model 1); 4) determining how easily multiple introduced variables can be used if a single introduced variable does not produce sufficient confounding amplification; and 5) determining whether sufficiently precise results can be routinely obtained from the ACCE Method despite the effects of random variability in treatment effect estimates, since this method requires the accurate detection of what may be fairly small changes in treatment effect estimates. It may prove that, for this reason, this method may be most useful when applied to particularly large databases; however, some recent studies using propensity score-based stratification do suggest that quite subtle changes in relative risk or hazard ratio from application of slightly different propensity score models can be detected^{9,10}. Finally, an obvious need exists for methodology to develop confidence

limits around the effect estimates emerging from the ACCE method. The procedure of bootstrapping would be one obvious candidate approach.

Simulation studies, given that the genuine treatment-outcome association is able to be specified by the investigator, may be the most immediate approach to addressing these research needs and evaluating the performance of this method in general. (Such simulations would be similar to the recent simulation study initially observing that confounding amplification may be predictable⁵, and others that have considered the impacts of unmeasured confounding^{13,14}). Real-world studies might investigate whether the method appears to accomplish the task of making results from nonrandomized studies better parallel results from randomized trials¹⁵.

Potential application of the method to comparative effectiveness and surveillance research

Regardless of its ultimate precision, this method may prove beneficial for nonrandomized comparative effectiveness research in general, as well as especially beneficial for studies in which substantial residual or unmeasured confounding is expected. For example, many studies of mental health and/or behavioral interventions might be expected to have substantial unmeasured confounding, since the important elements of the conversation between provider and patient that contributes to judgments of the severity of the patient's condition and helps influence treatment decisions often may go unrecorded even in the patient's chart, and thus becomes unmeasurable.

Another notable use would be to enhance medication surveillance efforts. By providing even a highly approximate estimate of unmeasured confounding in a few simple steps, the ACCE Method could help more accurately indicate which prominent "signals" (either in effectiveness or safety) observed during the screening of large datasets appear to be less confounded (and thus are a particular priority for further investigation).

Conclusions

This paper has outlined a relatively straightforward yet novel method to potentially obtain a quantitative estimate of total residual confounding. This total residual confounding estimate (which would include confounding from unmeasured as well as measured factors) then allows, in principle, for an estimate of unconfounded treatment effects to be calculated. This paper has described the steps involved in applying this method, offered a very preliminary examination of the performance of a simple, partial version of this method using published data, and outlined research needs for refinement and validation of this method. Given the importance of a method that may potentially help remove confounding from nonrandomized treatment effect estimates, further investigation of this method by multiple research groups is clearly warranted. Even if the ACCE method is eventually shown to have limitations or evolves from the form proposed here, the method's general approach of deliberately amplifying confounding to reveal existing residual confounding may have enduring analytic value. The ACCE Method and its underlying logic therefore have the potential to constitute a substantial advance for nonrandomized intervention research, and follow-up research should be rapidly conducted.

Endnotes

- A. Not addressed in this simple example is the fact that, in almost all implementation of this Method (i.e., all implementations other than introducing a true instrumental variable), these calculations would need to adjust for the association with outcome of the variable(s) introduced into Model 2 to produce the amplification. This is discussed subsequently in Steps 3 and 4.
- B. These regressions could be performed either within treatment arms or across both treatment arms while including an indicator for treatment arm, as well as a covariate(s) for treatment arm-introduced variable interaction(s). Comparing the results of all these approaches may be useful.
- C. A key area for additional investigation is whether the effects upon the treatment effect estimate of the increasing balance in Model 2 in variables correlated with the introduced variable is adequately reflected by the adjustment proposed in Steps 3 and 4. This proposed adjustment does separate the residual confounding associated, directly or indirectly, with the introduced variable (which is being controlled in Model 2 and therefore cannot amplify) from the residual confounding being amplified. However, whether this separation and calculation fully captures the change in confounding attributable to the resulting increase in control, even if modest, of unmeasured confounders correlated with the introduced variable is unclear. Even if this adjustment should prove only incompletely effective in capturing the change in confounding attributable to correlated covariates, it may be determined that sometimes this is a relatively small source of error. The method would be also expected to exhibit its strongest performance when introduced variable(s) can be chosen that are suspected to be largely uncorrelated with potential unmeasured confounders. Please see [Appendix 2.2](#) for further discussion.
- D. Subtraction of both these quantities is necessary because, as pointed out in Step 3, the process of adding the introduced variable to Model 2 means that the amplification that occurs in Model 2 is not amplification of all the residual Model 1 confounding, but only the *remaining* Model 1 residual confounding (i.e., minus the contribution of the introduced variable and its correlates). Therefore, the value for the original residual confounding in Model 1 that is extrapolated from the amplified value does not include the contribution of the yet-to-be-introduced variable(s) and its correlates. The contribution to Model 1's original residual confounding that is attributable to the yet-to-be-introduced variable(s) and its correlates must be subtracted, along with the extrapolated remaining residual confounding, from the Model 1 treatment effect estimate to estimate an unconfounded treatment effect.
- E. This example assumes a linear propensity score model but a logistic regression outcome model because the existing simulation demonstrating proportional confounding amplification is for a linear propensity score model⁵. Still to be determined is whether linear, rather than logistic, outcome models will need to be used for the ACCE method's estimates to be the most accurate, due to the need to compare risks of outcome between Model 1 and Model 2. A requirement for linear outcome models, if it exists, would add complications; however, it may prove that these complications are relatively minor drawbacks in the context of permitting the ACCE's Method's estimation of residual confounding and an unconfounded treatment effect estimate. This is another worthwhile area for additional research. Meanwhile, the published data examples suggests the ACCE Method may contribute useful information to guide inferences from nonrandomized studies even when only outcomes from nonlinear analyses (i.e., hazard ratios from Cox regression) are available. However, these examples are premised on the assumption that the c statistic can serve at least as an approximate index of confounding amplification.
- F. This comparison can be made in this straightforward fashion since for the two analyses, the change in the prediction of exposure (in this case the c statistic) was highly similar (all-cause mortality: Model 1 c = 0.82, Model 2 c = 0.90; hip fracture: Model 1 c = 0.81, Model 2 c = 0.89). Thus, the resulting confounding amplification would be expected to be generally similar.

Competing interests

No competing interests were disclosed.

Grant information

This material is based upon work supported by the Department of Veterans Affairs, Veterans Health Administration, Office of Research and Development, Health Services Research and Development (HSR&D). Specifically, this work was supported by a VA HSRD&D Career Development Award (09-216) and by support from the Center for Healthcare Organization and Implementation Research.

Acknowledgements

The author would like to thank the numerous individuals who provided important encouragement or support of this manuscript throughout its development. Specific thanks should go to the friends and colleagues who provided very helpful reviews of manuscript drafts, including Brian Sauer, James Burgess, Lawrence Herz, David Hoaglin, Susan Eisen, Katherine Hoggatt, Gunet Jasuja, Keith McInnes, Donald Miller, C. Arden Pope, Karen Quigley, Kevin Rader, David Smith, Marcia Valenstein, and Amy Borg. The author also wants to thank John Brooks for providing a timely and thoughtful email response clarifying aspects of his simulation, and Amanda Patrick for generously discussing her analyses and providing the quantitative value for the age-and-sex adjusted glaucoma diagnosis hazard ratio and thanks to Jeroan Allison for the suggestion to consider bootstrapping as an approach to generating confidence intervals. However, the author alone is responsible for the ideas advanced in this manuscript, as well as the final form of the manuscript and associated documentation and whatever errors or oversights they may contain. In addition, the author would like to specifically thank the Health Services Research and Development Office of the Veterans Health Administration for their generous funding of his Career Development Award that helped provide valuable protected time to dedicate to the development of the ideas in this manuscript.

Supplemental appendices

The essence of the proposed method described in the manuscript can be summarized in a very simple explanatory example. If one knew that turning up the volume on a television or radio doubled the volume (if, for instance, if the volume settings were genuinely proportional, so that turning up the volume from a setting of “20” to “40” doubled the sound produced), and one knows the actual volume obtained after this doubling occurred, it should be both possible and simple to extrapolate back to determine what the original volume was. That is, it would not be necessary to know the original volume if you knew these other two quantities (the final volume, and the proportion by which the volume changed). In nonrandomized intervention studies, one never knows exactly the “volume”, or amount, of total confounding, so an additional wrinkle is employed of measuring the *change* in the overall effect estimate that occurs between two models when only the amount of confounding is deliberately changed (through amplification). To make this example even more closely comparable, consider the scenario in which one knew that a certain sound system apparatus would unfortunately double (i.e., “amplify”) the static, or white noise, in whatever sound is being broadcast. If one knew the original *total, or overall*, volume (on a linear scale) was “90 units”, and this changed to 93 units when the particular apparatus was used, then we would know that the static made up 3 units of the original sound (since doubling it added 3 units). This would mean that the volume of sound devoid of any static was 87 units. (I deliberately refer to an imaginary linear unit of sound, rather than using the highly-familiar units of “decibels”. Decibels use a logarithmic scale, which would make the example less easily appreciated). It would not be necessary to know beforehand the value of the sound volume without static; rather, knowing the static had doubled and how much the sound volume had changed would permit extrapolation backwards to determine the unknown, devoid-of-static value. “Static” in this example can be seen as analogous to residual baseline confounding, and the sound volume without static as analogous to the unconfounded treatment effect.

When applying this approach to comparative effectiveness research, however, the details of the approach are very important. While only the amplification of confounding is being *deliberately* changed, it is very easy to unintentionally introduce changes to additional aspects of the models besides simply the amplification of confounding. These appendices therefore highlight the major points that I have been able to identify to date which appear to warrant some consideration in the application of the method.

Some readers may view this level of detail concerning important details of the method to be premature, since the method has not yet been extensively validated. I hope instead that these Appendices will both facilitate the method’s rigorous validation and promote sophisticated use of the method going forward. The details discussed below may not prove as important if the method is ultimately determined to provide only general, highly-approximate qualitative estimates of residual confounding. If, however, this method indeed appears to provide quantitative estimates of residual confounding in some circumstances, the details discussed below and even further details yet to be identified may prove important to consider or address.

Appendix 1: Other elements of the analysis that may produce changes in treatment effect estimates between Model 1 and Model 2

This method ultimately views the change in the treatment effect estimate between Model 1 and Model 2 (minus adjustment for the contribution of the introduced variable(s) and their correlates) as arising from confounding amplification. As a result, an obvious and crucial need exists to keep all differences between Model 1 and Model 2 other than the amplification of confounding to a minimum, to the extent feasible. Changes to the Model 2 treatment effect estimate, compared to Model 1, can be expected to occur in several areas in addition to confounding amplification. As discussed below, these areas of potential differences between the two models include changes in the control of the confounding from “included” covariates (covariates that are included in both propensity scores), the comparability of the patient sample, and the comparability of specific aspects of the intervention received by patients.

1.1. Differences in control of confounding from included covariates

Although these changes may be expected to be minor, at least in some settings, they deserve thorough consideration as part of an effort to anticipate sources of potential imprecision in the estimates resulting from the ACCE method. Some degree of change between Model 1 and Model 2 is expected in the balance of each of the propensity score covariates present in common between Model 1 and Model 2 (the “included covariates”). These changes are produced as a byproduct of the need to include an additional variable or variables in Model 2 to generate confounding amplification.

(NOTE: The additional “introduced variable” is also an “included” covariate in a limited sense, in that it is included in one of the two propensity scores. For clarity in terminology and because the introduced variable represent a genuinely special circumstance (please see [Appendix 2.2](#)), the term “included variable” or “included covariate” is reserved for the variables included in *both* models. The term “introduced variable(s)” is used for the variable(s) *added* to Model 2. The term “nonincluded covariate” or “nonincluded variable” will refer to covariates not included in either propensity score. Non-included covariates may either be unmeasured covariates (which inherently cannot be included), or measured covariates not selected for inclusion into the propensity score).

This change can be best visualized by considering the case of propensity score matching: including an additional variable in the propensity score used for matching would be expected to weaken at least slightly the tightness of the match on the other covariates. In some cases, the differences in covariate balance between models could be quite minimal. However, this balance needs to be explicitly compared between Model 1 and Model 2. This comparison is important because in some cases it may prove difficult to attain a control of confounding from included covariates that is strictly equivalent between Model 1 and Model 2 if sufficient confounding amplification is to be achieved. The impact of confounding amplification is to tend to create a greater number of individuals at the extremes of the propensity score distribution who are less comparable (and thereby, less similar in balance in the covariates included in the model)¹.

If these differences are not trivial, one approach possibly worthy of future exploration would be to examine whether it is feasible to adjust the stringency of the stratification or matching in Model 2 so that the balance in the included covariates in Model 1 and Model 2 are more equivalent. Another would be to use the Bross equation⁸ to attempt to estimate the change in confounding attributable to the observed changes in the balance of these measured covariates.

1.2. Differences in patient sample

The overall patient cohort for the study from which the samples for Models 1 and 2 are derived will obviously not change. Some degree of change, however, can be anticipated to occur (although it may be minimal), in the samples of individuals selected from that overall cohort by Model 1 and Model 2. These differences can arise from differences in the patients that fall under the “Common Support Area” of the propensity scores, and, if matching is employed, differences in the percent of patients matched. The “Common Support Area” refers to the range of propensity score values which include members of both treatment groups; it is often recommended that individuals outside the Common Support Area be “trimmed” (i.e., removed) from the analysis¹⁶. These differences in patient sample, however, will only influence the method’s estimates to the extent that they are extensive enough to produce substantively different compositions of patients between the Models and effect modification exists (whereby the treatments studied have different effects in different patients). In addition, possible strategies exist to minimize some of these potential differences. These are discussed below.

1.2.a. Differences in patient sample from differences in common support area/propensity score trimming. Because confounding amplification tends to make at least patients on the extremes of the propensity score distribution less comparable⁵, it might prove difficult in practice to maintain a highly similar Common Support Area between Model 1 and Model 2. Fortunately, the number and identity of individuals differing between Model 1 and Model 2 is measurable. In addition, different approaches might be compared (such as examining in both models only the subset of patients that fall under the Model 2 Common Support Area, or all the patients that fall under either Model’s Common Support Area). These comparisons might establish the sensitivity of the results to small differences in the Model 1 and Model 2 patient samples arising from different Common Support Areas.

1.2.b. Differences in patient sample in the two models from differences in percent matching. Regarding matching strategies, it may prove difficult for a similar proportion of matching to be preserved between the two models. (The Brooks and Ohsfeldt simulation⁵ showed that as unexplained variance decreased and amplification increased, the number of patients matched for a given caliper decreased). The alternative approach, propensity score stratification, may be determined to be preferable in some cases, since by design stratification retains all individuals from the trimmed sample. (The ACCE Method emphasizes stratification and matching because in the Brooks and Ohsfeldt simulation¹ propensity score weighting produced confounding amplification that was less predictable, at least by R^2).

1.3. Differences in specific aspects of the intervention received

While the general nature of the interventions received by the two treatment groups remains identical between Model 1 and Model 2, specific aspects of the intervention received can vary between the treatment groups in ways that are not immediately obvious. It is important to consider these possible differences because they may be another, non-amplification related, contributor to differences in the treatment effect estimates between Model 1 and Model 2.

1.3.a. Differences in dose. Unless the intervention is a single, one-time-only dosed treatment, such as a vaccine, either dose or other “quasi-dose” aspects of how the intervention is administered may vary at least slightly between the individuals receiving the intervention in Model 1 and those receiving the intervention in Model 2. Even for nonmedication-based interventions, such as a psychotherapy or educational intervention, the timing of visits or the number of visits may vary slightly among the individuals included in the intervention arm in Model 1 versus Model 2. Therefore, when implementing the ACCE method, it may prove important to examine whether the overall mean dosage, number, or timing of treatments is similar between Model 1 and Model 2, and potentially within strata for stratified analyses. If sufficient sample size exists in particularly large patient samples, an additional approach might be to restrict the analysis to patients only receiving one particular dosage of the medication.

1.3.b. Differences in discontinuation rates. Treatment effect estimates are likely to be sensitive to discontinuation rates, whether an intent-to-treat or as-initially-treated analysis (i.e., with follow-up censored upon alteration of the initial treatment) is conducted. Because of this, investigators should examine the rates of discontinuation observed in Model 1 and Model 2 to determine their similarity. Ideally, a determination that the *reasons* for discontinuation within the patient sample for Model 1 compared to Model 2 were also similar is the ideal, but such information is often not available¹⁷.

In many cases the difference in discontinuation rates for the treatment of interest between the two models may be quite small, and the practical impact of this difference unclear. Differences in discontinuation rates appear to be at least slightly more significant in this approach, however, than in a propensity score or regression analysis involving a single model. If any effect modification exists, then changes in the group remaining in treatment produced by differences in discontinuation between the two models could produce some degree of difference in the underlying treatment effect estimate between the models, even if the factors governing discontinuation in both cases were not related to outcome.

Addressing specific differences in the intervention received by patients between the models, if they exist, may be difficult. One modest strategy might involve simply evaluating the differences in intervention specifics when different strategies are explored to minimize differences in patient sample and/or in the control of confounding from included covariates. Then the approach that also minimizes differences in intervention specifics might be favored, or at least examined as a sensitivity analysis.

As indicated in the main manuscript (Endnote A), this manuscript generally does not consider confounding arising after treatment initiation from differences in patient characteristics of patients who remain receiving in the two treatment groups. However, three points should be made. First, what this means is that the “unconfounded treatment effect estimate” that is estimated from this method is not necessarily estimating a completely unconfounded effect estimate, in the strictest sense of the term. Rather, the method seeks to provide a treatment effect estimate unconfounded from *baseline* confounding. Second, if confounding between the treatment groups exists that arises after treatment initiation but is similar between the two models (which may be somewhat plausible, for instance, if discontinuation rates vary little between the two models), then a treatment effect estimate largely unconfounded from baseline confounding, at least, can still be estimated. (This estimate can occur since the confounding post-initiation would be expected to stay relatively constant and not be a major source of differences in the treatment effect estimates between the two models). Third, it is conceivable that the same approach used here – deliberately introducing confounding amplification to estimate the original confounding present – could be used, at least in theory, to also estimate confounding post-initiation, occurring from differential discontinuation during treatment. This is likely to be a substantially harder and more complex endeavor, however, since the most commonly used approach to addressing post-initiation confounding, generating a “pseudopopulation”, usually occurs by weighting, which has not been shown in simulation to produce very predictable confounding amplification⁵. Matching instead of weighting could be considered, but the results would become applicable to only a smaller and smaller subset of patients. One relatively simple pragmatic approach might be to conduct this matching at only a single additional time point – study completion. Clearly, much additional work is needed to fully understand if the ACCE method could also help address confounding occurring after treatment initiation.

Summary

The need to evaluate, and potentially address, these diverse factors that may contribute to a change in treatment effect estimates from Model 1 to Model 2 may initially seem daunting. However, it may be ultimately determined that in practice little difference between the Models in these aspects is typically observed. In theory, there may even be circumstances in which sizeable differences in some of these aspects do not prevent the method from providing accurate estimates (e.g., a difference in timing of an intervention whose effects have been shown to not be very sensitive to the timing of its administration). In most cases, however, substantive differences of the types described between the models would be a concern. If these differences cannot be minimized by the strategies suggested, or approaches to quantifying the likely effects of these differences cannot be identified, then caution in interpretation is clearly warranted. Validation studies using simulated or real-world datasets would be useful by providing information concerning both the frequency with which these differences occur and their impact on the ACCE Method’s estimates.

Appendix 2: Important considerations involved in the estimation of confounding amplification

Appendix 2.1. Approaches to estimating confounding amplification

In the lower ranges of exposure prediction (at least as measured by explained variance in terms of R^2), a simulation has shown that a predictable relationship exists between amount of remaining unexplained variance in the prediction of exposure and confounding amplification⁵. Differences in prevalence between treatment arms in covariates that are not included in the propensity score increase linearly with increases in R^2 . This increase in the imbalance of uncontrolled (i.e., nonincluded) factors is the phenomenon underlying the amplification of residual confounding. However, in this simulation in the upper portion of the range of R^2 the relationship becomes increasingly nonlinear, with changes in R^2 underestimating the increased imbalance in nonincluded covariates⁵. It is unclear whether this nonlinearity relates to particulars of how the simulation was designed; therefore it is difficult to judge whether similar nonlinearities will continue to be observed, even for the R^2 metric of exposure prediction. However, since the Brooks and Ohsfeldt simulation is the only data currently available, it is worth considering the implications of possible nonlinearity in the upper portions of the range of exposure prediction in detail.

If this nonlinearity in the upper portions of the range continues to be observed when using R^2 , but is reduced or not apparent for other metrics of prediction of exposure, then these metrics should be preferred. If this nonlinearity in the upper ranges of exposure prediction continues to hold for other metrics, then three strategies suggest themselves. The first approach, the “low amplification strategy”, would be to deliberately limit Models 1 and 2 so that the prediction of exposure these models achieve are in the lower end of the possible range, where the relationship is most linear. In some cases, propensity score models may already fall into this range. In other cases, this approach may involve reducing or minimizing the variables included in the propensity scores. Such reduction might entail including only variables with a significant *a priori* expectation, based on evidence, of being confounders¹⁸. An even more restrictive strategy would include those variables estimated (by using the Bross equation⁸) to be the most substantial confounders, or suspected *a priori* as being the most certain confounders (e.g., age, Charlson Comorbidity index, etc.). Reductions in the number of included covariates could increase residual confounding relative to some other models that could be constructed. To the degree genuine confounders were removed from the propensity score, the analysis would become increasingly dependent on the ACCE method to capture this larger amount of residual confounding and accurately remove it in order to obtain an unconfounded treatment effect estimate.

However, at least two other strategies suggest themselves that may prove feasible. One alternative would be to develop a formula that captures any nonlinearity in the chosen metric of exposure prediction. This could permit the amount of expected amplification to be relatively accurately predicted over larger portions of the range.

The second alternative strategy would be to develop an “internal marker” covariate that would reflect how much increased imbalance in the nonincluded confounders is occurring. The internal marker would be a measured covariate deliberately left out of the propensity scores. The increase in its imbalance in Model 2 could be measured and serve as an indicator of confounding amplification.

Intuitively, an internal marker strategy has some attractive qualities, since it might sidestep any uncertainties about the relationship between confounding amplification and metrics of prediction of exposure. There is also already some evidence to support this approach. In the Brooks and Ohsfeldt simulation study¹, it was shown that covariates not included in the propensity score and that are uncorrelated with the included covariates all amplify to a remarkably similar extent, at least in that simulation. Thus, in principle, it appears feasible to use an “internal marker”, produced by withholding one (or a few) measured covariates from the Model 1 and Model 2 propensity scores, to track and estimate the general amount of confounding amplification. A key practical consideration in real-world datasets, however, is the need for these internal markers to have a minimal correlation with any of the covariates included in the propensity scores. Any such correlations might “constrain” the ability of the internal marker to reflect the degree of confounding amplification that is influencing the nonincluded confounders. (As an aside, these included-nonincluded covariate correlations would also similarly influence the confounding amplification of the nonincluded covariates that are not being used as internal markers. This interference, however, does not appear (based on present information) to be problematic for the method, for reasons discussed in [Appendix 2.2](#)).

Since some degree of correlation is unavoidable, then in the strictest sense the internal marker strategy may intrinsically underestimate true confounding amplification, although potentially only minimally.

Fortunately, this correlation is readily measurable. Therefore it should be possible to deliberately select the nonincluded measured covariate with the least correlation with the included covariates and the introduced variable to serve as an internal marker. Alternatively, simulation research may suggest quantitative approaches to correct for this correlation.

In summary, multiple aspects of confounding amplification are worthy of investigation. These include the presence and predictability of nonlinear relationships between the metric estimating the prediction of exposure and confounding amplification, and the potential strategies to address these nonlinearities.

Appendix 2.2. An initial exploration of the impact of correlation on confounding amplification

The data that exists to date from the Brooks and Ohsfeldt simulation⁵ indicates that confounding amplification is uniform between simulated covariates that were not included in the propensity score. However, this similarity may be a byproduct of their simulation. In theory, aspects of real-world data might create heterogeneities in amplification between covariates. One aspect, the effect upon confounding amplification of correlations between covariates (expected to be a common feature of real-world data), is considered below.

This conceptual exploration of the impact of correlations, which also draws upon aspects of the Brooks and Ohsfeldt simulation⁵, tentatively concludes that most correlations do not appear to interfere with the ACCE Method. A special case is posed by correlations between nonincluded covariates and the introduced variable. In that case, at least a partial remedy is intended to be provided by the adjustments performed in Steps 3 and 4 of the method.

Correlations can be categorized into five types, based on whether the correlated covariates are or are not included in the propensity score model. Correlations can exist between two nonincluded covariates, between a nonincluded and an included covariate, between nonincluded covariates and the introduced variable(s), between included covariates and the introduced variable(s), and between two included covariates. For the latter two categories, substantial amplification involving either of the correlated variables is not expected, at least in the same sense as the term applies to nonincluded covariates. Correlations between two nonincluded covariates also do not appear to be problematic. Both of the correlated variables would constitute part of the residual confounding being amplified, and thus be expected to be amplified to a similar extent, based on the Brooks and Ohsfeldt simulation⁵.

Correlations between included covariates and nonincluded covariates, however, seemingly could pose the possibility of creating constraints to amplification for certain nonincluded covariates. The measured covariates that are included in the propensity score cannot amplify substantially, and thereby might seem to constrain amplification, to a degree, of the correlated nonincluded variable. Upon further consideration, however, it appears that while the correlated nonincluded variable would be expected to have an *overall* change in imbalance in Model 2 that is less than the estimated confounding amplification, this might not interfere with the method's performance. The implications of the Brooks and Ohsfeldt simulation suggests that the change in a correlated variable can be alternatively modeled as a fraction that is largely unchanging between the model (to the degree that it is correlated with included covariates whose balance is not appreciably changing), and a fraction (alternatively, a “residual”) that amplifies as much as any uncorrelated nonincluded covariate ([Reference 2, Supplemental Appendices](#)). (These elements will be termed the “included” *fraction* and “nonincluded” *fraction*, of the nonincluded covariate, respectively).

It is important to recognize that part of the goal for the nested models in the ACCE Method is to minimize change in the balance of included covariates, to the extent feasible ([Appendix 1.1](#)). If the nested models do successfully exhibit little change in the balance of included covariates, then the Brooks and Ohsfeldt simulation⁵ would suggest that the amount of imbalance observed in the “included fraction” of the correlated, non-included variable also would not change substantively between models. This would leave the amplification of the nonincluded fraction (which Brooks and Ohsfeldt have observed amplifies as completely as for the uncorrelated, nonincluded covariates)⁵ as the only contribution to the *change* in treatment effect estimate attributable to this covariate. Thus, in principle, the method would still provide accurate final effect estimates. Further research regarding this conclusion, however, would clearly be beneficial.

This consideration of the effects of included covariate-nonincluded covariate correlations reinforces the need to keep the balance of the included covariates as similar as possible between the two models. It also brings to attention the fact that a small fraction of confounding would be expected to exist, attributable to the residual imbalance in the included covariates (and the correlated fraction of their nonincluded covariates), that is not detected because it does not amplify. Such confounding would be somewhat addressable by achieving as close a balance as feasible between treatment groups for any variable included in the propensity score models. The need for tight control of any included variables could also conceivably be yet another practical reason for limiting the included variables in the propensity scores (as discussed in Appendix 1.2.c and Appendix 2.1). Another alternative might be to include all of the included propensity score covariates, as is done in Step 3, in the outcome model estimating the Model 1 and Model 2 treatment effects (Step 1 of the method). This might largely address residual confounding arising from the remaining imbalance in these included covariates.

Correlations between nonincluded covariates and the introduced variable are a particularly distinct circumstance that can be considered as a special case. In this case, the introduced variable is an included variable in only one Model (Model 2). As a consequence, its balance is being deliberately changed from Model 1 to Model 2 to produce the needed confounding amplification. The “included fraction” of any nonincluded covariate relating to its correlation with the introduced variable(s) is now being controlled in Model 2 and thus more closely balanced in Model 2 than in Model 1. This is because in Model 1 the nonincluded covariate’s correlate, the yet-to-be introduced variable, is not controlled at all. However, it is partly for this circumstance that the Step 3 and 4 procedure was designed involving deriving regression coefficients and applying the Bross equation⁸. The intent of Steps 3 and 4 is that the change in confounding of the introduced variable is estimated, *along with* the change in confounding resulting from the change in imbalance in the fraction of the correlated variable(s) that is balanced as fully as the introduced variable. Whether the effect of correlation upon regression coefficients, however, is sufficiently similar to the effects of correlation on the balancing of covariates using a propensity score, to permit a generally effective adjustment in typical practice has not been determined. Conceptually, this regression-based approach appears to be a reasonable strategy to provide at least some first approximation of the changes in confounding associated with the correlated variables, but the strict accuracy of this approximation is unclear.

This is an area of the ACCE method in which further research would be particularly beneficial. The comparability of the quantitative effects of correlation on regression coefficients versus on covariate balance in propensity score analyses could be examined further through simulation, and perhaps theoretically through frameworks based on the general location model¹⁹ or other methods. Such simulations would helpfully allow the strength of the correlation between correlated variables and the amount of confounding amplification existing between the two models to be varied. Finally, a valuable role in validation may exist for real-world studies as well. These studies could investigate how frequently the ACCE Method provides what appears to be improved treatment effect estimates (i.e.,

closer to the result that is expected based on randomized trials)¹⁵ than typical propensity score or regression methods. Any imprecision in the ability of the adjustment in Steps 3 and 4 to adequately reflect the change between the models in confounding attributable to the covariates correlated with the introduced variable would depend on the number and strength of such correlations. While this is impossible to quantify the correlations present for truly unmeasured covariates in real world data, it may turn out, based on the comparisons to randomized data discussed above, that in practice these correlations often do not appear to be numerous or strong enough to substantially affect the method’s estimates. Also, even if the regression coefficient-based adjustment was ultimately shown to only poorly capture the effects of correlation, it is possible this may not interfere markedly with the overall accuracy of the method if most of the residual confounding is not correlated with the introduced variable. Further research is clearly warranted.

Even in the “worst case” scenario in which the adjustments in Step 3 and 4 of the method do not perform well *and* comparisons with randomized data suggested that this limitation typically impairs the method’s estimates substantially, three special circumstances exist in which the ACCE Method’s performance would not be generally expected to be adversely impacted by this limitation. These special circumstances would include the use of a true instrumental variable as the introduced variable (although it is uncertain whether any significant advantages would exist for the ACCE Method compared to conventional, 2-stage instrumental variable analysis). These special circumstances would also include introduced variables that are either near-instrumental variables or an additional category of variable: variables with an independent association with outcome but little correlation with other confounders. As long as the Bross equation⁸ adequately captured the effects upon confounding of increased control of this type of introduced variable upon confounding, then imprecisions in how the regression-coefficient based adjustment in Steps 3 and 4 captured the effect of correlated covariates would be relatively immaterial (since little correlation would be present). The frequency of such variables, however, is unclear. In addition, as pointed out above, it is impossible to determine whether a variable is correlated with unmeasured confounders in real-world data. However, a first step towards applying this variant would be confirming its lack of significant association with *any* the measured covariates available (although such a lack of correlation could not be taken as conclusive evidence of a lack of association with unmeasured covariates).

In addition, even if the ACCE Method was ultimately determined to typically provide estimates of total residual confounding that routinely have substantial imprecision, at least four beneficial applications to the method suggest themselves. The first is simply determining the direction of the remaining residual confounding in an association after efforts to control for confounding. Second, even imprecise estimates from the method may be able to provide indication of whether residual confounding appears to be a small, moderate, or large contributor to the observed treatment effect estimate. Along similar lines, associations between treatments and multiple outcomes could be able to be investigated, with the method providing information concerning which treatment-outcome associations appear to be more confounded, and which less confounded. The published data example examined in the manuscript, with its

suggestion that the statin-mortality analysis is less confounded than the statin-hip fracture analysis, could be seen as an example of this application. For these reasons, this method may be a particular benefit to database surveillance research that seeks to identify promising associations for further, detailed investigation. Finally, it is conceivable that by concentrating at least part of the uncertainty concerning residual confounding to a particular focus upon the introduced variable and its potential correlates, the method may suggest, in some instances, beneficial future investigations. This may be helpful, for instance, if additional information about the introduced variable and its likely correlates is expected to be more easily gathered (e.g., through chart review) for a particular dataset than efforts to obtain information about new potential confounders.

Appendix 2.3. Other considerations, such as the impact of the form (exponential versus linear) of the treatment effect estimate

Other important considerations in the application of the ACCE Method can be envisioned. It may, for instance, be determined that the method works better for linear, rather than nonlinear, outcome models (such as either continuous outcomes, or probabilities of dichotomous outcomes rather than the outcome itself, despite the well-known shortcomings of such an outcome measure). Some other innovations in nonrandomized treatment research have been introduced as approaches applicable to linear regression models²⁰.

Perhaps one valid measure of the value of the ACCE Method, however, should not be whether it derives an estimate that is 100% accurate in theory, but rather whether it, in practice, simply helps

to substantially further the goal of achieving nonrandomized study results that more closely approximate randomized trials. (To be specific, more closely approximating the findings which would be obtained if a randomized trial could be done on precisely that particular population). It is hoped that publication of this proposed method will spur research on this question as well as the other questions highlighted in these Appendices and the main manuscript.

Overall summary

As Appendix 1 and Appendix 2 have shown, there are a number of aspects of the ACCE Method that require attention for its optimal implementation. These are also the aspects of the method that most warrant research attention. Nevertheless, the ACCE Method may provide a novel approach for estimating residual confounding either quantitatively or qualitatively, and thus provide treatment effect estimates that may be an improvement over what has been achieved by conventional propensity score or regression methods. Stated another way, although numerous potential complications for the ACCE Method can be envisioned, the practical significance of these potential complications (i.e., the extent to which they routinely interfere with the accuracy of the estimates from the ACCE Method) remains to be determined. Most existing comparative effectiveness techniques are unable to address unmeasured confounding quantitatively or estimate residual confounding easily. Therefore, even if the ACCE Method provides only a very crude estimate of residual confounding and of an unconfounded treatment effect estimate, it may prove a substantial and highly useful advance on current methodology. This potential promise of the ACCE Method to provide an estimate of something largely not estimated by many current methodologies clearly justifies its further investigation.

Supplementary Table 1. Step-by-Step Application of the ACCE Method (Hypothetical Example).

Scenario: Initially (i.e., in Model 1), a genuine treatment effect odds ratio (OR) = 1.10 exists for the investigated intervention (e.g., medication, surgery, psychotherapy, etc.), *concealed* by a larger amount of residual confounding (OR = 1.15). This combination of genuine treatment effect and baseline confounding produces a biased association between the intervention and outcome (treatment effect estimate of OR = 1.265). Model 1 has an R^2 of 0.25.

Applying the ACCE Method, an additional variable or variable(s) is identified that is substantially associated with treatment. This identified variable has an association of $RR = 1.05$ with the outcome, and a 4:1 imbalance (80% versus 20%) between the treatment groups in Model 1. Upon introduction into Model 2, this imbalance changes to a 1.08:1 imbalance (52% versus 48%) once this variable is included in the propensity score and balanced through stratification or matching. Model 2 (which has an R^2 of 0.5) has a treatment effect estimate of OR = 1.2985.

Step	Description	Verbal and Symbolic Formula	Example
1a	Construct Model 1 ("M1") & derive its Treatment Effect Estimate	Model 1 Treatment Effect Estimate = EE_{M1} Model 1 prediction of exposure metric value = a	$EE_{M1} = OR_{M1} = 1.265$ <i>alternatively:</i> $EE_{M1} = \beta_{M1} = \ln(1.265) = 0.2351$ $a = 0.25$ (R^2)
1b	Construct Model 2 ("M2") by adding variable(s) to Model 1 & derive its Treatment Effect Estimate	Model 2 Treatment Effect Estimate = EE_{M2} Model 2 Prediction of Exposure metric value = b	$EE_{M2} = OR_{M2} = 1.2985$ <i>alternatively:</i> $EE_{M2} = \beta_{M2} = \ln(1.2985) = 0.2612$ $b = 0.5$ (R^2)
2	Estimate Confounding Amplification ("CAmp") (either through use of a Prediction of Exposure metric or an "Internal Marker") ^a	For R^2 , if both $R^2 < 0.56$, THEN: Confounding Amplification = Unexplained Variance of Exposure _{Model 1} / Unexplained Variance of Exposure _{Model 2} (1-a)/(1-b) = CAmp	CAmp = $0.75/0.5 = 1.5$

Step	Description	Verbal and Symbolic Formula	Example
Appendix Table 1 (continued)			
3a	Determine if an association ("IntV:O") exists between the Introduced Variable ^b ("IntV") and the Outcome ("O") by examining the association within the treatment arms for each Model ^c	<p>IntV:Outcome Association = IntV:O or, as mean value, "OR_{i:o}"</p> <p>More specifically:</p> $\text{IntV:O}_{\text{Model 1 Treatment Group 1}} = \text{IntV:O}_{\text{M1 Tx Grp 1}}$ $\text{IntV:O}_{\text{Model 1 Treatment Group 2}} = \text{IntV:O}_{\text{M1 Tx Grp 2}}$ <p>AND</p> $\text{IntV:O}_{\text{Model 2 Treatment Group 1}} = \text{IntV:O}_{\text{M2 Tx Grp 1}}$ $\text{IntV:O}_{\text{Model 2 Treatment Group 2}} = \text{IntV:O}_{\text{M2 Tx Grp 2}}$	$\text{IntV:O}_{\text{M1 Tx Group 1}} = 1.045 \text{ (as OR)}$ $\text{IntV:O}_{\text{M1 Tx Group 2}} = 1.055 \text{ (as OR)}$ $\text{Mean IntV:O}_1 = \text{OR}_{\text{IoM1}} = 1.05^{\text{d}}$ $\text{IntV:O}_{\text{M2 Tx Group 1}} = 1.046 \text{ (as OR)}$ $\text{IntV:O}_{\text{M2 Tx Group 2}} = 1.056 \text{ (as OR)}$ $\text{Mean IntV:O}_2 = \text{OR}_{\text{IoM2}} = 1.051^{\text{d}}$
3b	Determine the degree to which the change in Treatment Effect Estimate observed between Model 1 and Model 2 is attributable to the increased balance in the Introduced Variable (resulting from stratification or matching on the propensity score). This is done through use of the Bross equation ^e . (i.e., how much imbalance occurs in the Introduced Variable in Model 2 * the coefficient from Substep 3a minus the original imbalance in Model 1 * the coefficient from Substep 3a)	<p>Estimated the Confounding attributable to Original (Model 1) Imbalance in Introduced Variable ("CIntV_{M1}"): p = probability (e.g., 80%)</p> $\ln [(p_{\text{M1 Tx Grp 1}} * (\text{OR}_{\text{IoM1}} - 1) + 1) / (p_{\text{M1 Tx Grp 2}} * (\text{OR}_{\text{IoM1}} - 1) + 1)]$ <p>Estimated Confounding attributable to the Subsequent (Model 2) Imbalance in Introduced Variable ("CIntVM2")</p> $\ln [(p_{\text{M2 Tx Grp 1}} * (\text{OR}_{\text{IoM2}} - 1) + 1) / (p_{\text{M2 Tx Grp 2}} * (\text{OR}_{\text{IoM2}} - 1) + 1)]$ <p>Change in Effect Estimate (Model 2 - Model 1) attributable to increased balance of Introduced Variable ("Δ_{EE(IntV)}"): CIntV_{M2} - CIntV_{M1} = Δ_{EE(IntV)}</p>	$\text{CIntV}_{\text{M1}} = \ln[(0.8 * (1.05-1)+1) / (0.2 * (1.05-1)+1)] = 0.0293$ $\text{CIntV}_{\text{M2}} = \ln[(0.52 * (1.051-1)+1) / (0.48 * (1.051-1)+1)] = 0.002$ $\Delta_{\text{EE(IntV)}} = 0.002 - 0.0293 = -0.0291$
3c	Calculate the Change in the observed Treatment Effect Estimate (Model 2 versus Model 1)	<p>Model 2 Effect Estimate – Model 1 Effect Estimate = Change in Effect Estimate ("Δ_{EE}"): EE_{M2} – EE_{M1} = Δ_{EE}</p>	$\Delta_{\text{EE}} = 0.2612 - 0.2351 = 0.0261$
3d	Adjust the Change in the observed Treatment Effect Estimate from Model 1 to Model 2 by the amount of change accounted for by increased balance in the Introduced Variable	<p>Change in Effect Estimate (Model 2 - Model 1) – Difference in Treatment Effect Estimate attributable to the Introduced Variable = Adjusted Treatment Effect Estimate Change ("AdjΔ_{EE}"): Δ_{EE} – Δ_{EE(IntV)} = AdjΔ_{EE}</p>	$\text{Adj}\Delta_{\text{EE}} = 0.0261 - (-0.0291) = 0.0552$
4a	Calculate estimate of Residual Confounding in Model 1 except for the Introduced Variable by dividing the adjusted change in Treatment Effect Estimate by the amount of confounding amplification (beyond 1.0)	<p>Adjusted Treatment Effect Estimate Change/ (Confounding Amplification – 1) = Residual Confounding_{Model 1 except for IntV} ("CRes_{M1-IntV}"): AdjΔ_{EE} / (CAMP - 1) = CRes_{M1-IntV}</p>	$\text{CRes}_{\text{M1-IntV}} = 0.0552 / (1.5 - 1) = 0.1105$
4b1	Derive an estimate of Total Residual Confounding in Model 1	<p>Residual Confounding_{Model 1 except for IntV} + IntV_{M1} Confounding = Total Residual Confounding_{Model 1} ("CTotRes_{M1}"): CRes_{M1-IntV} + CIntV_{M1} = CTotRes_{M1}</p>	$\text{CRes}_1 = 0.1105 + 0.0293 = 0.1398$ $e^{0.1398} = 1.15$ <p>OR_{Total Confounding (Model 1)} = 1.15^f</p>

Step	Description	Verbal and Symbolic Formula	Example
Appendix Table 1 (continued)			
4b2	Derive an Estimate of the Unconfounded Treatment Effect	Model 1 Effect Estimate – Total Residual Confounding _{Model 1} = Unconfounded Treatment Effect Estimate (“EE _{Unc} ”): $EE_{M1} - CTotRes_{M1} = E_{EUnC}$	$EE_{Unc} = 0.2351 - 0.1398 = 0.0953$ $e^{0.0953} = 1.10$ OR_{Unconfounded (Model 1)} = 1.10^f

^a “Internal Marker” = a measured covariate deliberately not included in the propensity score that is generally uncorrelated with other propensity score covariates. The internal marker serves to index the amount of confounding amplification between treatment groups that occurs between Model 1 and Model 2. If an internal marker is used, then Confounding Amplification (Camp) = Final Internal Marker Covariate Imbalance/Initial Internal Marker Covariate Imbalance.

^b For all associations involving the Introduced Variable, the association would include the association of the Introduced Variable plus the associations of its correlates, to the extent that these associations influence the observed association between the Introduced Variable and outcome. When balance in the Introduced Variable is referenced, this also refers to balance in both the Introduced Variable and, to a lesser extent, its correlates.

^c Either within-treatment arm or overall regressions can be performed. Examining the association within treatment arms prevents the association with the intervention, which may be substantial, from influencing the estimation of the IntV-Outcome association. The association between Introduced Variable and Outcome is an aggregate of direct and indirect associations. This aggregate association is then used in Step 3b to estimate the quantitative effect on the treatment effect estimate of adding the introduced variable into the propensity score (and increasing its balance through stratification or matching) that is independent of confounding amplification.

^d Based on averaging the coefficients (i.e., ln(OR)). The most straightforward circumstance for the within-treatment arm approach is if the observed association is highly comparable in both treatment arms and both models. If so, as an approximation these values can be averaged. In this hypothetical example, the observed association is varied slightly to illustrate that it can differ between the models.

^e Since the Introduced Variable(s) is a measured covariate or covariates, it is possible to determine its initial imbalance in Model 1, and the degree to which this imbalance changes in Model 2. This information can then be combined (through use of the Bross equation) with the coefficients derived in Step 3a to estimate the component of the change in the Treatment Effect Estimate between Model 1 and Model 2 that is attributable to increased balance in the Introduced Variable(s) and its correlates.

^f The congruence between the scenario's genuine treatment effect and total confounding and these values should not be seen as validating the method. The scenario's effects estimates were selected based upon what would be expected from confounding amplification operating consistent with the system described here. However, this step-by-step example does illustrate the mechanics of how, absent the effects of random variability, this series of calculations would function to produce the desired values (i.e., Total Residual Confounding and an Unconfounded Treatment Effect Estimate) from the initial values (i.e., the confounded Model 1 and 2 treatment effect estimates and an estimate of confounding amplification).

References

- Bhattacharya J, Vogt W: **Do instrumental variables belong in propensity scores?** In: *NBER Technical Working Paper no 343*. Cambridge, MA: National Bureau of Economic Research. 2007.
[Reference Source](#)
- Wooldridge J: **Should instrumental variables be used as matching variables?** East Lansing, MI: Michigan State University; Unpublished manuscript. Accessed July 21, 2014. 2009.
[Reference Source](#)
- Pearl J: **On a class of bias-amplifying variables that endanger effect estimates.** In: *Proceedings of the Twenty-Sixth Conference on Uncertainty in Artificial Intelligence (UAI 2010)*; Corvallis, OR: Association for Uncertainty in Artificial Intelligence; Accessed November 8, 2013. 2010; 2425–2432.
[Reference Source](#)
- Pearl J: **Invited commentary: understanding bias amplification.** *Am J Epidemiol.* 2011; **174**(11): 1223–1227; discussion pg 1228–1229.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brooks JM, Ohnsfeldt RL: **Squeezing the balloon: propensity scores and unmeasured covariate balance.** *Health Serv Res.* 2013; **48**(4): 1487–1507.
[PubMed Abstract](#) | [Publisher Full Text](#)
- DeMaris A: **Explained variance in logistic regression: A Monte Carlo study of proposed measures.** *Social Methods Res.* 2002; **31**(1): 27–74.
[Publisher Full Text](#)
- Steyerberg EW, Vickers AJ, Cook NR, et al.: **Assessing the performance of prediction models: a framework for traditional and novel measures.** *Epidemiology.* 2010; **21**(1): 128–138.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Bross ID: **Spurious effects from an extraneous variable.** *J Chronic Dis.* 1966; **19**(6): 637–647.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Schneeweiss S, Rassen JA, Glynn RJ, et al.: **High-dimensional propensity score adjustment in studies of treatment effects using health care claims data.** *Epidemiology.* 2009; **20**(4): 512–522.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Patrick AR, Schneeweiss S, Brookhart MA, et al.: **The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration.** *Pharmacoepidemiol Drug Saf.* 2011; **20**(6): 551–559.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- 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.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Toh S, Hernandez-Diaz S: **Statins and fracture risk. A systematic review.** *Pharmacoepidemiol Drug Saf.* 2007; **16**(6): 627–640.
[PubMed Abstract](#) | [Publisher Full Text](#)
- Sturmer T, Schneeweiss S, Rothman KJ, et al.: **Performance of propensity score calibration—a simulation study.** *Am J Epidemiol.* 2007; **165**(10): 1110–1118.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Brookhart MA, Schneeweiss S, Rothman KJ, et al.: **Variable selection for propensity score models.** *Am J Epidemiol.* 2006; **163**(12): 1149–1156.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Schneeweiss S, Patrick AR, Sturmer T, et al.: **Increasing levels of restriction in pharmacoepidemiologic database studies of elderly and comparison with randomized trial results.** *Med Care.* 2007; **45**(10 Supl 2): S131–142.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Sturmer T, Rothman KJ, Avorn J, et al.: **Treatment effects in the presence of unmeasured confounding: dealing with observations in the tails of the propensity score distribution—a simulation study.** *Am J Epidemiol.* 2010; **172**(7): 843–854.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Hernan MA, Robins JM: **Authors' response, part I: observational studies analyzed like randomized experiments: best of both worlds.** *Epidemiology.* 2008; **19**(6): 789–792.
[Publisher Full Text](#)
- Toh S, Garcia Rodriguez LA, Hernan MA: **Confounding adjustment via a semi-automated high-dimensional propensity score algorithm: an application to electronic medical records.** *Pharmacoepidemiol Drug Saf.* 2011; **20**(8): 849–857.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)
- Oikari J, Tate RF: **Multivariate correlation models with mixed discrete and continuous variables.** *Ann Math Statist.* 1961; **32**(2): 448–465.
[Publisher Full Text](#)
- VanderWeele TJ, Shpitser I: **A new criterion for confounder selection.** *Biometrics.* 2011; **67**(4): 1406–13.
[PubMed Abstract](#) | [Publisher Full Text](#) | [Free Full Text](#)

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 05 January 2015

<https://doi.org/10.5256/f1000research.5125.r7091>

© 2015 Matthews G. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Gregory Matthews

Department of Mathematics and Statistics, Loyola University Chicago, Chicago, IL, USA

The authors present a manuscript describing a procedure that allows for the quantification of the total amount of residual confounding prior to bias amplification caused by propensity score models. I believe the procedure described is reasonable, and my biggest concerns with this manuscript are the presentation of the approach, which I had a hard time following initially. I think this paper is deserving of indexing as it is, but could be substantially improved with clearer presentation.

Specific Comments:

- The authors talk about creating two models (Model 1 and Model 2) that are nest within each other in such a way that Model 2 contains all the variables in Model 1 plus one/several extra variable/s. It seems like there are many choices for this extra variable/s from among the possible variables. Do the authors have any specific advice on how this or these should be chosen? They do mention that this variable should be chosen to have "discernible confounding amplification", but isn't it possible that there are many acceptable choices that will satisfy this criteria? In that case is there any advice on how to choose between the good candidate variables?
- In Step 2 of the description of the method, the authors mention that when R^2 is between 0.04 and 0.56 there is a linear relationship between unexplained variance and confounding amplification. I believe that this threshold is then used in Supplementary table 1 when they state that the step should be taken only if R^2 is less than 0.56. Should this step not be taken if R^2 is less than 0.04? Do the authors have any advice on what to do when R^2 is greater than 0.56?

Minor Comments:

- Should the outcome in Table 1B be hip fracture rather than all cause mortality?
- Supplementary Table 1, 3a I think this is a typo: "IntV:"

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 02 Mar 2015

Eric Smith, Edith Nourse Rogers Memorial Medical Center, Bedford, USA

I would like to thank both reviewers for their thoughtful, insightful, and encouraging reviews. I particularly appreciate their openness to a new methodology to attempt to estimate residual/unmeasured confounding. I am very glad to see that they recognized the value in disseminating and exploring a methodology that takes a very different approach (and possibly an approach that is more broadly applicable) than some of the limited number of alternatives currently available to tackle the problem of unmeasured confounding. Their specific comments were also extremely valuable.

Both reviewers suggested that the manuscript would benefit from greater clarity; therefore I have revised and enhanced the presentation of the method quite substantially. The major ways I have done this is to: 1) expand the description of the method in the text and adding cross-references to the exact steps in the Appendix Table (which has also been expanded); 2) adding 3 additional hypothetical examples to communicate more incrementally the rationale for the method; 3) reorganized the manuscript Table so it reads more vertically than horizontally; 4) attempted to be more precise and detailed in my language; and, perhaps most importantly, 5) expressed the entire method mathematically in a single Summary Equation to help facilitate its understanding. The main manuscript text is substantially longer as a result of this increased explanation, but hopefully less ambiguous at key points. Some of the increase in length results from the more detailed description of the method, but much of the increase relates to the more detailed hypothetical examples, which some readers may not even feel a need to review. Similarly, the Appendices are considerably longer, but the reader is encouraged to pick and choose whether they want to review some, none, or all of these based entirely on their interest.

Another important comment was Dr. Lunt's comment that considerably further work needed to be done on the method. I couldn't agree more, and it is my hope that the dividend that results from laying out the method in such detail is that multiple research groups can quickly advance this research. As I try to anticipate and highlight as fully as possible, there are a number of important uncertainties. These uncertainties range from such fundamental points as how consistently predictable the phenomenon of confounding amplification actually is, how accurately the difference between effect estimates can be determined, and how accurate are the proposed Bross equation-based corrections for the contribution of the Introduced Variable and, to a partial degree, its correlates, on the estimates of the change in treatment effect estimate as well as the starting Model 1 treatment effect estimate. Indeed, it is not even certain whether the method can be applied to some common logistic model effect estimates (e.g., odds ratio). I have even identified two more potential sources of uncertainty that are now included and discussed in the text and appendices: whether the introduced variable-outcome regression coefficient would potentially also suffer from at least some confounding amplification, and whether possible

“constraints” might exist to achievable confounding amplification in real-world settings. So I am in complete agreement with Dr. Lunt that this manuscript represents only the very start of what hopefully will be steady advance of knowledge about this method and its value relative to other proposed approaches addressing unmeasured confounding. To my point of view, this is all the more reason to seek to enlist the greater research community in this effort.

Nevertheless, it is important to note that approaches suggest themselves to address or minimize many of these uncertainties, although much investigation is needed. In addition, I want to emphasize a key point: while a number of uncertainties exist relevant to the actual performance of the method, it is my intention that, with this version of the manuscript, that there be no substantial uncertainty concerning the specific approach that is actually being proposed. I paid close attention to the fact that Dr. Matthews and Dr. Lunt (who has published on bias amplification) appeared uncertain about how to apply the method as described in Version 1. I hope in this version that I have communicated the method clearly enough that the vital next step can take place: testing the method in simulated and real-world datasets.

It is for this reason – to facilitate the ability of as many interested research teams as possible to contribute to the method’s evaluation and evolution – that I have taken particular pains to expand communication concerning the overall logic, and underlying rationale, of the method and each of its steps. There are certainly places in which my proposed solutions to potential challenges for the method may prove imperfect or suboptimal (some possibilities might include the use of a regression coefficient and the Bross equation to take account confounding from the Introduced Variable-outcome relationship, the suggested approach to addressing possible confounding amplification in the Introduced Variable-outcome coefficient, and/or the favoring of stratification over matching to increase comparability of Model 1 and Model 2 mentioned in Appendix 2). It is my firm hope that other research groups can contribute by suggesting other approaches to accomplishing that particular objective within in the method, or even other angles concerning how to exploit confounding amplification to help estimate residual confounding. Therefore I wanted to be particularly clear in explaining the method so that the objective to be accomplished in each step was clear. This communication has been done through expanded text, calculations, examples, metaphors, technical Appendices, and the Summary Equation. I also outline the clear initial and subsequent steps for research as I see them (most centered on simulation) in the Discussion. Hopefully the manuscript is now sufficiently clearer so that collaborative investigation and elaboration of this method can take place.

I thank the reviewers for encouraging me to much more carefully clarify the logic and approach of the method, and I hope they think that I have succeeded in that task.

In closing, I would like to address the remaining specific points brought up by the reviewers:

Dr. Lunt (Reviewer 1):

1. As mentioned above, I am extremely grateful for Dr. Lunt's observation for noting that the denominator of equations 3-6 in Reference 4 (Pearl, 2011) does indeed appear to support the $1-R^2$ relationship predicting the proportional amount of confounding amplification separate from the Brooks and Ohsfeldt (2013) simulation. This is potentially quite important, for it suggests that application of the technique might not need to be limited to an R^2 of ≤ 0.56 (one of the concerns of the 2nd reviewer, Dr. Matthews). It does, however, increase the need to understand why the Brooks and Ohsfeldt simulation begins to exhibit nonlinear confounding amplification above R^2 of 0.56.

Dr. Matthews (Reviewer 2):

1. Dr. Matthews asked a number of helpful questions concerning important details involved in implementing the method that I see now were not addressed as directly and thoroughly as they might have been. So that many readers can easily benefit from his helpful inquiry concerning recommendations on how to choose an instrumental variable without having to access my response to this comment, I have added an entire Appendix (Appendix 7) devoted in large part to this topic. In addition to offering practical suggestions on implementing the method, based on current knowledge, this Appendices also attempts to anticipate the likely trade-offs involved in optimizing one characteristic of the method potentially at the cost of another characteristic (e.g., wanting to maximize confounding amplification while minimizing differences between the two models that are separate from confounding amplification).
2. Regarding Dr. Matthew's 2nd major point, the simulation research that I hope follows this manuscript will likely provide the best guidance on what approaches should be taken if the R^2 is < 0.04 or > 0.56 . It should be noted, however, that, until that research is available, it is to be hoped that almost all propensity score models will succeed in achieving an R^2 of at least 0.04. Furthermore, one remedy for circumstances in which Model 2 exceeds an R^2 of 0.56 seemingly would be simply to remove measured covariates from the propensity score model until Model 2's R^2 is ≤ 0.56 . This is a pragmatic, but not a perfect solution, since as pointed out in Appendix 3.2, such a step places extra weight on the method achieving an accurate estimate of residual/unmeasured confounding, since more of that type of confounding now exists. Also, as discussed in Appendix 4, if variables have to be removed from the propensity score, priority should be given to removing variables with little or no correlation with the Introduced Variable(s) and retaining in the propensity scores, to the extent possible, variables that correlate with the Introduced Variable(s).
3. I also thank Dr. Matthews for pointing out the mislabeling of the outcome in Table 1. As mentioned, in addition to correcting this error, I have entirely restructured this Table to make it read more vertically than horizontally, at least in regard to the information pertaining to Model 1 versus Model 2.
4. Regarding the "IntV" terminology in Supplementary Table 1, I have retained this

abbreviation. "IntV" is my attempt to propose a nomenclature (abbreviation) for the introduced variable that will separate it from instrumental variables (which, unfortunately, share the same initials). "InV" might also be useable, but I felt the extra letter of "IntV" as an abbreviation for the term "Introduced Variable" made sense because the abbreviation was less likely to appear to be simply an erroneous typing of "IV."

I have also made the following minor changes:

1. Capitalized "Introduced Variable(s)" to make each of its mentions more noticeable, since this variable or variables plays a key role in the method.
2. Expanded the discussion of the potential impacts of correlations between various types of variables on the method's estimates, and added Appendices that explore potential threats to the accuracy of the Introduced Variable-outcome regression coefficient, that provide explanation of the method's components (and key uncertainties) in reference to the terms of the ACCE Method Summary Equation, and that begin to explore the use of sets of Introduced Variables and the practical trade-offs to be considered when implementing the method.
3. Tried to be consistent with my language concerning "confounding amplification": "proportional confounding amplification" refers to the percentage increase in residual confounding predicted by $1-R^2$, some other measure of exposure prediction, or an internal marker, while "quantitative confounding amplification" refers to the numerical change in the treatment effect estimate (technically, the change in the treatment effect estimate adjusted for the impact of increased balance in the Introduced Variable(s)).
4. Replaced the term "multiple" Introduced Variable(s) with the term "set of Introduced Variables" to make it clearer I am referring to simultaneous insertion of several to many Introduced Variables, rather than the sequential use of different single Introduced Variables.
5. Clearly labeled the Hypothetical Examples as Hypothetical Examples, moving them out of "Results."
6. Changed the examples from "odds ratio" to "risk ratio" due to concerns that noncollapsibility of the odds ratio might interfere with the subtraction of the Model 1 and Model 2 treatment effect estimates necessary to estimate the quantitative effect of confounding amplification.
7. Invented the term "amplifiable fraction of residual confounding" to hopefully better communicate that (if the Introduced Variable(s) has any association with outcome) it is only the residual confounding separate from that which is attributable to the Introduced Variable(s) (which is not amplified) that is able to be amplified. Hopefully this has made this clearer.
8. Removed the somewhat redundant word "Supplementary" from "Supplementary

Appendix Table.”

9. Corrected a minor subtraction error in the Appendix Table, Equation 3b (and subsequent steps), that had no substantive impact on the estimates of total residual confounding and the unconfounded treatment effect estimate. Also corrected a notation error in Step 4a where “M2” had been written “M3” by mistake.

Competing Interests: No competing interests were disclosed.

Reviewer Report 27 November 2014

<https://doi.org/10.5256/f1000research.5125.r6843>

© 2014 Lunt M. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mark Lunt

Arthritis Research UK Epidemiology Unit, University of Manchester, Manchester, UK

This article outlines a very interesting approach to using propensity score methods to correct for unmeasured confounding. That was not the aim of the propensity score, and current methods are not able to do this, so it potentially represents a considerable advance.

The idea is conceptually a simple one, related to the well-established use of instrumental variables to control for unmeasured confounding. However, I have not come across this idea before, and the author is to be congratulated on his originality.

Having said that, I was a little disappointed in the presentation of the method. I do not feel that I am in a position to apply this method to any of my own data. One reason that I took so long over the review was that I wanted to be certain I fully understood the method by applying it myself, but it has become obvious that I will not be able to in a reasonable timescale.

Greater precision in the presentation would have been welcome, whether that was explicit mathematical formulae, or simply causal diagrams showing how the various biases arose and which causal paths contributed to which estimates. This has been done very well in some of the references. For example, the relation between bias amplification and $1-R^2$ is given a clear mathematical basis in reference 4, and I would regard this as more convincing than simulation evidence.

I'm sure that the author would agree with me that there is a lot of work to be done on this method before it can be applied routinely. I hope that this paper does spark that research, and that the author gets the credit he deserves for coming up with this potentially very useful idea.

Competing Interests: No competing interests were disclosed.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Author Response 02 Mar 2015

Eric Smith, Edith Nourse Rogers Memorial Medical Center, Bedford, USA

I would like to thank both reviewers for their thoughtful, insightful, and encouraging reviews. I particularly appreciate their openness to a new methodology to attempt to estimate residual/unmeasured confounding. I am very glad to see that they recognized the value in disseminating and exploring a methodology that takes a very different approach (and possibly an approach that is more broadly applicable) than some of the limited number of alternatives currently available to tackle the problem of unmeasured confounding. Their specific comments were also extremely valuable.

Both reviewers suggested that the manuscript would benefit from greater clarity; therefore I have revised and enhanced the presentation of the method quite substantially. The major ways I have done this is to: 1) expand the description of the method in the text and adding cross-references to the exact steps in the Appendix Table (which has also been expanded); 2) adding 3 additional hypothetical examples to communicate more incrementally the rationale for the method; 3) reorganized the manuscript Table so it reads more vertically than horizontally; 4) attempted to be more precise and detailed in my language; and, perhaps most importantly, 5) expressed the entire method mathematically in a single Summary Equation to help facilitate its understanding. The main manuscript text is substantially longer as a result of this increased explanation, but hopefully less ambiguous at key points. Some of the increase in length results from the more detailed description of the method, but much of the increase relates to the more detailed hypothetical examples, which some readers may not even feel a need to review. Similarly, the Appendices are considerably longer, but the reader is encouraged to pick and choose whether they want to review some, none, or all of these based entirely on their interest.

Another important comment was Dr. Lunt's comment that considerably further work needed to be done on the method. I couldn't agree more, and it is my hope that the dividend that results from laying out the method in such detail is that multiple research groups can quickly advance this research. As I try to anticipate and highlight as fully as possible, there are a number of important uncertainties. These uncertainties range from such fundamental points as how consistently predictable the phenomenon of confounding amplification actually is, how accurately the difference between effect estimates can be determined, and how accurate are the proposed Bross equation-based corrections for the contribution of the Introduced Variable and, to a partial degree, its correlates, on the estimates of the change in treatment effect estimate as well as the starting Model 1 treatment effect estimate. Indeed, it is not even certain whether the method can be applied to some common logistic model effect estimates (e.g., odds ratio). I have even identified two more potential sources of uncertainty that are now included and discussed in the text and appendices: whether the introduced variable-outcome regression coefficient would potentially also suffer from at least some confounding amplification, and whether possible

“constraints” might exist to achievable confounding amplification in real-world settings. So I am in complete agreement with Dr. Lunt that this manuscript represents only the very start of what hopefully will be steady advance of knowledge about this method and its value relative to other proposed approaches addressing unmeasured confounding. To my point of view, this is all the more reason to seek to enlist the greater research community in this effort.

Nevertheless, it is important to note that approaches suggest themselves to address or minimize many of these uncertainties, although much investigation is needed. In addition, I want to emphasize a key point: while a number of uncertainties exist relevant to the actual performance of the method, it is my intention that, with this version of the manuscript, that there be no substantial uncertainty concerning the specific approach that is actually being proposed. I paid close attention to the fact that Dr. Matthews and Dr. Lunt (who has published on bias amplification) appeared uncertain about how to apply the method as described in Version 1. I hope in this version that I have communicated the method clearly enough that the vital next step can take place: testing the method in simulated and real-world datasets.

It is for this reason – to facilitate the ability of as many interested research teams as possible to contribute to the method’s evaluation and evolution – that I have taken particular pains to expand communication concerning the overall logic, and underlying rationale, of the method and each of its steps. There are certainly places in which my proposed solutions to potential challenges for the method may prove imperfect or suboptimal (some possibilities might include the use of a regression coefficient and the Bross equation to take account confounding from the Introduced Variable-outcome relationship, the suggested approach to addressing possible confounding amplification in the Introduced Variable-outcome coefficient, and/or the favoring of stratification over matching to increase comparability of Model 1 and Model 2 mentioned in Appendix 2). It is my firm hope that other research groups can contribute by suggesting other approaches to accomplishing that particular objective within in the method, or even other angles concerning how to exploit confounding amplification to help estimate residual confounding. Therefore I wanted to be particularly clear in explaining the method so that the objective to be accomplished in each step was clear. This communication has been done through expanded text, calculations, examples, metaphors, technical Appendices, and the Summary Equation. I also outline the clear initial and subsequent steps for research as I see them (most centered on simulation) in the Discussion. Hopefully the manuscript is now sufficiently clearer so that collaborative investigation and elaboration of this method can take place.

I thank the reviewers for encouraging me to much more carefully clarify the logic and approach of the method, and I hope they think that I have succeeded in that task.

In closing, I would like to address the remaining specific points brought up by the reviewers:

Dr. Lunt (Reviewer 1):

1. As mentioned above, I am extremely grateful for Dr. Lunt's observation for noting that the denominator of equations 3-6 in Reference 4 (Pearl, 2011) does indeed appear to support the $1-R^2$ relationship predicting the proportional amount of confounding amplification separate from the Brooks and Ohsfeldt (2013) simulation. This is potentially quite important, for it suggests that application of the technique might not need to be limited to an R^2 of ≤ 0.56 (one of the concerns of the 2nd reviewer, Dr. Matthews). It does, however, increase the need to understand why the Brooks and Ohsfeldt simulation begins to exhibit nonlinear confounding amplification above R^2 of 0.56.

Dr. Matthews (Reviewer 2):

1. Dr. Matthews asked a number of helpful questions concerning important details involved in implementing the method that I see now were not addressed as directly and thoroughly as they might have been. So that many readers can easily benefit from his helpful inquiry concerning recommendations on how to choose an instrumental variable without having to access my response to this comment, I have added an entire Appendix (Appendix 7) devoted in large part to this topic. In addition to offering practical suggestions on implementing the method, based on current knowledge, this Appendices also attempts to anticipate the likely trade-offs involved in optimizing one characteristic of the method potentially at the cost of another characteristic (e.g., wanting to maximize confounding amplification while minimizing differences between the two models that are separate from confounding amplification).
2. Regarding Dr. Matthew's 2nd major point, the simulation research that I hope follows this manuscript will likely provide the best guidance on what approaches should be taken if the R^2 is < 0.04 or > 0.56 . It should be noted, however, that, until that research is available, it is to be hoped that almost all propensity score models will succeed in achieving an R^2 of at least 0.04. Furthermore, one remedy for circumstances in which Model 2 exceeds an R^2 of 0.56 seemingly would be simply to remove measured covariates from the propensity score model until Model 2's R^2 is ≤ 0.56 . This is a pragmatic, but not a perfect solution, since as pointed out in Appendix 3.2, such a step places extra weight on the method achieving an accurate estimate of residual/unmeasured confounding, since more of that type of confounding now exists. Also, as discussed in Appendix 4, if variables have to be removed from the propensity score, priority should be given to removing variables with little or no correlation with the Introduced Variable(s) and retaining in the propensity scores, to the extent possible, variables that correlate with the Introduced Variable(s).
3. I also thank Dr. Matthews for pointing out the mislabeling of the outcome in Table 1. As mentioned, in addition to correcting this error, I have entirely restructured this Table to make it read more vertically than horizontally, at least in regard to the information pertaining to Model 1 versus Model 2.
4. Regarding the "IntV" terminology in Supplementary Table 1, I have retained this

abbreviation. "IntV" is my attempt to propose a nomenclature (abbreviation) for the introduced variable that will separate it from instrumental variables (which, unfortunately, share the same initials). "InV" might also be useable, but I felt the extra letter of "IntV" as an abbreviation for the term "Introduced Variable" made sense because the abbreviation was less likely to appear to be simply an erroneous typing of "IV."

I have also made the following minor changes:

1. Capitalized "Introduced Variable(s)" to make each of its mentions more noticeable, since this variable or variables plays a key role in the method.
2. Expanded the discussion of the potential impacts of correlations between various types of variables on the method's estimates, and added Appendices that explore potential threats to the accuracy of the Introduced Variable-outcome regression coefficient, that provide explanation of the method's components (and key uncertainties) in reference to the terms of the ACCE Method Summary Equation, and that begin to explore the use of sets of Introduced Variables and the practical trade-offs to be considered when implementing the method.
3. Tried to be consistent with my language concerning "confounding amplification": "proportional confounding amplification" refers to the percentage increase in residual confounding predicted by $1-R^2$, some other measure of exposure prediction, or an internal marker, while "quantitative confounding amplification" refers to the numerical change in the treatment effect estimate (technically, the change in the treatment effect estimate adjusted for the impact of increased balance in the Introduced Variable(s)).
4. Replaced the term "multiple" Introduced Variable(s) with the term "set of Introduced Variables" to make it clearer I am referring to simultaneous insertion of several to many Introduced Variables, rather than the sequential use of different single Introduced Variables.
5. Clearly labeled the Hypothetical Examples as Hypothetical Examples, moving them out of "Results."
6. Changed the examples from "odds ratio" to "risk ratio" due to concerns that noncollapsibility of the odds ratio might interfere with the subtraction of the Model 1 and Model 2 treatment effect estimates necessary to estimate the quantitative effect of confounding amplification.
7. Invented the term "amplifiable fraction of residual confounding" to hopefully better communicate that (if the Introduced Variable(s) has any association with outcome) it is only the residual confounding separate from that which is attributable to the Introduced Variable(s) (which is not amplified) that is able to be amplified. Hopefully this has made this clearer.
8. Removed the somewhat redundant word "Supplementary" from "Supplementary

Appendix Table.”

9. Corrected a minor subtraction error in the Appendix Table, Equation 3b (and subsequent steps), that had no substantive impact on the estimates of total residual confounding and the unconfounded treatment effect estimate. Also corrected a notation error in Step 4a where “M2” had been written “M3” by mistake.

Competing Interests: No competing interests were disclosed.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research