

On the Use of Covariate Supersets for Identification Conditions

Paul N. Zivich,^a Bonnie E. Shook-Sa,^b Jessie K. Edwards,^a Daniel Westreich,^a and Stephen R. Cole^a

Abstract: The union of distinct covariate sets, or the superset, is often used in proofs for the identification or the statistical consistency of an estimator when multiple sources of bias are present. However, the use of a superset can obscure important nuances. Here, we provide two illustrative examples: one in the context of missing data on outcomes, and one in which the average causal effect is transported to another target population. As these examples demonstrate, the use of supersets may indicate a parameter is not identifiable when the parameter is indeed identified. Furthermore, a series of exchangeability conditions may lead to successively weaker conditions. Future work on approaches to address multiple biases can avoid these pitfalls by considering the more general case of nonoverlapping covariate sets.

Keywords: Causal inference; Exchangeability; Identification; Multiple biases

(*Epidemiology* 2022;33: 559–562)

Point identification of parameters, like the average causal effect (ACE), relies on several assumptions. In many cases, exchangeability and positivity are assumed. Exchangeability is a statement about when an action (e.g., treatment and exposure, etc.) and potential outcomes are independent,¹ and positivity is a statement about the opportunity for different actions.² When there are multiple sources of bias, a series of identification assumptions are warranted for the observed data to stand-in for unobserved data. A union of the covariate sets, or the superset, that satisfies the separate identification assumptions is often used for identification or statistical consistency results. Perhaps due to its convenience and notational simplicity, supersets have been commonly used.^{3–24} Here, we

Submitted November 29, 2021; accepted March 28, 2022

From the ^aDepartment of Epidemiology, UNC Gillings School of Global Public Health, Chapel Hill, NC; ^bDepartment of Biostatistics, UNC Gillings School of Global Public Health, Chapel Hill, NC.

This work was supported in part by T32-AI007001 (Paul N Zivich), R01-AI157758 (Stephen R Cole, Bonnie E Shook-Sa, Jessie K Edwards), and K01AI125087 (Jessie K Edwards).

The authors report no conflicts of interest.

Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Paul Zivich, Department of Epidemiology, Gillings School of Global Public Health, UNC Campus Box 7435, Chapel Hill, NC 27599-7435. E-mail: zivich.5@gmail.com.

ISSN: 1044-3983/22/334-559

DOI: 10.1097/EDE.0000000000001493

provide two illustrative examples where important nuances are obscured by such supersets.

Example 1: Average Causal Effect with Missing Data

For the first example, we are interested in the ACE with missing data for the outcome (Figure). Using capital letters to denote random variables, lower-case letters for constants, and lower-case Greek letters for parameters; the parameter of interest is

$$\theta = E[Y^{a=1}] - E[Y^{a=0}]$$

where $E[\cdot]$ represents the expected value function and Y^a is the potential outcome for Y under treatment a . Further, let $A_i \in \{0,1\}$ indicate the treatment received, Z_i, X_i represent observed variables, U_1, U_2 indicate unobserved variables, and $R_Y = 1$ indicate that Y is missing and $R_Y = 0$ otherwise. For example, we could be interested in the effect of selective serotonin reuptake inhibitors (A) on incidence lung cancer over the following 10 years (Y), where smoking status (W) and coronary artery disease (X) were measured but depression (U_1) and occupational exposures (U_2) were not collected.²⁵ Furthermore, those diagnosed with coronary artery disease often later switched to a specialty hospital, and thus lung cancer diagnoses were unavailable (i.e., $R_Y = 1$). The identification strategy used here relies on two sets of sufficient identification assumptions. Equations 1–3 indicate when the treated are expected to appropriately represent the untreated, and vice versa.

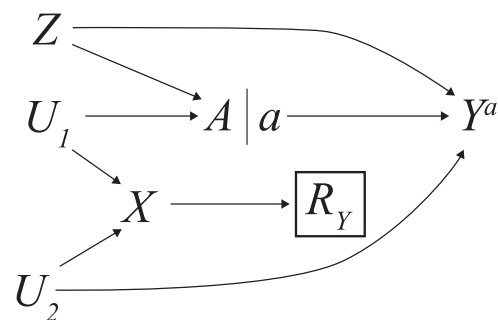


FIGURE. Single world intervention graph for confounding and missing data on Y . Let A indicate the treatment of interest, Y indicate the outcome, and R indicate missing data for the outcome. Note U_1 and U_2 are both unobserved.

$$E[Y^a | Z = z] = E[Y^a | A = a, Z = z] \text{ for } a \in \{0,1\} \quad (1)$$

$$\Pr(A = a | Z = z) > 0 \text{ for } a \in \{0,1\}$$

and all z such that $\Pr(Z = z) > 0$ (2)

$$Y_i = Y_i^a \text{ if } a = A_i \quad (3)$$

where Equation 1 is conditional exchangeability, Equation 2 is positivity, and Equation 3 is causal consistency.¹ The second set of assumptions relate to when the observations with nonmissing outcomes can stand-in for those with missing outcomes.

$$E[Y^a | X = x] = E[Y^a | R_Y = 0, X = x] \quad (4)$$

$$\Pr(R_Y = 0 | X = x) > 0 \text{ for all } x \text{ such that}$$

$$\Pr(X = x) > 0 \text{ and } a \in \{0,1\} \quad (5)$$

where Equation 4 is conditional exchangeability of missingness and Equation 5 is positivity. For ease of presentation, we assume discrete Z, X , but analogous conditions exist for continuous variables.

As depicted in Figure A, independence of A and Y^a requires conditioning on $\{Z\}$ and independence of R_Y and Y^a requires conditioning on $\{X\}$. However, X is a collider on the $A \leftarrow U_1 \rightarrow X \leftarrow U_2 \rightarrow Y^a$ path.^{26,27} Since neither U_1 or U_2 are observed, a superset that blocks all backdoor paths does not exist, and it appears that θ is not identified given the superset $\{Z, X\}$. Yet, consider the following identification proof without reliance on a superset:

$$\begin{aligned} E[Y^a] &= \sum_z E[Y^a | Z = z] \Pr(Z = z) \\ &= \sum_z E[Y^a | A = a, Z = z] \Pr(Z = z) \\ &= \sum_z \Pr(Z = z) \sum_x E[Y^a | A = a, Z = z, X = x] \\ &\quad \Pr(X = x | A = a, Z = z) \\ &= \sum_z \Pr(Z = z) \sum_x E[Y^a | R_Y = 0, A = a, Z = z, X = x] \\ &\quad \Pr(X = x | A = a, Z = z) \\ &= \sum_{z,x} E[Y | R_Y = 0, A = a, Z = z, X = x] \\ &\quad \Pr(X = x | A = a, Z = z) \Pr(Z = z) \end{aligned}$$

where the first and third steps follow from the law of total probability, the second follows from Equations 1–2, the fourth follows from Equations 4–5, and the final follows from Equation 3. Therefore, θ is identified and can be estimated with g-computation. Alternatively, $E[Y^a]$ can be re-expressed as

$$E[Y^a] = \sum_{w,x} \frac{E[Y, R_Y = 0, A = a, W = w, X = x]}{\Pr(R_Y = 0 | X = x) \Pr(A = a | W = w)}$$

which is amenable to estimation using inverse probability weighting (eAppendix 1; <http://links.lww.com/EDE/B921>), with independently constructed weights.

Example 2: Transport of the Average Causal Effect

As a second example, consider the following transportability problem where we are interested in estimating the ACE for a target population using information from a second population:

$$\psi = E[Y^{a=1} | s = 1] - E[Y^{a=0} | s = 1]$$

where $s = 1$ denotes membership in the target population and $s = 2$ denotes membership in the second population. Further, let W_i, V_i represent two sets of covariates for participant i . Here, we consider the case where W and V are nonoverlapping (see eAppendix 2; <http://links.lww.com/EDE/B921> for partial overlap). In a random sample of the target population, W, V were measured but information on A or Y is unavailable. Instead, data on both A and Y are available for $s = 2$. Therefore, the primary challenge is to demonstrate the identification conditions for ψ using the observed data [i.e., $s = 2 : (Y, A, W, V)$ and $s = 1 : (W, V)$].

Again, the identification strategy relies on two sets of identification assumptions. The first set indicates when the sample is expected to appropriately represent the target population. The identification assumptions for sampling can be written as:

$$E[Y^a | V = v, s = 1] = E[Y^a | V = v, s = 2]$$

for v where $\Pr(V = v | s = 1) > 0$ (6)

$$\Pr(V = v | s = 2) > 0 \text{ for } v \text{ where } \Pr(V = v | s = 1) > 0 \quad (7)$$

where Equation 6 is conditional exchangeability and Equation 7 is positivity for sampling.¹⁰ The second set of assumptions regard identification of the ACE in the second population:

$$E[Y^a | W = w, s = 2] = E[Y^a | A = a, W = w, s = 2]$$

for $a \in \{0,1\}$ and w where $\Pr(W = w | s = 2) > 0$ (8)

$$\Pr(A = a | W = w, s = 2) > 0 \text{ for } a \in \{0,1\}$$

and w where $\Pr(W = w | s = 2) > 0$ (9)

$$Y_i = Y_i^a \text{ if } a = A_i \quad (10)$$

where Equation 8 is conditional exchangeability, Equation 9 is positivity, and Equation 10 is causal consistency.

To demonstrate that these assumptions are sufficient to identify Ψ , we may define Q as the superset (i.e., $Q = W \cup V$) and replace both W and V with Q in Equations 6–10, provided this does not result in collider stratification bias.²⁷ Therefore,

$$\begin{aligned} E[Y^a | s = 1] &= \sum_q E[Y^a | Q = q, s = 1] \Pr(Q = q | s = 1) \\ &= \sum_q E[Y^a | Q = q, s = 2] \Pr(Q = q | s = 1) \\ &= \sum_q E[Y^a | A = a, Q = q, s = 2] \Pr(Q = q | s = 1) \\ &= \sum_q E[Y | A = a, Q = q, s = 2] \Pr(Q = q | s = 1) \end{aligned}$$

where the first step follows from the law of total probability over Q , the second from Equations 1–2, the third from Equations 3–4, and the fourth from Equation 5. One may note that the conditions in this proof are stronger than necessary and restate that only V is necessary for Equations 6–7 and only W for Equations 8–10. However, Ψ is identified under an even weaker set of assumptions. Reconsider the proof without the superset:

$$\begin{aligned} E[Y^{a=1} | s = 1] &= \sum_v \Pr(V = v | s = 1) E[Y^a | V = v, s = 1] \\ &= \sum_v \Pr(V = v | s = 1) E[Y^a | V = v, s = 2] \\ &= \sum_v \left(\frac{\Pr(V = v | s = 1) \sum_w E[Y^a | W = w, V = v, s = 2]}{\Pr(W = w | V = v, s = 2)} \right) \\ &= \sum_v \left(\frac{\Pr(V = v | s = 1) \sum_w E[Y^a | A = a, W = w, V = v, s = 2]}{\Pr(W = w | V = v, s = 2)} \right) \\ &= \sum_v \left(\frac{\Pr(V = v | s = 1) \sum_w E[Y | A = a, W = w, V = v, s = 2]}{\Pr(W = w | V = v, s = 2)} \right) \end{aligned}$$

where the first and third steps follow from the law of total probability, the second follows from Equations 6–7, the fourth follows from Equations 8–9, and the final step follows from Equation 10 (see eAppendix 3; <http://links.lww.com/EDE/B921> for the weighted estimator). Therefore, Equations 8–9 can be weakened to:

$$\begin{aligned} E[Y^a | W = w, s = 2] &= E[Y^a | A = a, W = w, s = 2] \\ &\quad \text{for } a \in \{0, 1\} \text{ and} \\ &\quad w \text{ where } \Pr(W = w | V = v, s = 2) > 0 \text{ for all } v \text{ with} \\ &\quad \Pr(V = v | s = 1) > 0 \end{aligned} \quad (8^*)$$

$\Pr(A = a | W = w, s = 2) > 0$ for $a \in \{0, 1\}$ and w where

$\Pr(W = w | V = v, s = 2) > 0$ for all v with

$$\Pr(V = v | s = 1) > 0 \quad (9^*)$$

These weakened conditions clarify that identification in $s = 2$ is only necessary for strata of v seen in $s = 1$.

For further intuition, consider the population-level information displayed in the Table. Among the second population, $s = 2$, no individuals with covariate values of $V = 0, W = 0$ receive the treatment ($A = 1$), a violation of Equation 9. This information indicates that Ψ is not identifiable. However, the target population only consists of individuals with $V = 1$ or $V = 2$. As Equation 9* indicates, positivity for treatment in the second population is only necessary for strata of V in the target population. Put another way, the identifiability in the $V = 0$ strata in the $s = 2$ population is irrelevant to the identification of Ψ . As indicated by Equations 8* and 9*, Ψ is indeed identifiable.

A remaining point is whether Equations 6–7 could have been weakened instead of Equations 8–9. Here, the answer is no. To move from the parameter to the observed data, the sampling identification conditions ought to be applied first because the treatment assumptions correspond to $s = 2$ for which information on A, Y is available.

DISCUSSION

We provided two illustrative examples where the use of covariate supersets obscures identification results or the related identification assumptions. Our examples indicate three key messages. First, relying on supersets may simplify the presentation of identification or statistical consistency proofs, but identifiable parameters may not appear to be identifiable when supersets are applied. Second, identification assumptions tailored to specific covariate sets may indicate weaker conditions than either the superset or separately considering each source of bias. Last, the order in which sets of

TABLE. Distribution of Variables in the Target and Second Populations

| | $s = 1$ | | $s = 0$ | | | |
|---------|---------|---------|----------------|----------------|----------------|----------------|
| | Overall | Overall | $A = 1, W = 1$ | $A = 1, W = 0$ | $A = 0, W = 1$ | $A = 0, W = 0$ |
| $V = 0$ | 0 | 0.45 | 0.05 | 0 | 0.10 | 0.30 |
| $V = 1$ | 0.50 | 0.35 | 0.08 | 0.05 | 0.07 | 0.15 |
| $V = 2$ | 0.50 | 0.20 | 0.05 | 0.08 | 0.02 | 0.05 |

Let $s = 1$ indicate the target population and $s = 2$ indicate the second population. For $s = 1$, information on A (received treatment) and Y (outcome) is unavailable. W is a covariate that provides exchangeability of the potential outcomes and treatment among $s = 2$. V is a covariate that provides exchangeability of the potential outcomes and sampling of populations.

identification conditions are applied may matter for weakened conditions,²⁸ with the context guiding their ordering.

Therefore, we recommend that those proposing new methods or conducting simulation studies not rely on supersets of covariates, as this can easily lead to confusion. Instead, methodological work that uses nonoverlapping covariate sets allows for epidemiologists to apply those methods regardless of the specific background knowledge. A prominent example of confusion attributable to supersets is methodological work arguing whether or not survey sampling weights should be included in the propensity score model in the survey sampling literature.^{8,21,22} In a recent online seminar, Daniel McCaffrey²⁹ clarified when sampling weights should be included by considering disjoint covariate sets. Another example showcasing the clarifying role of non-overlapping sets is Ross et al.'s recent discussion on the construction of inverse probability weights to address confounding and missing data.³⁰ When lacking formal proofs or empirical studies that avoid supersets, researchers can attempt to use graph-based algorithms,^{31–33} or may need to work through the proofs themselves.

REFERENCES

- Hernán MA, Robins JM. Estimating causal effects from epidemiological data. *J Epidemiol Community Health*. 2006;60:578–586.
- Westreich D, Cole SR. Invited commentary: positivity in practice. *Am J Epidemiol*. 2010;171:674–677.
- Seaman S, White I. Inverse probability weighting with missing predictors of treatment assignment or missingness. *Commun Stat Theory Methods*. 2014;43:3499–3515.
- Schaubel DE, Wei G. Double inverse-weighted estimation of cumulative treatment effects under nonproportional hazards and dependent censoring. *Biometrics*. 2011;67:29–38.
- Chakladar S, Rosin SP, Hudgens MG, et al. Inverse probability weighted estimators of vaccine effects accommodating partial interference and censoring [published online ahead of print March 25, 2021]. *Biometrics*. doi:10.1111/biom.13459.
- Zhang M, Schaubel DE. Estimating differences in restricted mean life-time using observational data subject to dependent censoring. *Biometrics*. 2011;67:740–749.
- Gran JM, Røysland K, Wolbers M, et al. A sequential Cox approach for estimating the causal effect of treatment in the presence of time-dependent confounding applied to data from the Swiss HIV Cohort Study. *Stat Med*. 2010;29:2757–2768.
- Ridgeway G, Kovalchik SA, Griffin BA, Kabeto MU. Propensity score analysis with survey weighted data. *J Causal Inference*. 2015;3:237–249.
- Breskin A, Cole SR, Edwards JK, Brookmeyer R, Eron JJ, Adimora AA. Fusion designs and estimators for treatment effects. *Stat Med*. 2021;40:3124–3137.
- Westreich D, Edwards JK, Lesko CR, Stuart E, Cole SR. Transportability of trial results using inverse odds of sampling weights. *Am J Epidemiol*. 2017;186:1010–1014.
- Penning de Vries BBL, Groenwold RHH. Cautionary note: propensity score matching does not account for bias due to censoring. *Nephrol Dial Transplant*. 2018;33:914–916.
- Ackerman B, Lesko CR, Siddique J, Susukida R, Stuart EA. Generalizing randomized trial findings to a target population using complex survey population data. *Stat Med*. 2021;40:1101–1120.
- Matsuyama Y, Yamaguchi T. Estimation of the marginal survival time in the presence of dependent competing risks using inverse probability of censoring weighted (IPCW) methods. *Pharm Stat*. 2008;7:202–214.
- Xie J, Liu C. Adjusted Kaplan-Meier estimator and log-rank test with inverse probability of treatment weighting for survival data. *Stat Med*. 2005;24:3089–3110.
- Wang A, Arah OA. G-computation demonstration in causal mediation analysis. *Eur J Epidemiol*. 2015;30:1119–1127.
- Kisbu-Sakarya Y, MacKinnon DP, Valente MJ, Çetinkaya E. Causal mediation analysis in the presence of post-treatment confounding variables: A Monte Carlo Simulation Study. *Front Psychol*. 2020;11. Available at: <https://www.frontiersin.org/article/10.3389/fpsyg.2020.02067>. Accessed 29 January 2022.
- Vansteelandt S, Bekaert M, Lange T. Imputation strategies for the estimation of natural direct and indirect effects. *Epidemiol Methods*. 2012;1:131–158.
- Farbmacher H, Huber M, Laffers L, Langen H, Spindler M. *Causal mediation analysis with double machine learning*. ArXiv200212710 Econ. Published online February 16, 2021. Available at: <http://arxiv.org/abs/2002.12710>. Accessed 29 January 2022.
- Gillaizeau F, Sénage T, Le Borgne F, et al. Inverse probability weighting to control confounding in an illness-death model for interval-censored data. *Stat Med*. 2018;37:1245–1258.
- Karim ME, Petkau J, Gustafson P, Platt RW, Tremlett H; BeAMS Study Group. Comparison of statistical approaches dealing with time-dependent confounding in drug effectiveness studies. *Stat Methods Med Res*. 2018;27:1709–1722.
- Austin PC, Jembere N, Chiu M. Propensity score matching and complex surveys. *Stat Methods Med Res*. 2018;27:1240–1257.
- Lenis D, Nguyen TQ, Dong N, Stuart EA. It's all about balance: propensity score matching in the context of complex survey data. *Biostatistics*. 2019;20:147–163.
- Lenis D, Ackerman B, Stuart EA. Measuring model misspecification: application to propensity score methods with complex survey data. *Comput Stat Data Anal*. 2018;128:48–57.
- Tchetgen EJ, Shpitser I. Semiparametric theory for causal mediation analysis: efficiency bounds, multiple robustness, and sensitivity analysis. *Ann Stat*. 2012;40:1816–1845.
- Liu W, Brookhart MA, Schneeweiss S, Mi X, Setoguchi S. Implications of M bias in epidemiologic studies: a simulation study. *Am J Epidemiol*. 2012;176:938–948.
- Greenland S. Quantifying biases in causal models: classical confounding vs collider-stratification bias. *Epidemiology*. 2003;14:300–306.
- Cole SR, Platt RW, Schisterman EF, et al. Illustrating bias due to conditioning on a collider. *Int J Epidemiol*. 2010;39:417–420.
- Breskin A, Westreich D, Cole SR, Edwards JK. Using bounds to compare the strength of exchangeability assumptions for internal and external validity. *Am J Epidemiol*. 2019;188:1355–1360.
- McCaffrey D. Nonrandom Samples and Causal Inference. Presented at: Online Causal Inference Seminar; January 25, 2022; Virtual. Available at: <https://www.youtube.com/watch?v=et3eUSM0mu0>. Accessed 2 February, 2022.
- Ross RK, Breskin A, Breger TL, Westreich D. Reflection on modern methods: combining weights for confounding and missing data. *Int J Epidemiol*. 2021;1–6.
- Nabi R, Bhattacharya R, Shpitser I. *Full law identification in graphical models of missing data: completeness results*. ArXiv200404872 Cs Stat. [Published online ahead of print August 31, 2020]. Available at: <http://arxiv.org/abs/2004.04872>. Accessed 31 January 2022.
- Bhattacharya R, Nabi R, Shpitser I, Robins JM. Identification in missing data models represented by directed acyclic graphs. In: *Proceedings of The 35th Uncertainty in Artificial Intelligence Conference*. PMLR. 2020;1149–1158. Available at: <https://proceedings.mlr.press/v115/bhattacharya20b.html>. Accessed 31 January 2022.
- Bareinboim E, Pearl J. Transportability from multiple environments with limited experiments: completeness results. In: *Advances in Neural Information Processing Systems*. vol 27. Curran Associates, Inc.; 2014. Available at: <https://proceedings.neurips.cc/paper/2014/hash/69adc1e107f7f7d035d7baf04342e1ca-Abstract.html>. Accessed 31 January 2022.