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, and positivity is a statement about the opportunity for different actions. 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. Here, we 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

$$\theta = E(Y_1) - E(Y_0)$$

where $E[\cdot]$ represents the expected value function and $Y_1$ is the potential outcome for $Y$ under treatment $a$. Furthermore, let $A \in \{0,1\}$ indicate the treatment received, $Z$, $X_i$ 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. 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](image-url)

FIGURE. Single world intervention graph for confounding and missing data on $Y$. Let $A$ indicate the treatment of interest, $Y$ indicate the outcome, and $R_Y$ indicate missing data for the outcome. Note $U_1$ and $U_2$ are both unobserved.
Here, we consider the case where or and and if and independence of are available for . However, and . Therefore, the primary challenge is to demonstrate the denotes membership in the second population. But analogous conditions exist using the observed data [i.e., if and as with g-computation. Alternatively,

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

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 \( \psi \) 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)

\[ \Pr(V = v | s = 2) > 0 \text{ for } v \text{ where } \Pr(V = v | s = 1) > 0 \]

(7)

where Equation 6 is conditional exchangeability and Equation 7 is positivity for sampling. 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)

\[ Y_i = Y_i^a \text{ if } a = A_i \]

(10)
where Equation 8 is conditionally exchangeability, Equation 9 is positivity, and Equation 10 is causal consistency.

To demonstrate that these assumptions are sufficient to identify \( \psi \), 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 results in collider stratification bias.\(^{27}\) Therefore,

\[
E[Y^a | s = 1] = \sum_{q} E[Y^a | Q = q, s = 1] Pr(Q = q | s = 1) \\
= \sum_{q} E[Y^a | Q = q, s = 1] Pr(Q = q | s = 1) \\
= \sum_{q} E[Y^a | Q = q, s = 1] Pr(Q = q | s = 1) \tag{8}
\]

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, \( \psi \) is identified under an even weaker set of assumptions. Reconsider the proof without the superset:

\[
E[Y^a = 1 | s = 1] = \sum_{v} Pr(V = v | s = 1) E[Y^a | V = v, s = 1] \\
= \sum_{v} \left( \sum_{W} E[Y^a | W = w, V = v, s = 1] \right) \\
= \sum_{v} \left( \sum_{W} Pr(V = v | s = 1) E[Y^a | W = w, V = v, s = 1] \right) \tag{9*}
\]

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 \( \psi \) 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. But another way, the identifiability in the \( V = 0 \) strata in the \( s = 2 \) population is irrelevant to the identification of \( \psi \). As indicated by Equations 8* and 9*, \( \psi \) 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           |
| \( A = 1, W = 1 \)     | \( A = 1, W = 0 \) | \( A = 0, W = 1 \) | \( A = 0, W = 0 \) |
| \( V = 0 \)            | 0.45              | 0.05              | 0.10              | 0.30              |
| \( V = 1 \)            | 0.50              | 0.35              | 0.08              | 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,\textsuperscript{28} with the context guiding their ordering.

Therefore, we recommend that those proposing new methods or conducting simulation studies not rely on super-sets 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 super-sets is methodological work arguing whether or not survey sampling weights should be included in the propensity score model in the survey sampling literature.\textsuperscript{8,21,22} In a recent online seminar, Daniel McCaffrey\textsuperscript{29} 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.\textsuperscript{30} When lacking formal proofs or empirical studies that avoid super-sets, researchers can attempt to use graph-based algorithms,\textsuperscript{31–33} or may need to work through the proofs themselves.

REFERENCES